Unit-1

Introduction

IntrotoCourse

Welcome to Privacy and Security in Online Social Media course on NPTEL. I am PK. I am faculty at IIIT, Delhi. I received my PhD from Carnegie Mellon University and my primary area of interest is Privacy, Security and Computational Social Science, Data Science, Social Computing and topics surrounded. I am a part of Cyber Security Education and Research Center at IIIT,Delhi. I am also a part of Research Group called Precog, which primarily works on privacy and security in online social media, computational social science, data science, social computing and usable technologies which are around these topics.

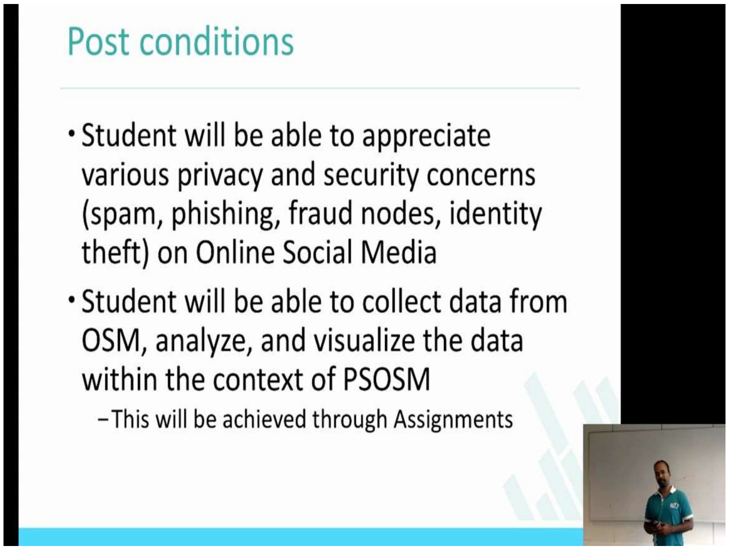
(ReferSlideTime:00:52)

Ponnurangam Kumaraguru with Madhur Rawat and 31 others.
December 14, 2015 at 10:40pm ·

Received 4.5/5 in "Course Instruction", 4.3 in "Course Management", and 4.4 in "Instructor to Students Feedback" for the CSE 648: Privacy and Security in Online Social Media, which I taught this Fall semester. 30 out of 38 students filled the course feedback. HUGE thanks to every student who took the course; was fun teaching this course for the first time in a different format. Thanks to my Teaching Assistants (Srishti, Prateek) and external evaluators. #LovingMyFacultyLife #IIITDStudentsRock

This is the Facebook post, I did about a year and half back which is regarding the course feedback that students give us at IIIT, Delhi. I have been teaching this course PSOSM Privacy and Security in Online Social Media at IIIT,Delhi for a couple of times. So, this isthefeedbackandwearetherewiththecourse.We actuallyhaveapostsessionwhere students actually present their work, what they have done over the semester in the form of a poster, in the form of a demo. So, this is the picture with all the students from the class, wearing the same T shirt and with actually the external evaluators who to came to evaluate these projects.

(ReferSlideTime:01:33)



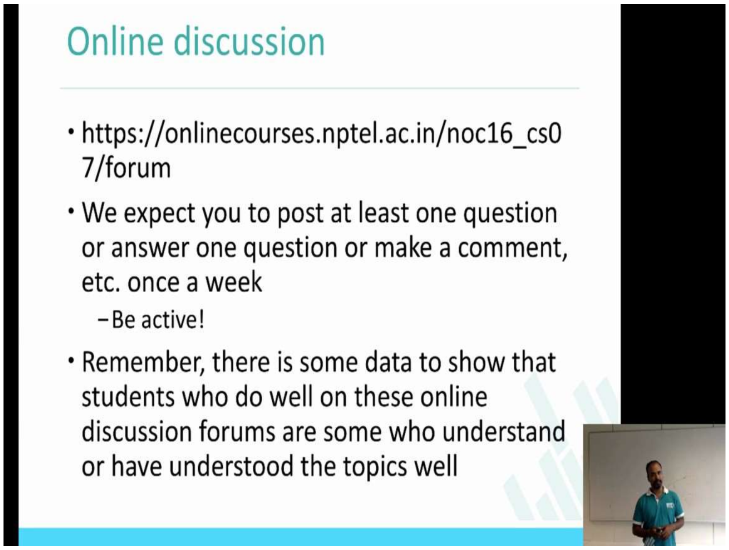
Post conditions

- Student will be able to appreciate various privacy and security concerns (spam, phishing, fraud nodes, identity theft) on Online Social Media
- Student will be able to collect data from OSM, analyze, and visualize the data within the context of PSOSM
  - This will be achieved through Assignments

So, what do you get out of this course? Some of the post conditions for this course is going to be at the end of the course, you will be able to appreciate various privacy and securityconcerns, spam, phishing,fraud, identitytheftand related issues on online social networks. Then the primary focus of this course is going to be different aspects of security and privacy on online social media.

Throughout the course, you will also be exposed to actually collecting data from online socialnetworkslikeFacebook,Twitter analyzingthese content andvisualizing thisdata in terms of the question thatyou are trying to ask, forexample, 1percent could be Iwant to understandwhether thefollowersthatIhaveonTwitterareactuallylegitimateorfake. Wecan do actually have achieve this goal, this post condition in terms of able to collect data, analyze data and visualize data through the assignments that you could be getting across these course.

(ReferSlideTime:02:36)

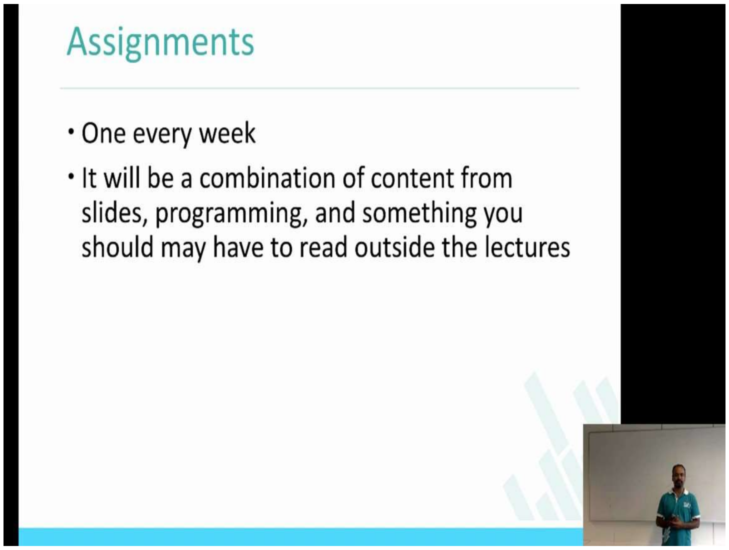


Online discussion

- https://onlinecourses.nptel.ac.in/noc16_cs07/forum
- We expect you to post at least one question or answer one question or make a comment, etc. once a week
  - Be active!
- Remember, there is some data to show that students who do well on these online discussion forums are some who understand or have understood the topics well

So, one of things that we primarily want to actually focus on is also about discussion around the topics that we will be discussing in this course. Here is a link to the online forum; we hope that you would actually participate. We expect you to actually post at least one question or answer or have a question, make comment, etcetera once per week. There is already research literature to show that people were active outside the class to perform well in the topics that are actually discussed inside the class.

(ReferSlideTime:03:11)

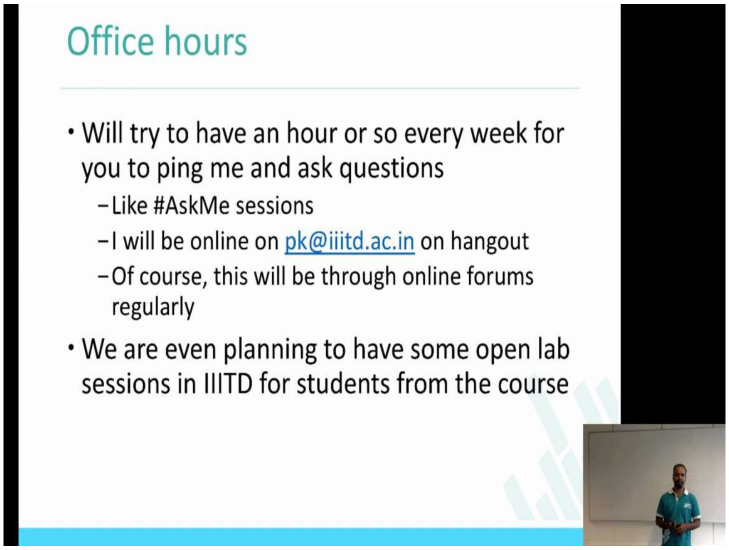


Assignments

- One every week
- It will be a combination of content from slides, programming, and something you should may have to read outside the lectures

So, what we plan to do in some of assignments is, we plan to have one homework assignment per weekcapturingthe topics thatwehave coveredin theclass,in thelecture and we will actually have homework questions around that every week and we hope to actually get you a sense of these kind of different topics same as collecting data from onlinesocialnetwork,whatkindof analysiscanbedonewiththisdata?Howtovisualize a data and things like that? Mostly these questions will be from slides, some programming; something that should actually be able to answer if you actually read content outside the lectures in the pointers that we discuss in the forum also.

(ReferSlideTime:04:00)



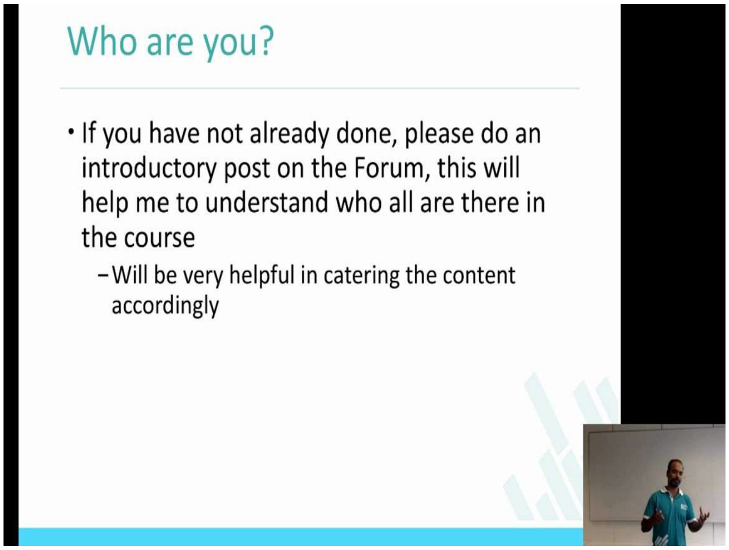
Office hours

- Will try to have an hour or so every week for you to ping me and ask questions
  - Like #AskMe sessions
  - I will be online on pk@iiitd.ac.in on hangout
  - Of course, this will be through online forums regularly
- We are even planning to have some open lab sessions in IIITD for students from the course

Ialsoplantohavesomethingtomakeitmoreexiting,tomakeitmore interestingforyou to actually participate in the topic and I also plan to have some office hours where I would be online on Hangout, where you can actually ask me questions or of course, the forum also can be used for asking these questions, it's basically like hashtag AskMe sessions that you may have heard about in the past or read it on twitter and other social networks. This will also allow you to interact with me directly, probably on hangout it could even try video sessions.

Iamalsoplanningtohave someopenlab sessions at IIIT,Delhi, where you canactually, this is only for students who are going to be mostly in and around Delhi. If you are in Delhi, you could actually show up on campus sometime at decided time that we will let you know. Join the open lab sessions where we could actually have the TAand others answer some questions for you, help you do the course better,help you understand some concepts even more deeply.

(ReferSlideTime:05:14)

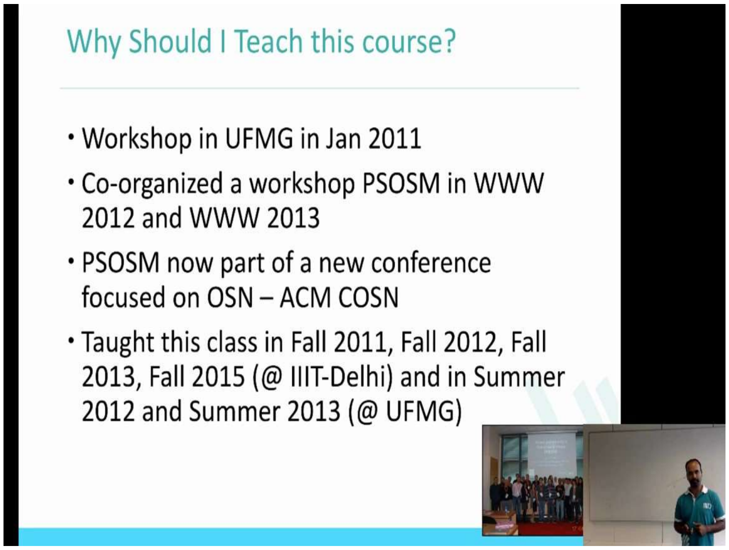


Who are you?

- If you have not already done, please do an introductory post on the Forum, this will help me to understand who all are there in the course
  - Will be very helpful in catering the content accordingly

So, I already saw some of you introducing itself on the mailing list. It is actually great to see that whether more than 1500 students who have registered for this course. It will be nice to actually have most of you introducing yourself on the forum and the main reason forme tounderstand whoyouareis actuallyhelpingme tocater thecontentaccordingly. It will help me to create content, it will help me actuallygive you appropriate pointers, if I know the proportion of the distribution of the students who are taking this class.

(ReferSlideTime:05:52)



Why Should I Teach this course?

- Workshop in UFMG in Jan 2011
- Co-organized a workshop PSOSM in WWW 2012 and WWW 2013
- PSOSM now part of a new conference focused on OSN – ACM COSN
- Taught this class in Fall 2011, Fall 2012, Fall 2013, Fall 2015 (@ IIIT-Delhi) and in Summer 2012 and Summer 2013 (@ UFMG)

Why should I teach this class? As I said earlier, I have been teaching this class at IIIT, Delhi couple of times, but before that; starting to teach this class I did things which actually makes very interesting way into teaching this course. I did a work shop at UFMG which is Universidade Federal de Minas Gerais in Brazil. I have done some workshops around the topic of privacy and security in online social media whose work converted into a conference called aconference on online social network. I also have taught this class in Brazil, which is a full credit course over the summers in 2012 and 2013.

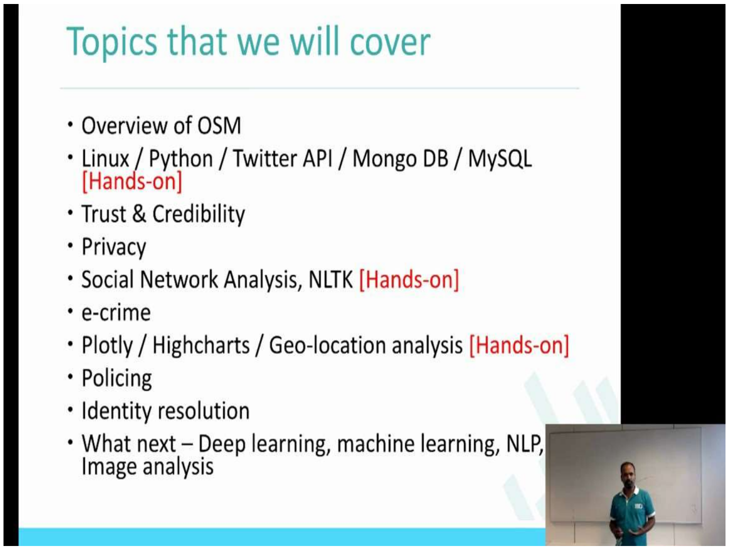
(ReferSlideTime:06:39)

Teaching Assistants!

Teaching Assistants, it is not going to be just me, you are not going to just listen to me over the entire course. Teaching assistants will help us in creating the questions, doing some lab sessions helping us in giving you more content wherever necessary,helping us even creating the home works and things like that and for now we have full fabulous TA’s, who are all my PhD students at IIIT, Delhi that is Anupama Agarwal, whose primary interest is actually understanding social reputation on social networks.

ThatisSrishtiGupta,whoseprimaryinterestisstudyingtheonlinesocialmediaandwith the phone numbers and OTT kind of technologies that are available. That is Prateek, whoseinterestis onstudying malicious contentonFacebook.Thatis NiharikaSachdeva, whose interest is primarily on visible security and studying how technology and social networks have been used by police organizations around the world and particularly in India.So,these4TA’swillhelpusindoing labsessions,settingupquestionsanswersfor the homework and helping us in generalmaking the coursemore exciting and interesting for us.

(ReferSlideTime:08:11)



Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- Trust & Credibility
- Privacy
- Social Network Analysis, NLTK [Hands-on]
- e-crime
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Policing
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

Now, let us look at topics that we will cover over this course. Initially, I will describe differentaspectsofonlinesocialnetworks,whatarethedifferentsocialnetworksthatare available which are popular? What kind of terms you need to know before we actually start delving into these social networks more deeply? Then we will also get hands-on experience with setting up python; understand how to collect data from Twitter API? How do we store the data in Mongo DB, MySQL? This will be more like a lab session where we walk you through on how to set this up.

Later in the course we will start looking at trust and credibility which is how much can you actually delete the content that are posted on Twitter or Facebook? What kind of problems exists? What kind of techniques are available to actually identify whether the post is credible or not.

Then we will also look at privacy issues on online social network. Privacy is becoming such a big topic because of the proliferation of online social networks, what information is leaked? What information can be actually collated? What information can be stitched together to create a profile user or which can be actually misused against. Then we also look at social network analysis, text analytics that can be done using the content for socialnetwork.NLTKisoneoftheplatformswhichwewillalsoexposeyoutoanalyze thecontentfromsocialnetworks.Thisagainwouldbeahands-onsessionwhereyouwill get experience on using these tools, techniques from the content in the course. E-crime which is also much related and very important topic in the context of online social networks is something we had also covered.

In this part of the course you will actually look at phishing. Wewill actuallylook at fake content, fake accounts and related topics.Wewillalso giveyou somehands-on details in terms of actually drawing graph with plotly,analyzing the data with highcharts, creating graph with high charts and also geo-location analysis because some of the content thatwe will be analyzing during the course and looking at during the course will be actually, will have information of geo-location, which is latitude longitude from a particular location where the post has been done. So, tools like these which is probably high chart, NLTK and social network analysis tools like ora will be actually very, very hands-on experience for you where you get a whole lot of ways to actually analyze the data anything that is relevant to online social networks.

Next, we look at policing which is in India, particularly if you see online social network has become such a big platform for police organizations to use in terms of interacting with citizens. So, we will actually study how police organizations are using this online social media for increasing their effectiveness of keeping this society safely and we will also look at how citizens have beem using social networks to interact with police organizations. We will not just look at only Indian context, we will also look at broader context in the world, how organizations are actually using it. There is also this whole topic of identity resolution which we will cover, which is in my case my Facebook handle is ponnurangam.kumaraguru, my Twitter handle is ponguru and my YouTube account is PK.

If you were to understand whether these three accounts are actually same is actually avery hard problem. So, we will actually look at some of the identity resolutiontechniques that people have created, how we can actually stitch these accounts together? It can be actually very useful for multiple reasons, one it could be useful for advertising agencies to actually present the ads appropriately. You could also be useful for making decisions on whether it is the same person talking about in multiple social networks.

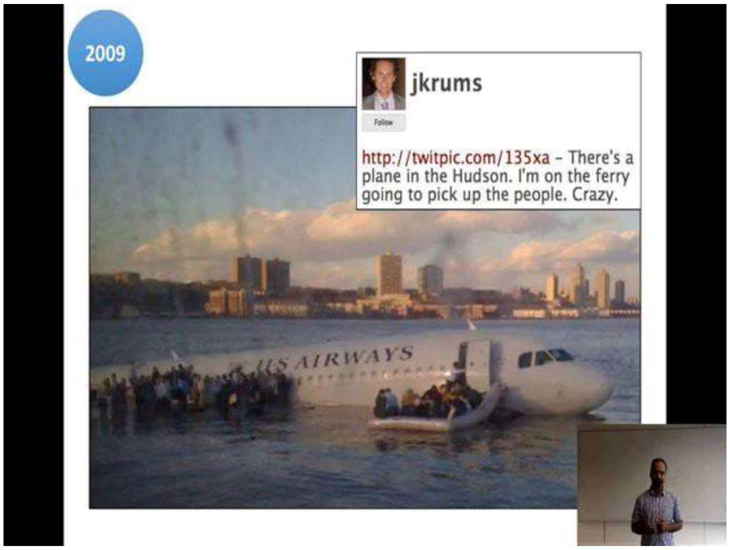
At the end of the course, we will actually review with some very broad questions and verybroadtopicswhichareconnectedtosocialnetworkswhichwewillnotabletocover in this course like deep learning, machine learning, national language processing, image analysis. These are the topics are becoming very, very popular in terms of using social network data because of the proliferation of the social network and understanding what data is available and how we can use this data is becoming a very important topic.

Mostly if you all see, image analysis is also becoming an important one because mostly these days theposts arecomingwith images. Itis notjust text only; it is actuallytext and images or sometimes only images. So, this will be a very broad tour of these new topics that are popping up around the online social network topic.

Incidents

Welcome back, until now in the course, we have seen some logistics about the course what online social media is? What is the impact and numbers and some ways in whichwe can actually use this social media services and some examples of these social media services? What I will cover today is actually looking at some of the incidences, both positive and negative where social media has actually a played role.

(ReferSlideTime:00:42)

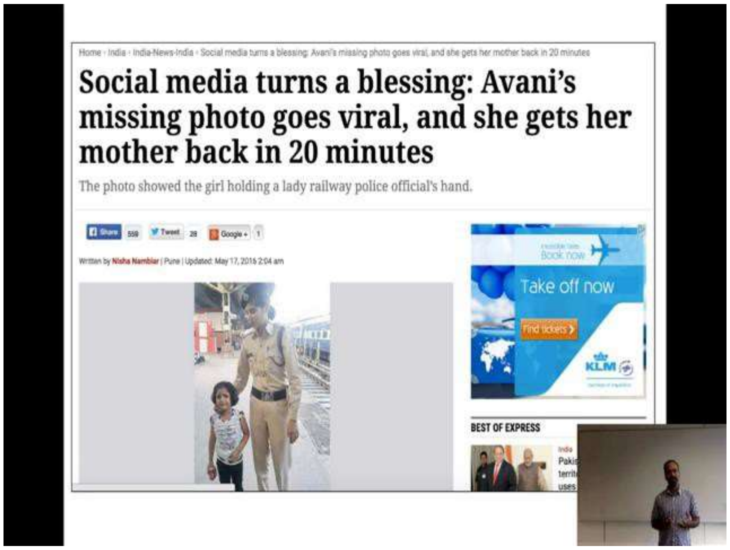


http://twitpic.com/135xa - There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy.

Here is a first one, as I have put in the slide, it happened in 2009. This is the first time ever social media service like Twitter was actually used for crisis management. Here is a tweet which actually jkrums, j k rums actually posted, which reads as, ‘There is a planein the Hudson. I am on the ferry going to pick up the people. Crazy’. Until then Twitter was basicallyused for conversations, basicallyused for saying what I am doing in life in the morning, that we are posting, ‘Morning Monday’, ‘I am having coffee’, ‘I am traveling here’ and things like that.

Whereas first time in 2009, jkrums actually posted this tweet, where when the US Airways flight landed in the Hudson river this post came out and before the first responders could reach, there was actually public, citizens who were actually helping in the situation. Therefore, Twitter and social media services have started being used in many different ways and I am going to talk to you about some examples, someincidenceswere socialmedias' playedabout positive andthe negativerole. Inthiscase it is verypositive because it helps in actually solving crisis and help in first responder.

(ReferSlideTime:02:06)



Home › India › India-News-India › Social media turns a blessing: Avani's missing photo goes viral, and she gets her mother back in 20 minutes

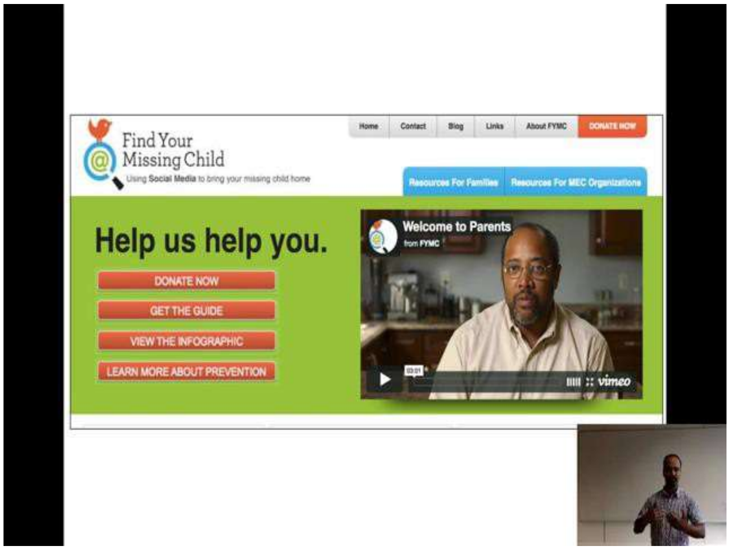
Social media turns a blessing: Avani's missing photo goes viral, and she gets her mother back in 20 minutes

The photo showed the girl holding a lady railway police official's hand.

Written by Nisha Nambiar | Pune | Updated: May 17, 2016 2:04 am

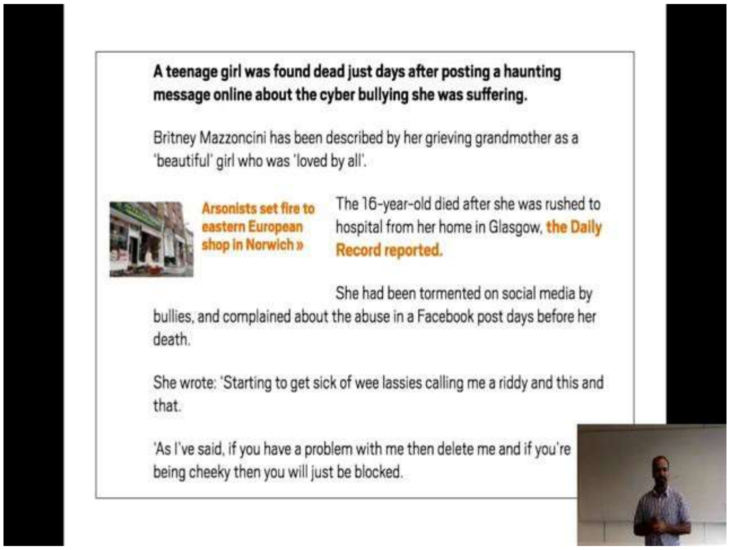
Here is an example, where in India there was a kid who is actually lost in a railway station and somebody took a picture of this kid with railway the police officer and in 20 minutes she was actually able to connect with her parents. The primary way by which it was done was actually the picture was posted on social media particularly tagging, mentioning the concern that I said earlier in my lecture mentioning the Indian railways minister and therefore this tweet got viral, this picture got viral and the kid was able to connect with her parents within 20 minutes.

(ReferSlideTime:02:50)



Keeping some of these examples, particularly with missing child there has been also organization which has been started only to help finding out missing child and parents through social media. So, here is one example which is 'find your missing child' using only social media services, and this website actually provides lot oftips on how one can use these social media services to connect with children to find out the missing children.

(ReferSlideTime:03:21)



A teenage girl was found dead just days after posting a haunting message online about the cyber bullying she was suffering.

Britney Mazzoncini has been described by her grieving grandmother as a 'beautiful' girl who was 'loved by all'.

Arsonists set fire to eastern European shop in Norwich »

The 16-year-old died after she was rushed to hospital from her home in Glasgow, the Daily Record reported.

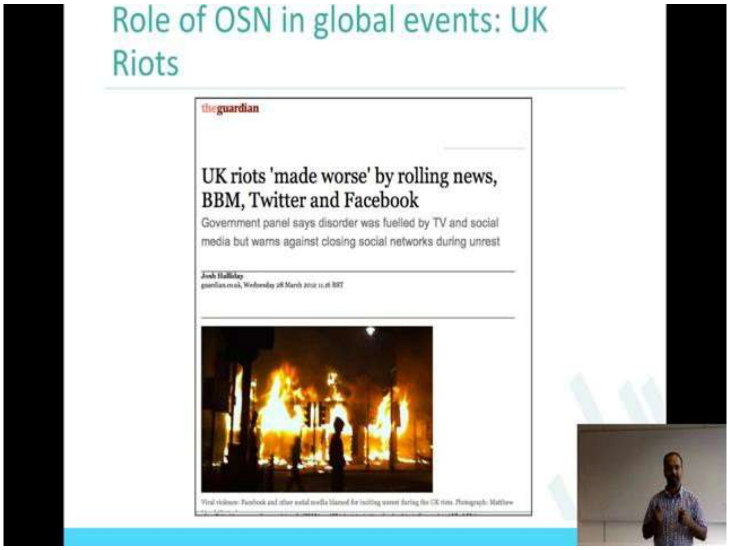
She had been tormented on social media by bullies, and complained about the abuse in a Facebook post days before her death.

She wrote: 'Starting to get sick of wee lassies calling me a riddy and this and that.

'As I've said, if you have a problem with me then delete me and if you're being cheeky then you will just be blocked.

Now, here is another example where a teenager was found dead just days after posting a messageonsocialmedia.So,thisismorelikeanegativething,wherethegirlposted information on her Facebook account, where she was saying about, she is being bullied. Cyber bullying is one of the big problems on social media also. We will talk about itlittle later in the course, but because of the cyber bullying people have actually killed themselves and many a times they actually leave a message on their own socialnetworks. In this case, she is being posting information about being cyberbullied done on Facebook, at the end of the day, she actually gives her life.

(ReferSlideTime:04:08)



Role of OSN in global events: UK Riots

theguardian

UK riots 'made worse' by rolling news, BBM, Twitter and Facebook

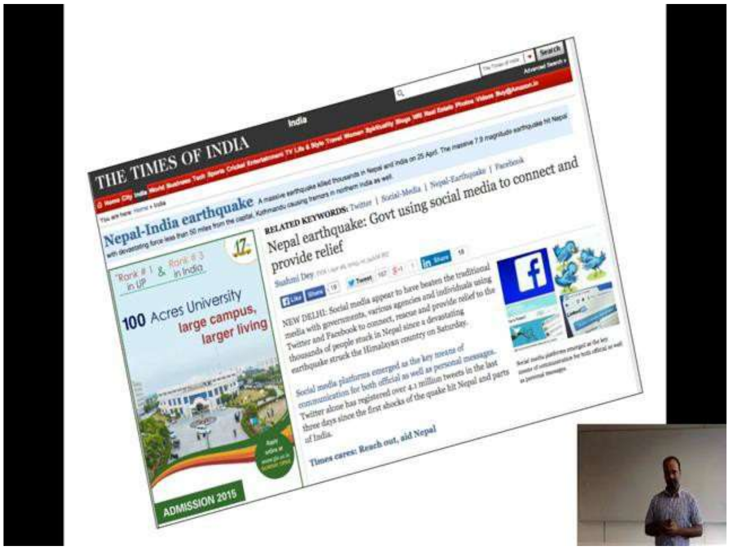
Government panel says disorder was fuelled by TV and social media but warns against closing social networks during unrest

Josh Halliday
guardian.co.uk, Wednesday 28 March 2012 11.16 BST

Viral violence: Facebook and other social media blamed for inciting unrest during the UK riots. Photograph: Matthew

Also, inthe world if you lookat it whensocialmedia was started using, were being used for crisis, this is the first time when social media was actually used to create or to propagate an incident. In the 2009 Hudson River, it was actually used for solving a problem, whereas in this case UK riots made worse by rolling news on Twitter and Facebook. So, the messages were sent on Twitter or Facebook saying that lets go to this site in this street.

(ReferSlideTime:04:47)



THE TIMES OF INDIA

India

Home City India World Business Tech Sports Cricket Entertainment TV Life & Style Travel Women Spirituality Blogs NRI Real Estate Photos Videos Buy@Amazon.in

Search

Advanced Search

You are here: Home » India

Nepal-India earthquake A massive earthquake killed thousands in Nepal and India on 25 April. The massive 7.9 magnitude earthquake hit Nepal with devastating force less than 50 miles from the capital, Kathmandu causing tremors in northern India as well.

RELATED KEYWORDS: Twitter | Social-Media | Nepal-Earthquake | Facebook

Nepal earthquake: Govt using social media to connect and provide relief

Sushmi Dey

Like Share 19 Tweet 167 +1 1 Share 18

NEW DELHI: Social media appear to have beaten the traditional media with governments, various agencies and individuals using Twitter and Facebook to connect, rescue and provide relief to the thousands of people stuck in Nepal since a devastating earthquake struck the Himalayan country on Saturday.

Social media platforms emerged as the key means of communication for both official as well as personal messages. Twitter alone has registered over 4.1 million tweets in the last three days since the first shocks of the quake hit Nepal and parts of India.

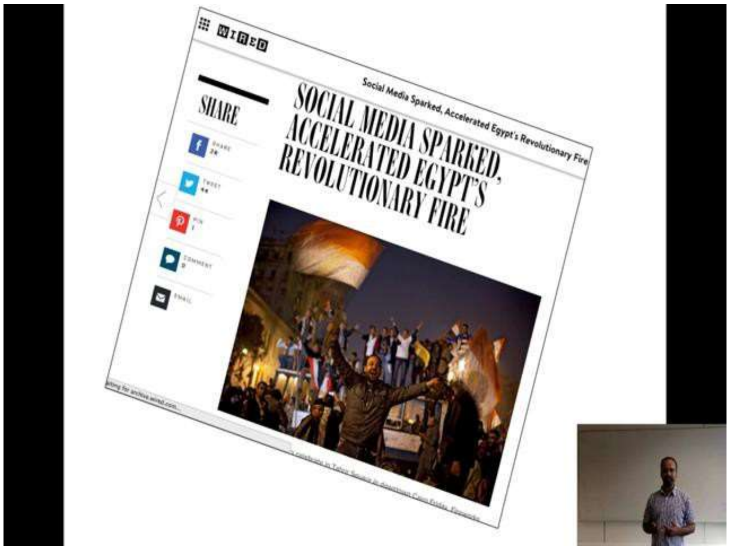
Social media platforms emerged as the key means of communication for both official as well as personal messages.

Times cares: Reach out, aid Nepal

There have been many incidences like this which you may have come across, where it is not that content on social media is talking about an incident that happens on in the real world, but these day social media itself as being used for organizing an event.

So in this example, if you see news articles which came of talking about ‘Nepal earthquake: Government using social media to connect and provide relief’. Particularlyin India, if you see there is lot more usage of social media for interacting with citizens and this is being done for the last at least a year or year and half, where it has become more, the social networks have proliferated the way that the government is organizing themselves and interacting with citizens.

(ReferSlideTime:05:39)

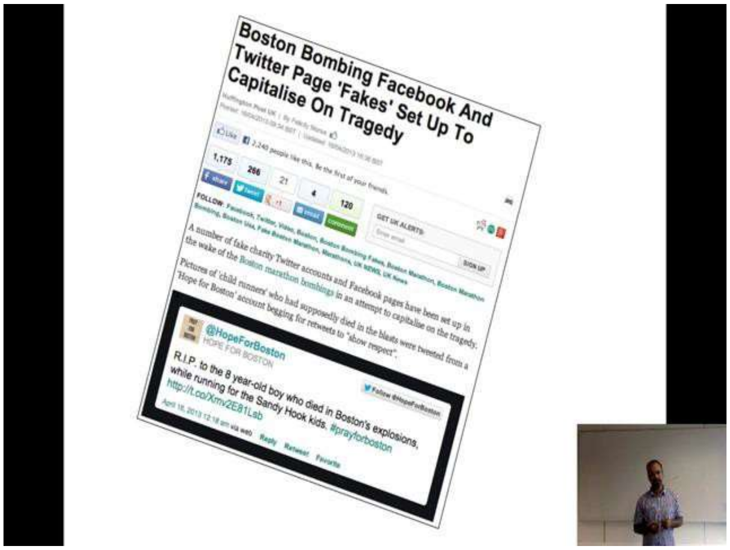


Social Media Sparked, Accelerated Egypt's Revolutionary Fire

SOCIAL MEDIA SPARKED, ACCELERATED EGYPT'S REVOLUTIONARY FIRE

Another incident that happened in the world which also is something that, if you read about the effect of social media, this incident would actually be one of the most phenomenal event that happened in terms of using social networks. So this was main social media that was used in this case was actually Twitter, where they were actually using twitter to connect with citizens giving them to one place and revolution happened becauseofusingthesesocialnetworkwhere it went veryviralthroughtheseservices like Twitter and Facebook.

(ReferSlideTime:06:19)



Boston Bombing Facebook And Twitter Page 'Fakes' Set Up To Capitalise On Tragedy

2,240 people like this. Be the first of your friends.

FOLLOW: Facebook, Twitter, Video, Boston, Boston Bombing Fakes, Boston Marathon, Boston Marathon Bombing, Boston Usa, Fake Boston Marathon, Marathons, UK NEWS, UK News

A number of fake charity Twitter accounts and Facebook pages have been set up in the wake of the Boston marathon bombings in an attempt to capitalise on the tragedy.

Pictures of 'child runners' who had supposedly died in the blasts were tweeted from a 'Hope for Boston' account begging for retweets to "show respect".

@HopeForBoston
HOPE FOR BOSTON

Follow @HopeForBoston

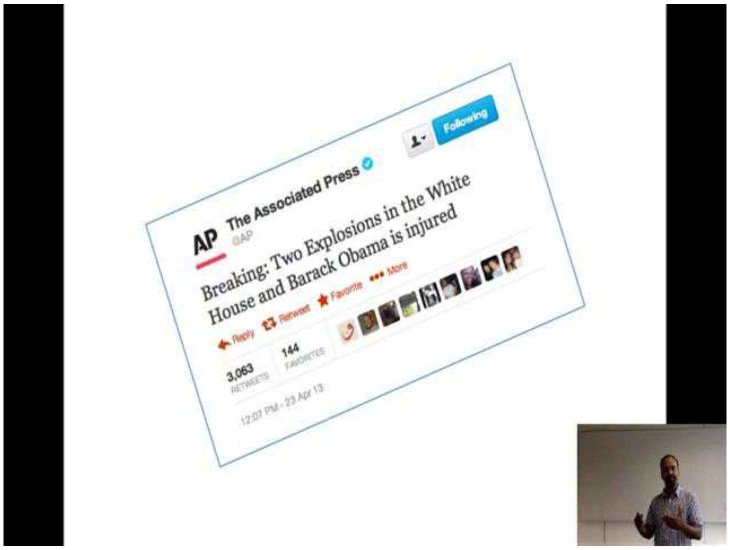
R.I.P. to the 8 year-old boy who died in Boston's explosions, while running for the Sandy Hook kids. #prayforboston http://t.co/Xmv2E81Lab

April 16, 2013 12:18 am via web · Reply · Retweet · Favorite

It is not only that social media services are being used or misused in situations like the Egypt or situations like finding kids. There is also a lot of misinformation that are actually floating around on social media services, for example, in this case Boston Bombing that happened, there was a tweet which said that, ‘RIP to the 8 year old boy who died in Boston’s explosions, while running for the Sandy Hook kids’. There was no 8 year old kid, who was actually participating in the Boston marathon. There was also another tweet during this Boston marathon said that, please RT this tweet, which is retweet this tweet, we will actually pay 1 dollars to Boston marathon. There was no money that was actually transferred to Boston marathon and this tweet get retweeted many many times, the Boston the RT tweet got retweeted more than few thousand times in couple of hours.

So, studying this misinformation is one of the main focus on this course. Also we talked about cyber bullying, cyber bullying will come back again the course. We also talked about incidences like these incidences, like Egypt revolution. We will look at these kind of incidences later in the course where we can collect data for these incidences and do some analysis around it. I also mentioned about how Indian government is using social media for their interactions with citizens. A specific module in the course we will also study about how police organization are actually using social networks to interact with citizens.

(ReferSlideTime:08:01)



AP The Associated Press
@AP

Breaking: Two Explosions in the White House and Barack Obama is injured

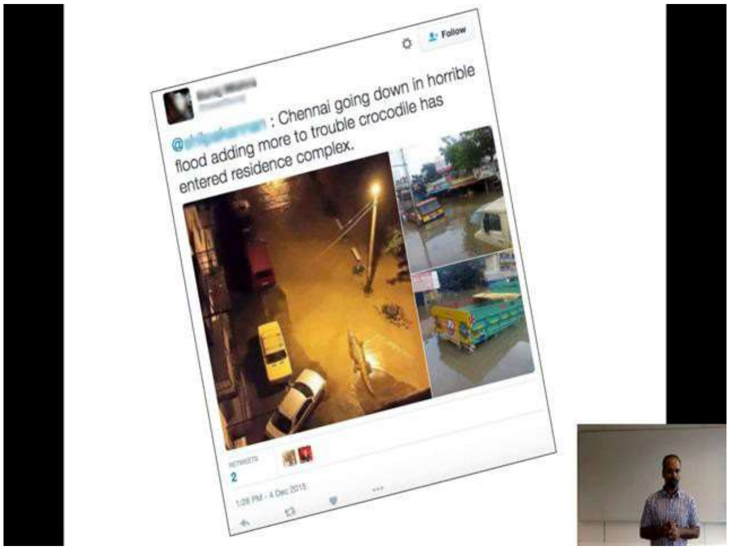
3,063 RETWEETS 144 FAVORITES

12:07 PM - 23 Apr 13

Here is another problem. Now, it is not about only this fake content, it is not only about the cyber bullying content that are being posted on social networks. Here is The Associated Press, which isa verified account; verified account means it is legitimate and it is actuallyassociatedtrust,twitter actuallyverifiesthe account if youare acelebrity, if you have lot more followers and if you are actuallya marketing companywhichcanpay for it, you can actuallyget your accounts verified. Inthis case, if you seeThe Associated Press is actually posting this tweet called, ‘Breaking: Two explosions in the white house and Barack Obama is injured’.

I am sure you already understand the implications of this specific tweet. This tweet is verified and therefore, people thought that it is a legitimate post, but unfortunately this account wascompromised, for a little bit ofthetimeand that iswhenthispost wasdone. Therefore, itis not only that the contents that are posted are not credible, butyou can also have these problems like compromised accounts. So, I am just enumeratingdifferent, using these examples and what I am trying to do is to enumerate the different problems that happen on social media.

(ReferSlideTime:09:18)

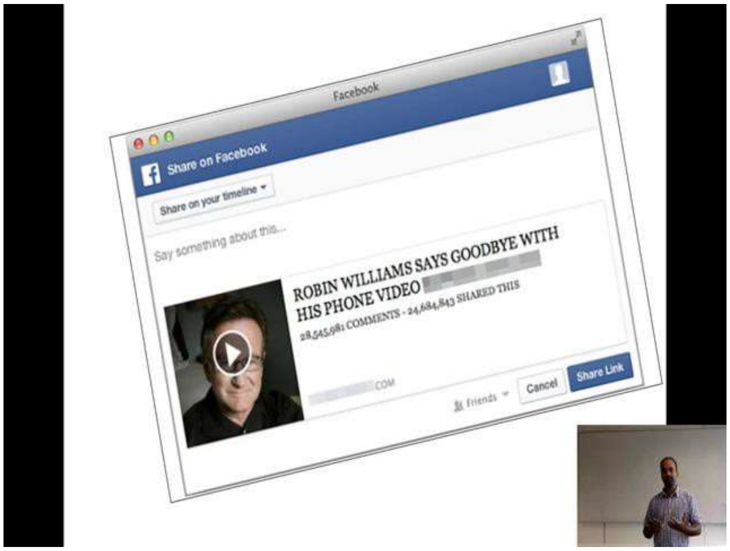


@ : Chennai going down in horrible flood adding more to trouble crocodile has entered residence complex.

Also here is another one that I was also following very closely. This is Chennai floods November-December 2015 and when the floods happen there was a lot of uproar about one of these pictures, which is crocodile on the streets and there is no crocodile, there is no alligator on streets. So, this is another way by which actually suchmis-information is beingspreadandthere have beenincidences inthe past wherehurricanesandyintheUS, where they had a picture with the shark in the water and this is crocodile in the water in the street in Chennai.

So, these kind of problems which is not only the text problem, there is also with the imageproblemwhere inthe fake imagesare being postedonsocialmedia whichbecome viral which also has impact in the society.

(ReferSlideTime:10:12)



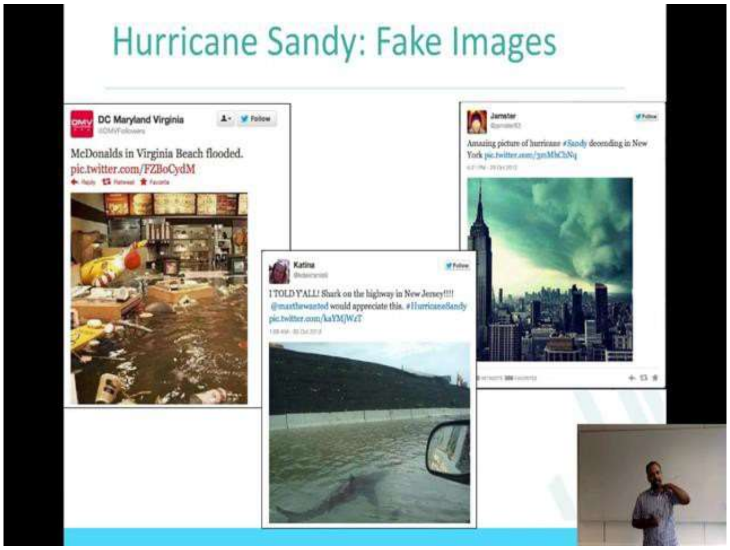
ROBIN WILLIAMS SAYS GOODBYE WITH HIS PHONE VIDEO [redacted]

28,545,981 COMMENTS - 24,684,843 SHARED THIS

[redacted].COM

So, here is another interesting problem that happened about a year or two years back. Robin Williams, when he passed away, there was a good bye message for him, which lookedasheactuallypostedthis videoonFacebook, but thiswasactuallya fakevideo, it was not a video that was taken before his death. So, situations like these, which are floods, incidences like these to death to Robin Williams, Boston marathon blast, people use these situations to create malicious content, content that is not credible on social media. So, this is another problem which we will actually study in detail.

(ReferSlideTime:10:52)



Hurricane Sandy: Fake Images

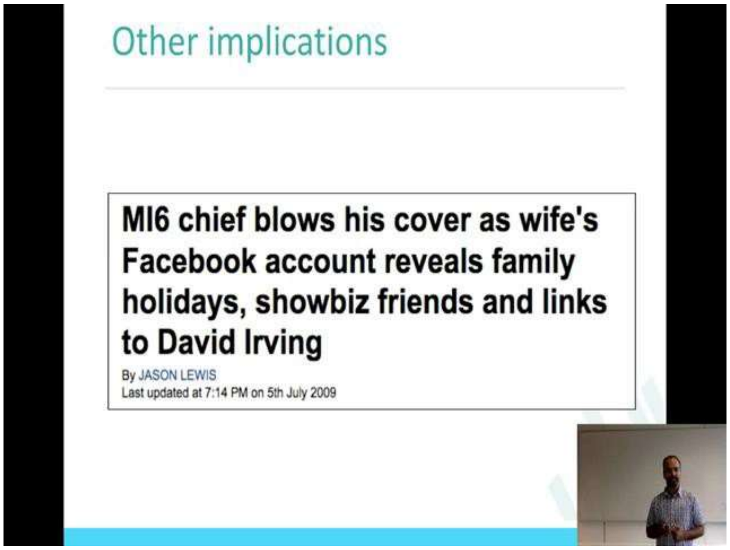
DC Maryland Virginia
@DMVFollowers
McDonalds in Virginia Beach flooded. pic.twitter.com/FZBoCydM

Katina
I TOLD Y'ALL! Shark on the highway in New Jersey!!!! @maxthewanted would appreciate this. #HurricaneSandy pic.twitter.com/kaYMJWzT

Jamster
Amazing picture of hurricane #Sandy decending in New York pic.twitter.com/3mMhChNq

So,this is againexample, that Italked fewsecondsbackwhich isthethree images inthis slide. The one on the left which is McDonalds the tweet, which is ‘McDonalds in Virginia beach flooded, which is actually a picture of McDonalds, but it was not taken from Virginia beach, it is actually taken somewhere else in the other part of the world few years before and the image in the middle is the image that I will talk about, which looks like a shark in the water, while the hurricane sandy was going on and the third image on the right hand top is actually the image from a movie which was used to say that is how it looks now, when this hurricane is going on. So, all these are fake images, but they were all used in an incident like hurricane sandy to propagate malicious intent,to propagate these kinds of information which is not credible.

(ReferSlideTime:11:50)



Other implications

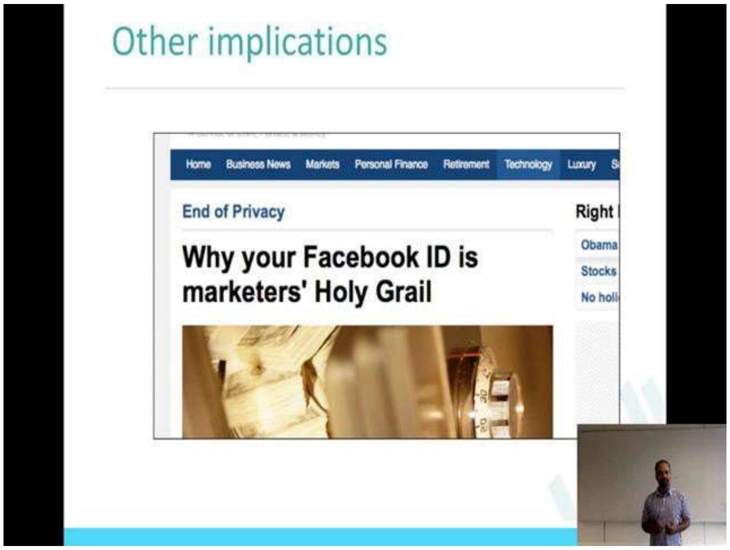
MI6 chief blows his cover as wife's Facebook account reveals family holidays, showbiz friends and links to David Irving

By JASON LEWIS
Last updated at 7:14 PM on 5th July 2009

So, until now we saw some events that where having talking about malicious content, how people connect with government organizations, how these information can be actually used to reconnect with parents, and things like that, but here is an examplewhich also happened couples of years, some years back which is MI6, which is military intelligence chief could not take his job because his wife actually posted some pictureson Facebook, which pictures also became public and the picture showed that this MI6 chief who was going to become a chief his being with some people whom he should not have been with.

Therefore, there is also a privacy issue which we will talk about which is how much information about people around you is getting revealed through these social networks. So, there is this whole idea of privacy the whole idea of using policing and online social networks, there is also credible information, misinformation on social network. So, this are the different topics as I said in one of my early lectures about the topics that we will cover.

(ReferSlideTime:13:07)



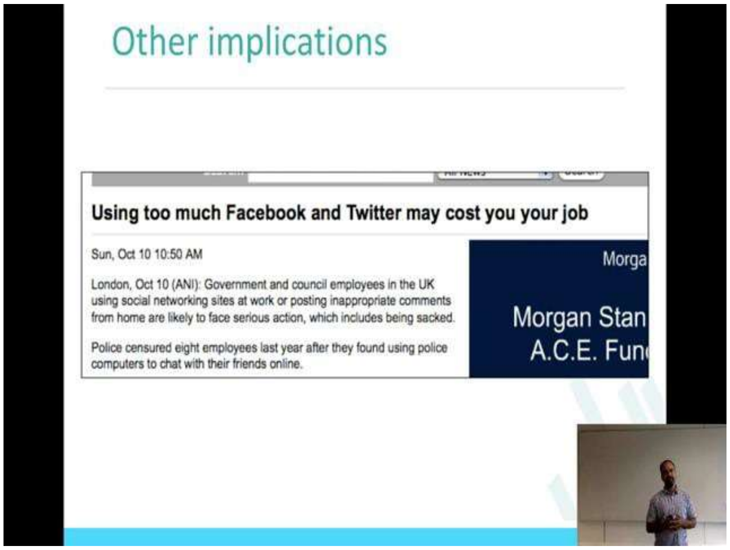
Other implications

End of Privacy

Why your Facebook ID is marketers' Holy Grail

Now, let me talk to you about whyFacebook id knowing what throughyour Facebookis important because I can actually use the information, who you are and where you are from,thingslikethat fromyourFacebookaccountandactuallysendrelevant information to you and that is why Facebook, understanding Facebook account, understandingFacebook handle becomes such an important topic also which is personalization, whichis understanding user behavior and things like that targetted advertisement and topics around it.

(ReferSlideTime:13:43)



Other implications

Using too much Facebook and Twitter may cost you your job

Sun, Oct 10 10:50 AM

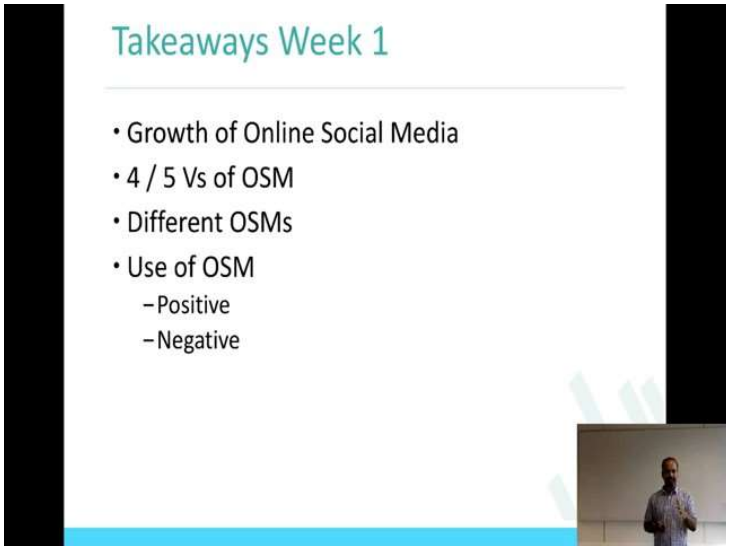
London, Oct 10 (ANI): Government and council employees in the UK using social networking sites at work or posting inappropriate comments from home are likely to face serious action, which includes being sacked.

Police censured eight employees last year after they found using police computers to chat with their friends online.

A few more implications about social media also, if you look at this one where there are many people around the world who lost the job because they have been using Facebook and Twitter too much.

So, therefore, this implication which is about usage of social networks itself and organizations keeping track about what you are posting and there have been incidences also where people are posted employee ofcompaniesposted about projectstheyworking on. They should not be talking about on social media and when these contents goes in hands of people whom they should not be looking and this information can actually be used against the company, the company actually takes very a strong thing around thing against the employee itself.

(ReferSlideTime:14:31)

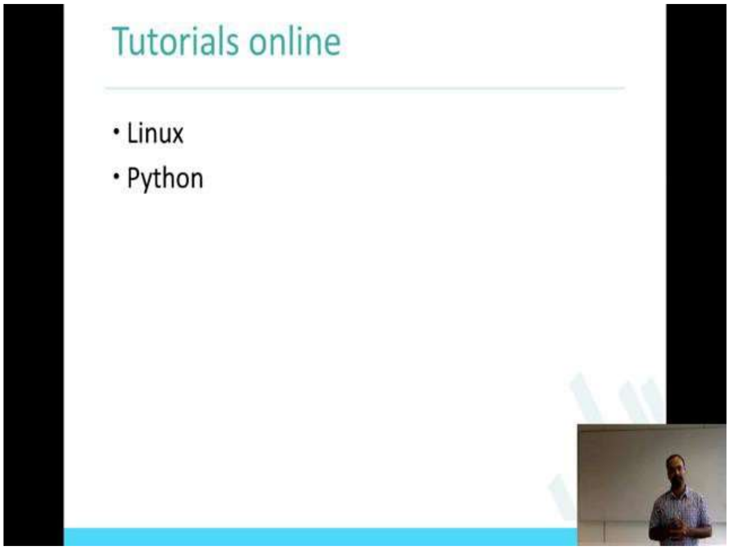


Takeaways Week 1

- Growth of Online Social Media
- 4 / 5 Vs of OSM
- Different OSMs
- Use of OSM
  - Positive
  - Negative

Now,just wrappinguptheweek1content,what allwehavedone, wehavedoneactually growth of online social media, which is how large it is? What is it? How much of contents is getting generated on 60 seconds something like that and we also looked at velocity,volume,variety,valueand veracity.So,thoseare5V'sthat wetalkedaboutand in different social networks like Facebook, Twitter, Linkedin, Google plus, Whisper, Periscope, Tinder, these arethe different social networks also we saw about. And now in this lecture I talked about some incidences that where social media is used for both positive and negative implications. So, this is the topics that we covered.

(ReferSlideTime:15:23)



Tutorials online

- Linux
- Python

We have also uploaded content on setting up your machine to use Linux and python, these are two tutorials that we have uploadedfor this week. So, as I was sayingin the first part of the introduction of the course we will actually have some hands-on tutorials overthesemester. So,thiswillactuallyhelp youto gethands-onand inparticularlysome of things that we are going talking early in the course will help you actually tounderstand and do things by yourself.

Aswe move forwardinthecourse, forexample, youneedto write Pythoncodeto collect the data fromTwitter and Facebook, you will have to get yourself familiarized withthat, it is not that the difficult the tutorials about 25 to 30minutes together Linux and python. If you can set things up if you can understand little bit about, how to set up Linux and python and set it up as earlier as possible in a machine that you would have a access to which you can use it for the entire course time.

(ReferSlideTime:16:32)

https://www.facebook.com/PreCog.IIITD/

Precog@IIIT-D

Our paper "Hiding in Plain Sight: Characterizing and Detecting Malicious Facebook Pages" accepted at IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) 2016 http://asonam.cpsc.ucalgary.ca/2016/ Authors: Prateek Dewan, Shrey

And last two slides I had was, there is Facebook page that we have, the group which is Precog where we actually talk about lot of things that we going to be discussing in the course also, which is in this page manyofthings that we talk about or share is the things activities that are around the topics that I have actually have been talking about in this course and things that the students do and then ifthere is interesting, that happens onthe topic,wewouldalso share it here. So,Iwould highlyrecommend youto lookatthispage Precog dot IIIT-D on Facebook.

(ReferSlideTime:17:10)

http://precog.iiitd.edu.in/

is a group of researchers who study, analyze, and build different aspects of social systems (e.g. social web systems like Twitter, Facebook), including their security and privacy. By understanding and measuring complex networks, we try and build solutions for social good. Our work primarily derives from Data Science, Computational Social Science, Social Computing, Machine Learning, and Natural Language Processing.

What's New!

New! July, 2016: Our paper Hiding in Plain Sight: Characterizing and Detecting Malicious Facebook Pages accepted at IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) 2016. Authors: Prateek Dewan, Shrey Bagroy and PK.

New! June, 2016: PK's Keynote Address on Online Social Media: Opportunities, Challenges, and Pitfalls at IGDTUW.

New! June, 2016: Our paper Social Media for Safety: Characterizing Online Interactions between Citizens and Police accepted at 30th British Human Computer Interaction Conference (BHCI) 2016. Authors: Niharika Sachdeva, PK and Munmun De Choudhury.

Hiding in Plain Sight: Characterizing and Detecting Malicious Facebook Pages

It is very similar to the page, this is the website that we have actually maintain precogdot iiitd dot edu, this is a website where we actually collate all the information that the group, the work ofthe group that we do together. So, I highly recommend you to look atthese two, Facebook page and this web page for any updates on this topics around if not only for this course, beyond this course also because I think it will always be good to have updated content or the latest contents on this topics.

(ReferSlideTime:17:43)

Thank you

pk@iiitd.ac.in

precog.iiitd.edu.in

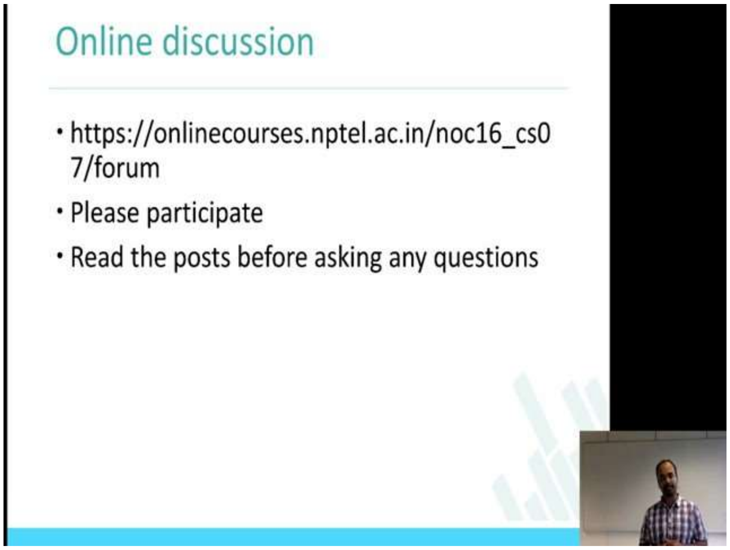
fb/ponnurangam.kumaraguru

With that I will wrap this for week 1. So, what we will do is the week is, next week we willtalk little bit aboutoneofthetopicsthat I coveredverybrieflywithincidencestoday and we will also have some hands-on session tutorials also uploaded for next week. So, there will be also homework upload for week 1.

Please tryto attempt it and if there are any questions, feel free to post it on the forumon NPTEL online course website and I'll be happy to actually answer some of them. If it is something that we could actually answer, you cannot be asking what is the answer to these questions. Therefore, I will sign-off from here and see you in the next lecture.

OSMAPIsandtoolsfordata collection

WelcomebacktothecourseonPrivacyandSecurityinOnlineSocialMedia,week2. (Refer Slide Time: 00:16)

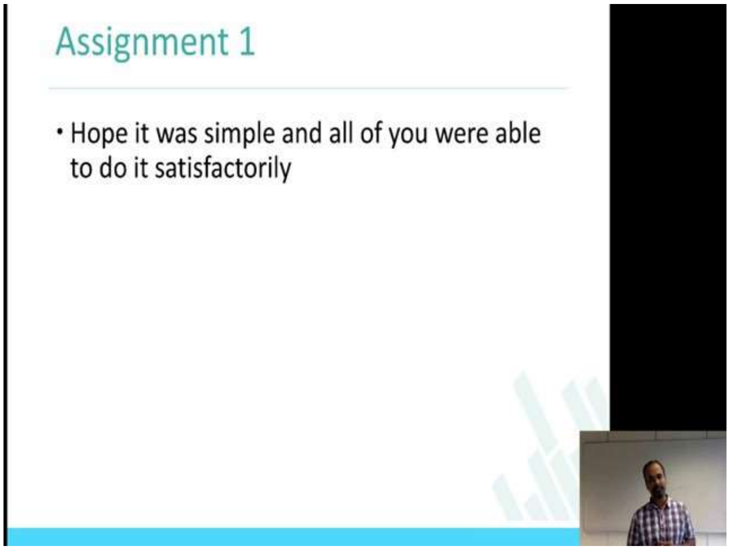


Online discussion

- https://onlinecourses.nptel.ac.in/noc16_cs07/forum
- Please participate
- Read the posts before asking any questions

I hope you are participating in the online forum that we have in the course. I already seea lotofpeopleaskingquestions, andtryingto answer.Mysincererequest will be,please, pleasereadthepostsbeforeyouactuallyaskthequestion;thatis,readthepoststhathave been already asked, the questions that have already been asked and the answers that has been already given, before asking the question. And, please participate also in the online forum, not just only asking questions; if you know the answers for the questions that others are asking, please try and answer them also.

(ReferSlideTime:00:52)



Assignment 1

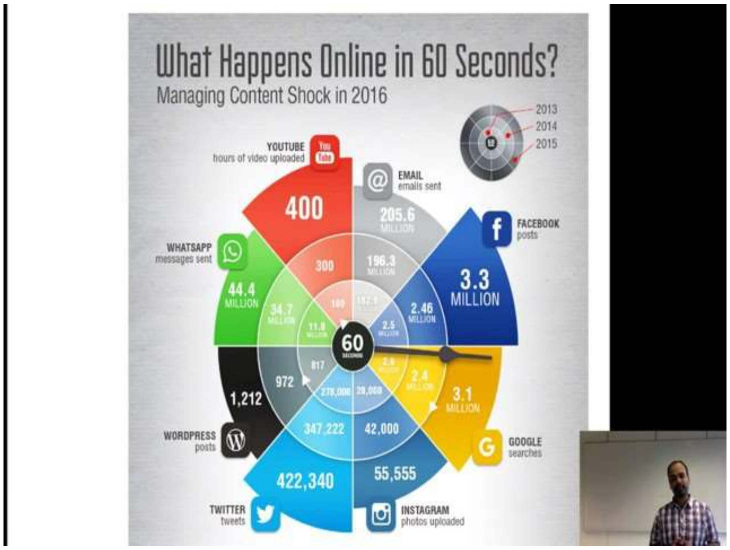
- Hope it was simple and all of you were able to do it satisfactorily

I hope most of you got to see the assignment 1 that we had posted. So, I think, the weekendoftheweek2isthedeadlinefortheassignment1.Pleasetrytoworkitout.The assignment 1 is actually pretty simple. We have just captured some questions from the slides that we did, and some from the tutorials.

(ReferSlideTime:01:15)

So,letmejustgiveyouaquicksummaryofwhatwehaveseenuntilnowandthen,Iwill goaheadwiththetopicsthatIwantedtocovertoday.So,first,wesawwhatsocialmedia is;differenttypesofsocialnetworks,differenttypes ofcontentthatgetsgeneratedonour social network; classical online social media services, and then some, which are more like ephemeral social networks and anonymous social networks.

(ReferSlideTime:01:41)



What Happens Online in 60 Seconds?

Managing Content Shock in 2016

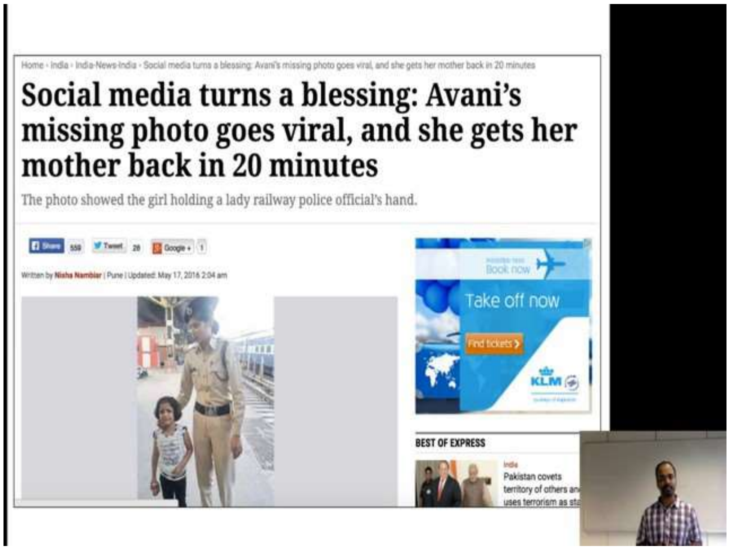
We also saw what online social media means in 60 seconds; so, how much of data is getting generated on social media in 60 seconds. We saw 400 hours of videos uploaded on YouTube, and 3.3 million posts are done on Facebook and things like that. This basically shows us that, large amount of content that are getting generated on online social media services.

(ReferSlideTime:02:09)

4 / 5 V's of Online Social Media

We also saw what 4 or 5 V's of online social media are, they are volume, velocity, veracity, variety and value - those are the 5 V's of online social media.

(ReferSlideTime:02:25)



Home › India › India-News-India › Social media turns a blessing: Avani's missing photo goes viral, and she gets her mother back in 20 minutes

Social media turns a blessing: Avani's missing photo goes viral, and she gets her mother back in 20 minutes

The photo showed the girl holding a lady railway police official's hand.

Written by Nisha Nambiar | Pune | Updated: May 17, 2016 2:04 am

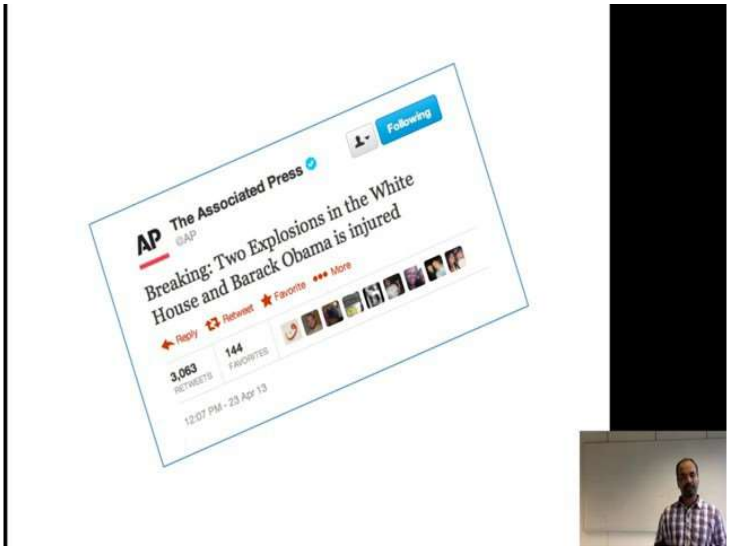
BEST OF EXPRESS

India

Pakistan covets territory of others an uses terrorism as sta

And then, I looked at, I showed you some events, where online social media has played an important role in the real world and in the society also.

(ReferSlideTime:02:37)



The Associated Press
@AP

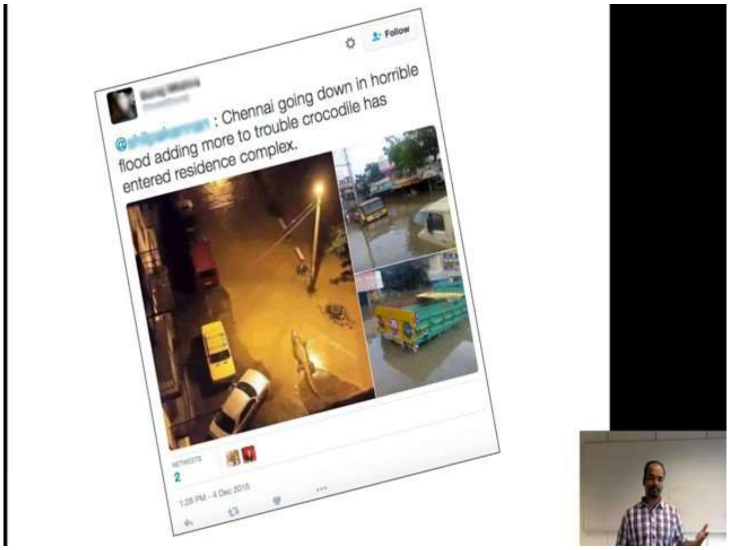
Breaking: Two Explosions in the White House and Barack Obama is injured

3,063 RETWEETS
144 FAVORITES

12:07 PM - 23 Apr 13

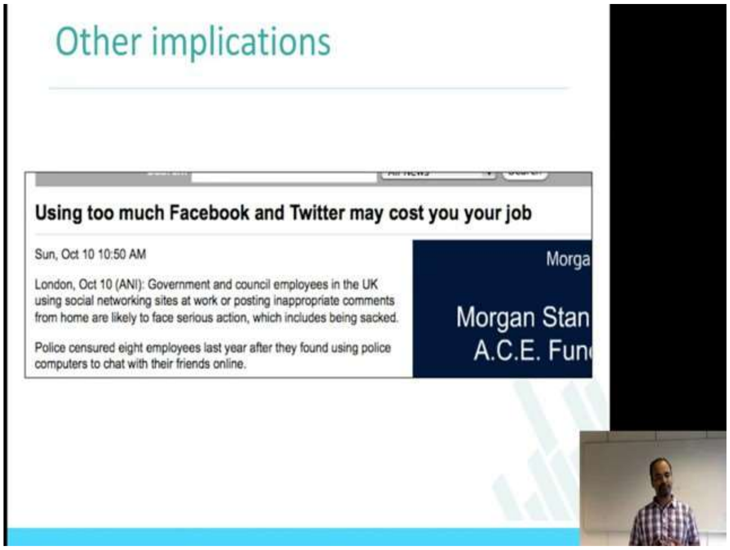
Telling you about different issues on online social media, for example, in this case, it is compromised account; an account was compromised, where the post said 'Two explosionsinWhiteHouseandBarrackObamainjured',and,therewas,therewasafter effects ofthis tweet. So,welookedat different issues thatare happening ononline social media; compromised account, fake content in this case; and image of a crocodile on the streets of Chennai, while Chennai floods was going on in December 2015, caused panic among citizens.

(ReferSlideTime:02:59)



@[redacted] : Chennai going down in horrible flood adding more to trouble crocodile has entered residence complex.

(ReferSlideTime:03:11)



Other implications

Using too much Facebook and Twitter may cost you your job

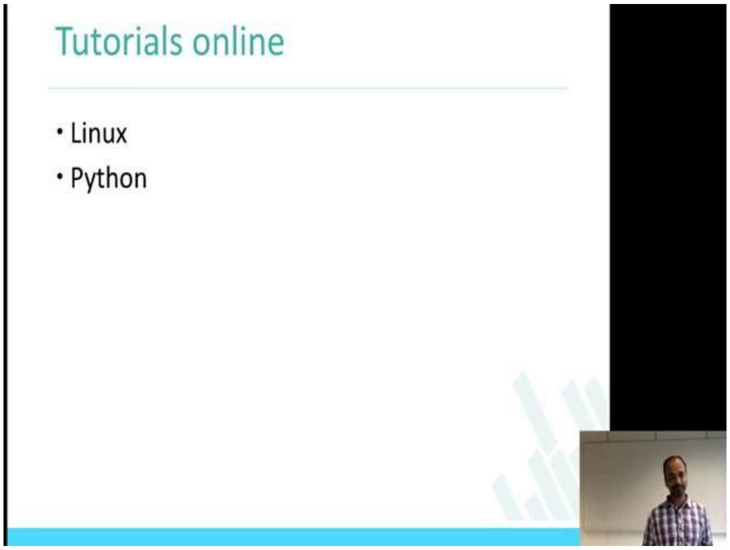
Sun, Oct 10 10:50 AM

London, Oct 10 (ANI): Government and council employees in the UK using social networking sites at work or posting inappropriate comments from home are likely to face serious action, which includes being sacked.

Police censured eight employees last year after they found using police computers to chat with their friends online.

And, there are also people who lose jobs and others issues, because of the usage ofonline social media.

(ReferSlideTime:03:19)

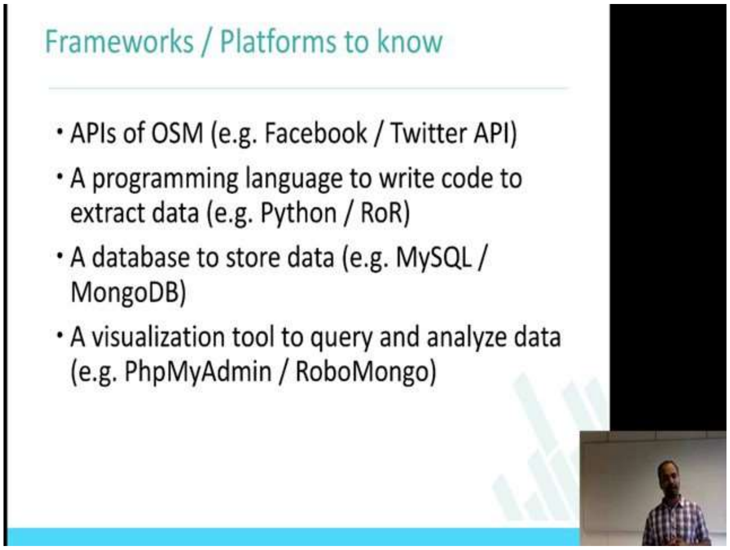


Tutorials online

- Linux
- Python

And, in the week 1, we also covered a little bit about Linux and python; hopefully, you are all set, in terms of using the platforms, because, I think, there were some questions about, ‘can we use windows?’You should be able to use windows, and do programs on python, but it just said our support will be mostly on Linux. And, of course learning Linux will also be good for you.

(ReferSlideTime:03:44)



Frameworks / Platforms to know

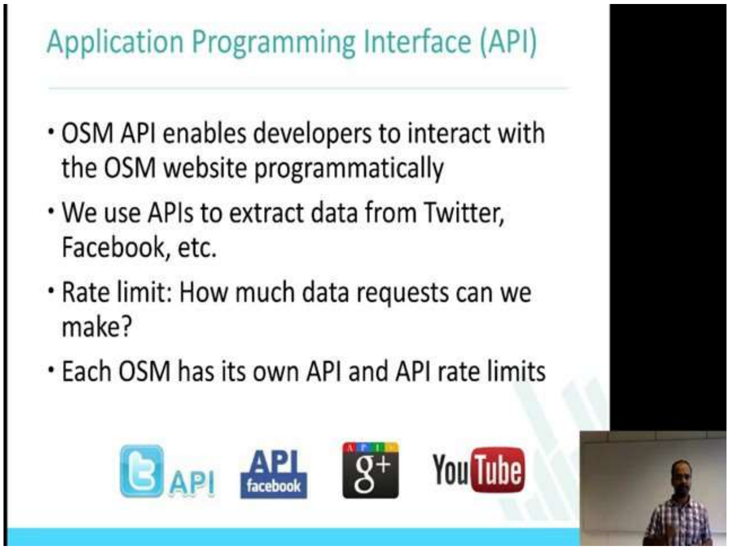
- APIs of OSM (e.g. Facebook / Twitter API)
- A programming language to write code to extract data (e.g. Python / RoR)
- A database to store data (e.g. MySQL / MongoDB)
- A visualization tool to query and analyze data (e.g. PhpMyAdmin / RoboMongo)

So, what I want to cover today is a couple of things; one, I wanted to actually look at differentframeworksorplatforms,thatyouwouldgettoknowwhiledoingthiscourse, or in another terms, you should know while doing this course, and collecting data from online social media, analyzing and making inferences.

We will look at what an API is; different kinds of APIs that are available for Facebook and Twitter.Then, we will also look at programming language. There has been a tutorial on python. So, I will just quickly go over.In any case, my work for this week 2.1, about these topics, are onlygenerally,to introduceand then, wewill haveatutorials, which are specifically focused on some of them.

Then, we look at programming languages; and then, we will also look at a little bit of database, how this data is stored, what kind of format that the data is coming out; and a little bit about visualization tool.

(ReferSlideTime:04:45)



Application Programming Interface (API)

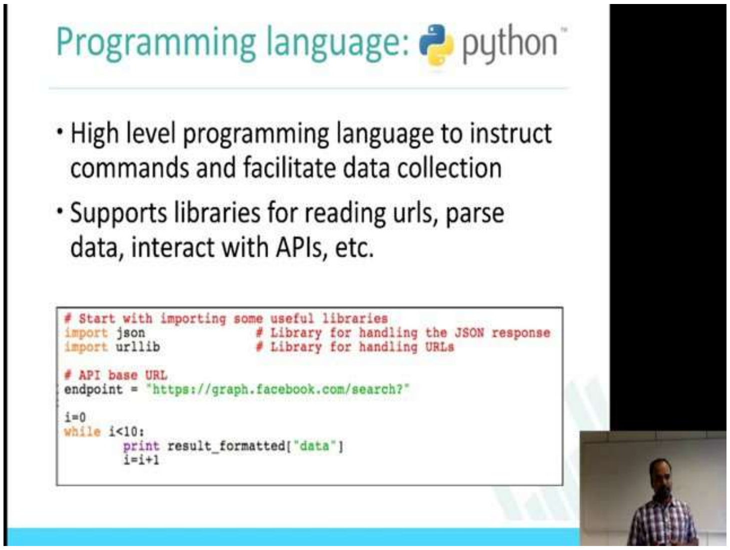
- OSM API enables developers to interact with the OSM website programmatically
- We use APIs to extract data from Twitter, Facebook, etc.
- Rate limit: How much data requests can we make?
- Each OSM has its own API and API rate limits

First, API, which is Application Programming Interface; this basically enables you to interact with the online social media, programmatically, and collect data from there. What does this mean? This basically means that, you can actually have a tunnel that is from your program to the social media services, to collect data. It just creates a tunnel between your program and the online social media services, where you are going to ask some data and then, the social media service is going to respond with saying, here is the data that you asked for, right.

Particularly, in our case, we will actually look at APIs for Facebook and Twitter, which will help you to collect data from Facebook and Twitter. There is other APIs also; all other social media services or majority of the social media services provide you with an API. We can't cover everything in this course. So, we are going to start looking at only the most popular ones, or the ones that we can actually use for this course, which will help you to understand howAPIs work, what data can be collected. So, you can actually do it for other social media services, yourself.

So, one of the important thing that you want to also keep in mind is that, about the rate limit, which is that in social media services when we want to collect data you cannot collect the data everything that is available on social, on these services. Because, I am sure, the companies do not want to give you all the data also. They have set it up, you know, by saying that, they have a rate limits for every social media service, and every piece of data that we want to collect from them. So, we will look at something in the tutorials about rate limits, particularly about each of the social media API , but I wanted to just give an idea about, there is going to be a rate limit, in terms of the data you can collect from these services.

(ReferSlideTime:06:33)



Programming language: python

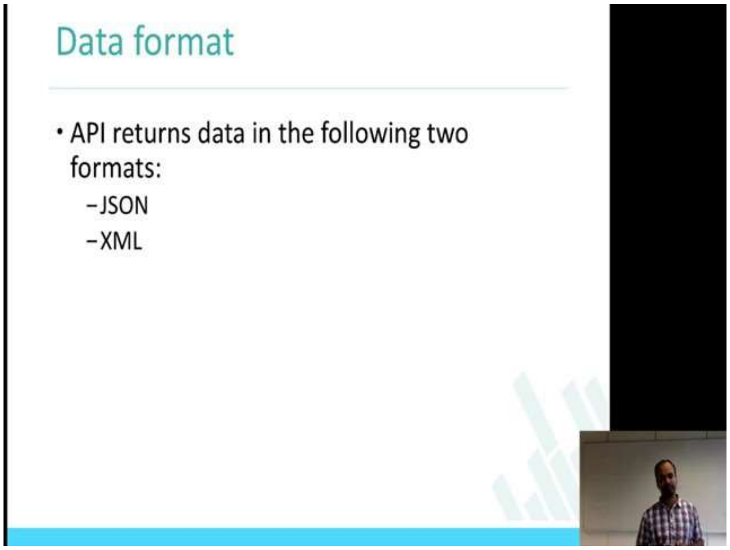
- High level programming language to instruct commands and facilitate data collection
- Supports libraries for reading urls, parse data, interact with APIs, etc.

```
# Start with importing some useful libraries
import json                 # Library for handling the JSON response
import urllib               # Library for handling URLs

# API base URL
endpoint = "https://graph.facebook.com/search?"

i=0
while i<10:
        print result_formatted["data"]
        i=i+1
```

Also then in python, since you have already done a tutorial on python, I will keep it reallyshort.Itisbasicallyaprogramminglanguage, thatisusedtocollectdata andisone of the popular languages currently used in terms of writing API requests to the social media services.

And, it also has a lot of libraries for reading URLs, parsing data, interact with API, and understanding the JSON objects, and things like that.

(ReferSlideTime:07:00)

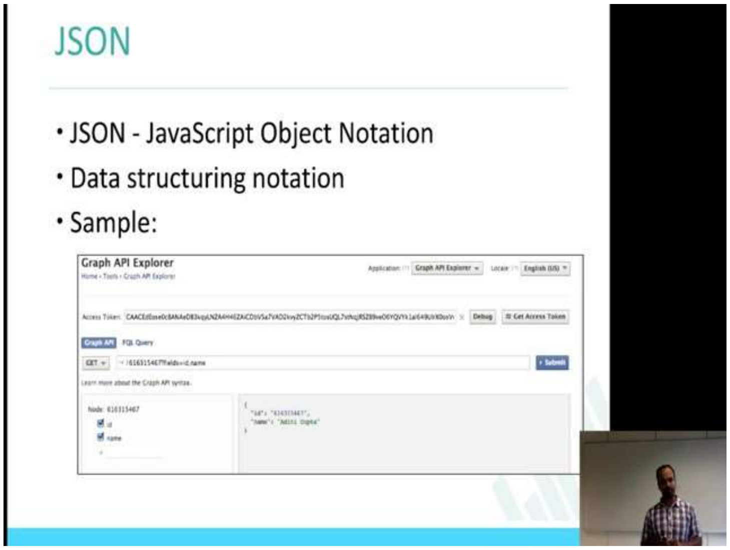


Data format

- API returns data in the following two formats:
  - –JSON
  - –XML

Data format, so particularlytheAPI, when you send the request to Facebook saying that, ‘please give me all the data about friends that PK has’, or, aboutthe date of birth of PK, orabout myfriends’network. So,what it is going togive you back is actually,it is going to give you in some format. One of the formats that it gives you is actually a JSON format, which we will see in brief what this format means and how we actually interpret the data that is coming back from Facebook, or Twitter. XML, which is also a format withsomesocialmediaservices give,ortheJSON,isalsoalittlebitlikeanXML,which is Extended Mark-up Language.

(ReferSlideTime:07:49)



JSON

- JSON - JavaScript Object Notation
- Data structuring notation
- Sample:

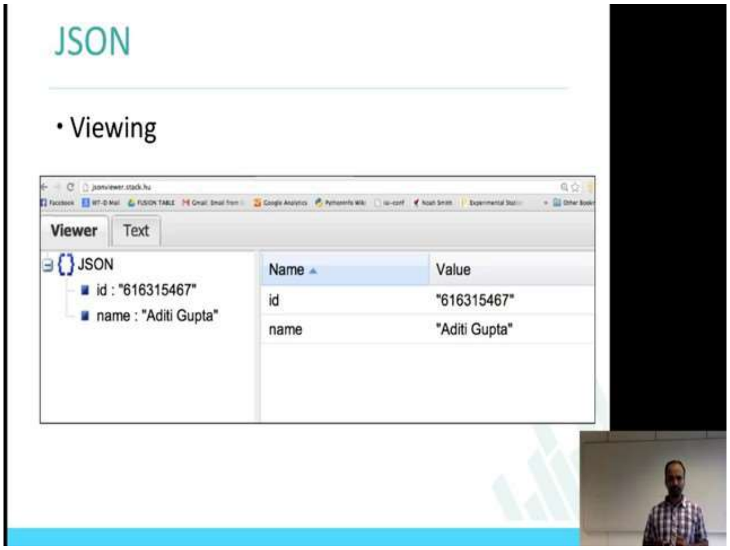
Graph API Explorer
Home › Tools › Graph API Explorer
Application: Graph API Explorer
Locale: English (US)
Access Token:
Debug
Get Access Token
Graph API
FQL Query
GET
/616315467?fields=id,name
Submit
Learn more about the Graph API syntax.
Node: 616315467
id
name

```
{
  "id": "616315467",
  "name": "Aditi Gupta"
}
```

So, here is what a JSON means. JSON means, JavaScript Object Notation, which is a data that you get back from the social media services. So, here is an example that I have in this slide, which just shows you about the JSON object that is returned, when you are askingforidandnameofaparticularuser.So,thisistheGraphAPIExplorer,whichyou will see in the tutorial in more detail but, it is essentially a through by browser you can actually look at the data, look at the JSON objects of the Facebook data of yourself, or whatever the FacebookAPI allows, which we will be able to see through this graphAPI.

So, again, that we emphasize JSON is the JavaScript Object Notation, which is the way that the data is stored in Facebook, data is stored in twitter when you request through the API, for saying, ‘give me this data about PK’, it is returning the data in JSON format. It is basically the format that most social media services use today.

(ReferSlideTime:08:53)



JSON

- Viewing

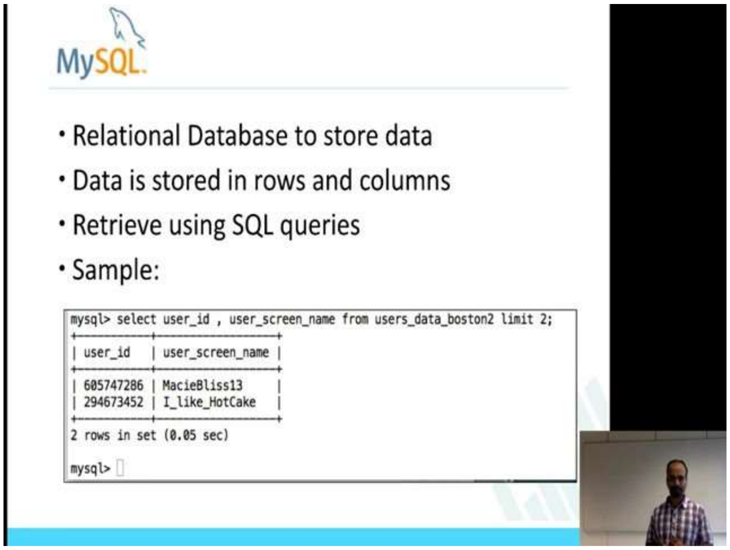
jsonviewer.stack.hu

```
{} JSON
   id : "616315467"
   name : "Aditi Gupta"
```

Name	Value
id	"616315467"
name	"Aditi Gupta"

So, when you take the data from JSON, and when you want to interpret the data that is available in this JSON, data that is coming back from Facebook or Twitter, you can actually use JSON dot viewer dot stack dot hu. This is only for you to see visually,what data is coming back; you can take the data that is coming out of Facebook, copy paste it into this JSON viewer,and you will beable to see, whatthe fields are. When you look at the data that is coming back from Facebook, it is generally a block of data; it is just a lot of data that comes back. So, you can actually take it, and put it into the JSON viewer,to see what are the fields that it is actually giving you. Wego through this slowly,when we do the tutorials.

(ReferSlideTime:09:36)



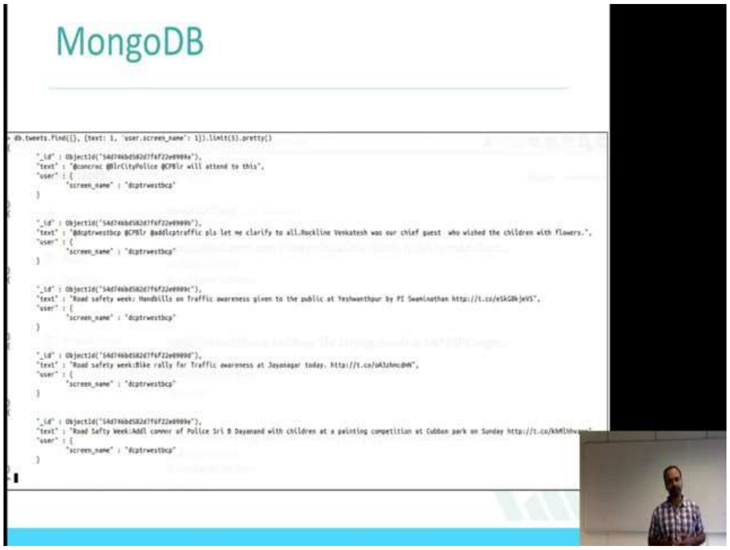
- Relational Database to store data
- Data is stored in rows and columns
- Retrieve using SQL queries
- Sample:

```
mysql> select user_id , user_screen_name from users_data_boston2 limit 2;
+-----------+------------------+
| user_id   | user_screen_name |
+-----------+------------------+
| 605747286 | MacieBliss13     |
| 294673452 | I_like_HotCake   |
+-----------+------------------+
2 rows in set (0.05 sec)

mysql> 
```

And, of course when you collect the data, so first is API which is a way by which you wanttocollectthedata,andthedataiscomingbackinJSON.Whenyoucollectthedata, you have to store it in some format, right. So, the format that majority of the times, the data is stored, is in MySQL. Basically, it is a relational database to store the data, and data is stored in rows and columns, and simple queries, you could use to get the data.

For example, in this case, I am just selecting user id, user screen name from the data that is being collected through Facebook, right. So, that helps meaning, again I am emphasizing that, this is not a course on MySQLitself; we will onlylook at some simple queries on how to look at the data that you have actually stored through the programsthat you have written.

(ReferSlideTime:10:28)

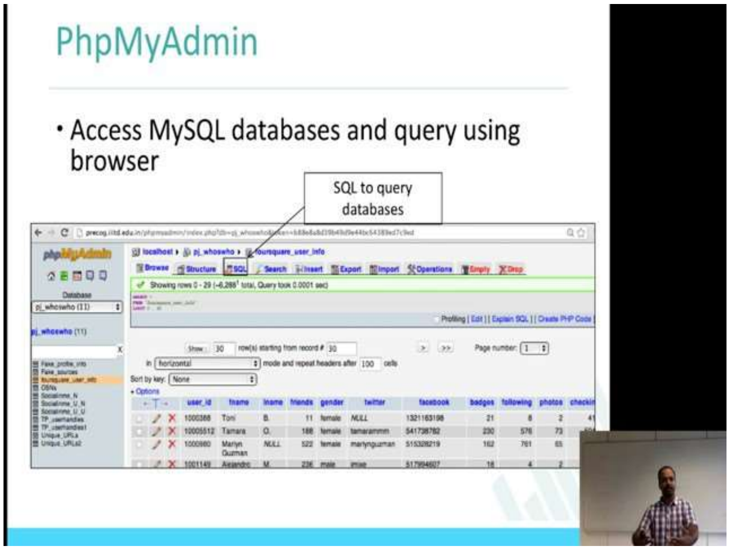


MongoDB

```
> db.tweets.find({}, {text: 1, 'user.screen_name': 1}).limit(5).pretty()
{
        "_id" : ObjectId("54d746bd582d7f6f22e0909a"),
        "text" : "@concrec @BlrCityPolice @CPBlr will attend to this",
        "user" : {
                "screen_name" : "dcptrwestbcp"
        }
}
{
        "_id" : ObjectId("54d746bd582d7f6f22e0909b"),
        "text" : "@dcptrwestbcp @CPBlr @addlcptraffic pls let me clarify to all.Rockline Venkatesh was our chief guest  who wished the children with flowers.",
        "user" : {
                "screen_name" : "dcptrwestbcp"
        }
}
{
        "_id" : ObjectId("54d746bd582d7f6f22e0909c"),
        "text" : "Road safety week: Handbills on Traffic awareness given to the public at Yeshwanthpur by PI Swaminathan http://t.co/eSkGBkjeVS",
        "user" : {
                "screen_name" : "dcptrwestbcp"
        }
}
{
        "_id" : ObjectId("54d746bd582d7f6f22e0909d"),
        "text" : "Road safety week:Bike rally for Traffic awareness at Jayanagar today. http://t.co/oA3zhmcdhN",
        "user" : {
                "screen_name" : "dcptrwestbcp"
        }
}
{
        "_id" : ObjectId("54d746bd582d7f6f22e0909e"),
        "text" : "Road Safty Week:Addl commr of Police Sri B Dayanand with children at a painting competition at Cubbon park on Sunday http://t.co/kbMlhhv
        "user" : {
                "screen_name" : "dcptrwestbcp"
        }
}
> 
```

MongoDB is one of the popular ones, more recently we have started looking at and people are actually using this. So, MongoDB is another way by which the data is stored and the data that is collected from Facebook is actually stored.

(ReferSlideTime:10:47)



PhpMyAdmin

- Access MySQL databases and query using browser

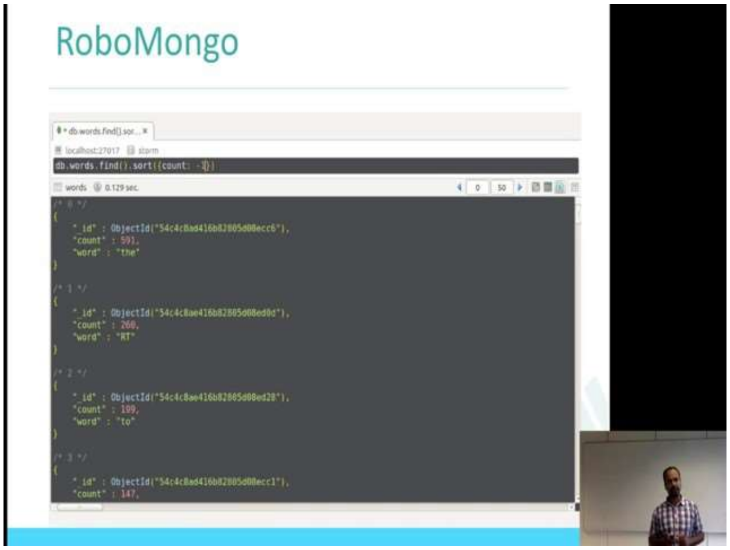
SQL to query databases

✓ Showing rows 0 - 29 (~6,288¹ total, Query took 0.0001 sec)

user_id	fname	lname	friends	gender	twitter	facebook	badges	following	photos	checkin
1000368	Toni	B.	11	female	NULL	1321163198	21	8	2	41
10005512	Tamara	O.	188	female	tamarammm	541738782	230	576	73	[illegible]
1000980	Marlyn Guzman	NULL	522	female	marlynguzman	515328219	162	761	65	[illegible]
1001149	Alejandro	M.	236	male	imixe	517994607	18	4	2	[illegible]

So, again, let me emphasize which is API; then, there is programming language; then, there is MySQL database or MongoDB, which is data is coming through an API, collected and dumpedinto this MySQL or MongoDB. So, now, we also need a way by whichtolookatthedatathatisbeingstored.So,oneofthewaysyoucouldusethis actually phpMyAdmin, which actually allows you to look at the data that you have in your own database.

(ReferSlideTime:11:20)



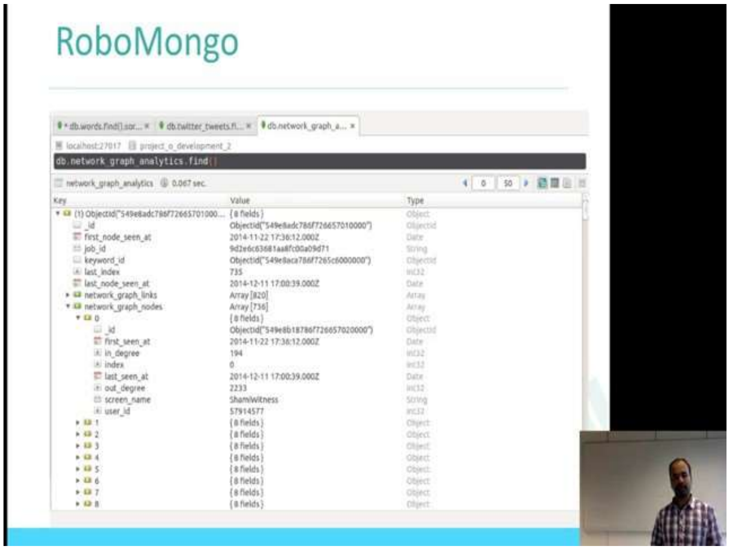
RoboMongo

```
db.words.find().sort({count: -1})
```

```
/* 0 */
{
    "_id" : ObjectId("54c4c8ad416b82805d08ecc6"),
    "count" : 591,
    "word" : "the"
}

/* 1 */
{
    "_id" : ObjectId("54c4c8ae416b82805d08ed0d"),
    "count" : 260,
    "word" : "RT"
}

/* 2 */
{
    "_id" : ObjectId("54c4c8ae416b82805d08ed28"),
    "count" : 199,
    "word" : "to"
}

/* 3 */
{
    "_id" : ObjectId("54c4c8ad416b82805d08ecc1"),
    "count" : 147,
```

So MySQLphpMyAdmin can look at the data from MySQL, and RoboMongo will help you to look at the data from a MongoDB. So, essentially, these are the ways by which you can collect the data, store the data and look at the data that is available with you.

(ReferSlideTime:11:38)



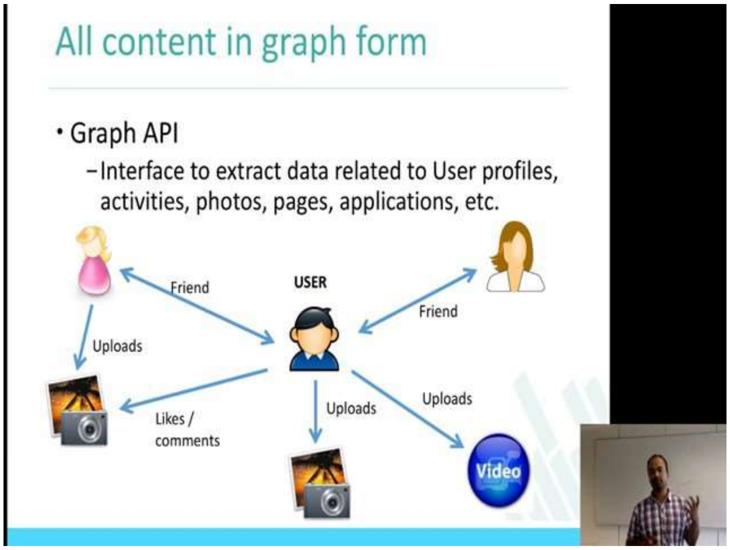
RoboMongo

```
db.network_graph_analytics.find()
```

Key	Value	Type
(1) ObjectId("549e8adc786f72665701000...	{ 8 fields }	Object
_id	ObjectId("549e8adc786f726657010000")	ObjectId
first_node_seen_at	2014-11-22 17:36:12.000Z	Date
job_id	9d2e6c63681aa8fc00a09d71	String
keyword_id	ObjectId("549e8aca786f7265c6000000")	ObjectId
last_index	735	Int32
last_node_seen_at	2014-12-11 17:00:39.000Z	Date
network_graph_links	Array [820]	Array
network_graph_nodes	Array [736]	Array
0	{ 8 fields }	Object
_id	ObjectId("549e8b18786f726657020000")	ObjectId
first_seen_at	2014-11-22 17:36:12.000Z	Date
in_degree	194	Int32
index	0	Int32
last_seen_at	2014-12-11 17:00:39.000Z	Date
out_degree	2233	Int32
screen_name	ShamiWitness	String
user_id	57914577	Int32
1	{ 8 fields }	Object
2	{ 8 fields }	Object
3	{ 8 fields }	Object
4	{ 8 fields }	Object
5	{ 8 fields }	Object
6	{ 8 fields }	Object
7	{ 8 fields }	Object
8	{ 8 fields }	Object

So, this is another view of RoboMongo, which shows you what are the different fields that are available; what data is stored in those fields.

(ReferSlideTime:11:53)

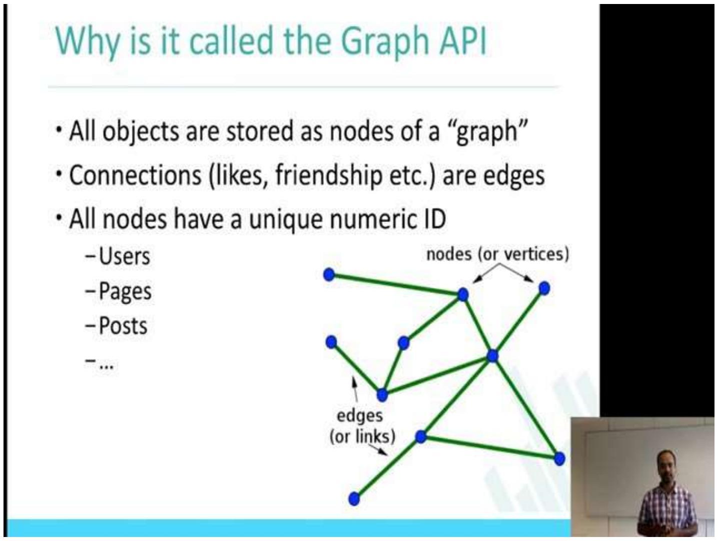


All content in graph form

- Graph API
  - Interface to extract data related to User profiles, activities, photos, pages, applications, etc.

All content on Facebook is actually stored in a graph format; that is, user - the friends that I would have, the pictures that I upload, the videos that I upload, and the status updates that I do, everything is actually a node in the graph. And, every interaction, which is basically like the comments, likes and things like that, becomes edges in this graph. Facebook actually stores all interactions, of all data that they have within the graph format; that is why the API that they have is also called as a graphAPI.

(ReferSlideTime:12:27)

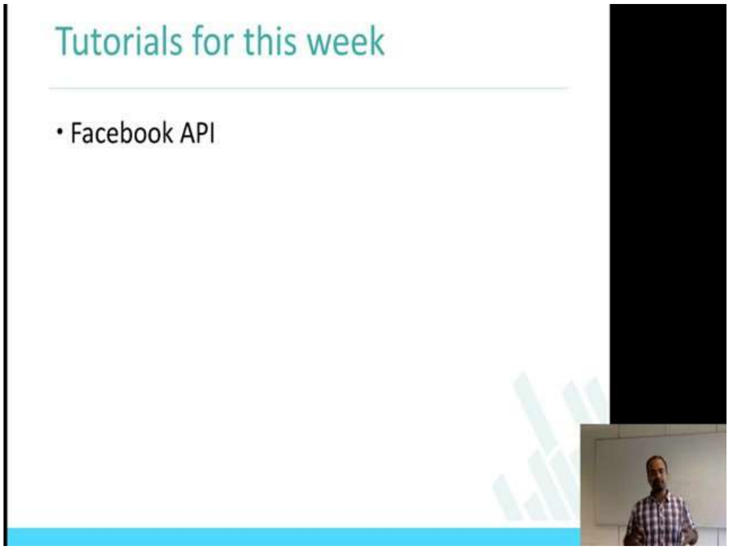


Why is it called the Graph API

- All objects are stored as nodes of a "graph"
- Connections (likes, friendship etc.) are edges
- All nodes have a unique numeric ID
  - Users
  - Pages
  - Posts
  - ...

Here is the another view ofthe same message, which is,all objects are stored as nodes in the graph; connections like friends, friendships, likes are edges and all nodes have a uniquenumericID,whichisusers,pagesandposts.And,wewillbetalkingmostlyabout users; we shall later talk about also, pages, which is one of the ways by which contentcan be generated on Facebook.

(ReferSlideTime:12:52)



Tutorials for this week

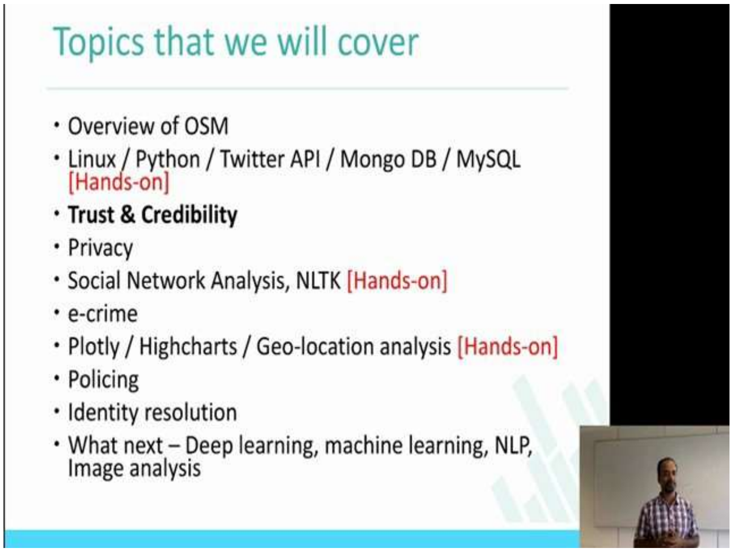
- Facebook API

In tutorials this week, you will actually look at in detail about what a Facebook API is, how do you actually create the secret key,what kind of authentication that you will have to provide Facebook, in terms of collecting data, what data can be collected and things like that.

TrustandCredibilityonOSM

So now, Let us look at week 2.2 of Privacy and Security in Online Social Media course on NPTEL.

(ReferSlideTime:00:12)

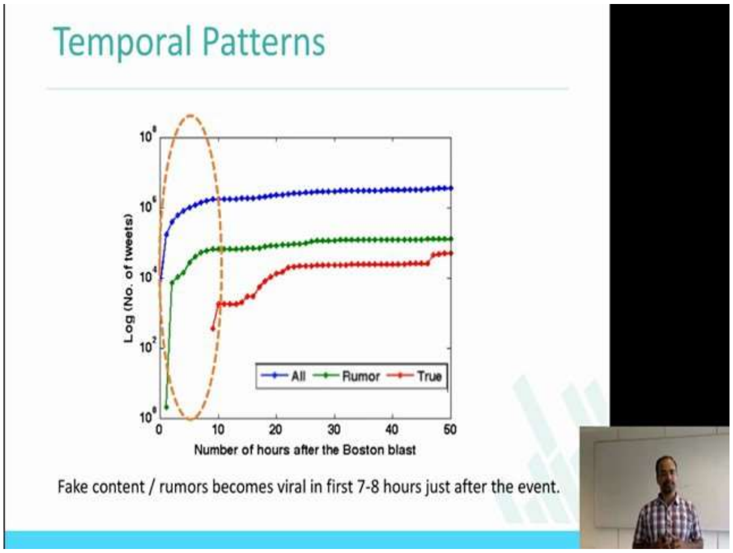


Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- **Trust & Credibility**
- Privacy
- Social Network Analysis, NLTK [Hands-on]
- e-crime
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Policing
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

In this part, what I am planning to cover is actually getting deeper into the topic called Trust and Credibility.This is the slide that I will probably tweet over the course multiple times just to tell you where we are and where we are going. We finished about overview of online social media and we in the lab sessions, we have done Linux and Python and you will actually get to see a little bit about Facebook, Twitter API's and now we will actually look at the topic trust and credibility in detail, and later we cover some topics like privacy in social network analysis, e-crime, policing, using online social media and also identityresolution.

(ReferSlideTime:00:53)



Temporal Patterns

Fake content / rumors becomes viral in first 7-8 hours just after the event.

Let us take a look at this graph. In this graph just to read the graph at the x-axis, this is the number of hours after the Boston blast, the data basically from the Boston blast that happened in the US. The x-axis is number of hours after the blast and the y-axis is log of tweetsbasicallywhatdoesitshow ,itshowsthatatanygivenpointintimewhichisafter the Boston blast, how much of tweets is being uploaded on Twitter.

So, there are 3 different colors in this graph which is blue, green and red. So, the one which is in the red is actually legitimate information, which you can call as the true information which is posted on Twitter. Green which is the rumor which is information that is not legitimate or untrustworthy, the non-credible content that was being postedlike the example, like the crocodile example that I mentioned in earlier lecture and the blue one which is the sum of the rumor on the true information.

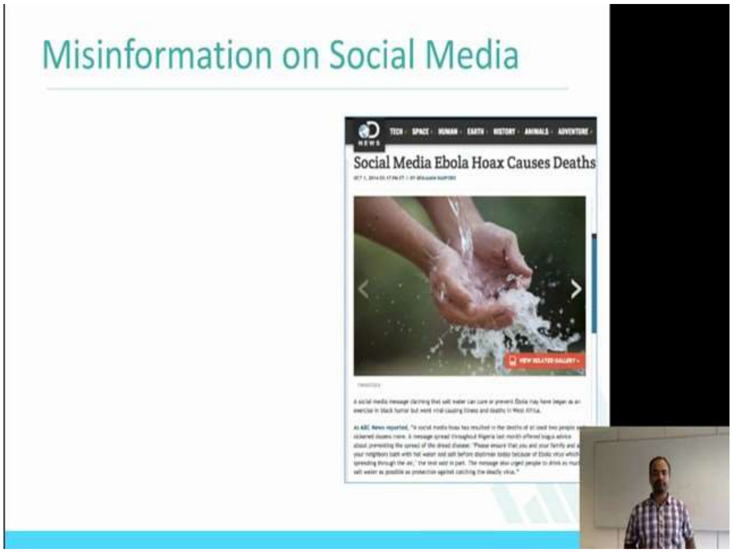
It clearly shows the messages, some implications from the step one that is the true information is actually coming later; it is taking much more time than the rumors that started. In this example, there was in this event Boston blast, there was actually multiple post which were related, which to this event was not actually legitimate, for example,one postwhich said that8year old kid was actuallypartof this Boston blastwhich when therewasnokidinvolvedintheBostonblast.Therewasalsoanothertweetwhichsaid that please RT this tweet and we will actually pay 1 dollar to Boston marathon league which also was not true.

There are many examples like this and these tweets got retweeted for more than thousands of times when Boston blast happened. This actually shows that there is multiple things one can actually look at one; how do you actually reduce the time in which the true information is coming, which is from currently it is about 9 hours or so, how can you actually get this true information come on to the social network as early as possible.

Theother solutionthat you could alsothink of this,how can you quicklyreduce thefalse information that is going on social media from, to reduce, for example in this case the green one is actually peaking in couple of hours and then its actually higher than the true information, how can you actually quickly reduce the effect of or the flow or the information propagation of this particular rumor on social networks.

So, those are the two things that you could actually do, atleast do to reduce the effect of rumorsyouneedtoactuallyunderstand,whattherumoris?Howcanweactuallyidentify these rumors on a Twitter that is what we basically look at in the section of this course. Which is to identify ways by which I will look at the tweets and identify whether theyare legitimate or not.

(ReferSlideTime:04:04)



Misinformation on Social Media

NEWS
TECH · SPACE · HUMAN · EARTH · HISTORY · ANIMALS · ADVENTURE ·
VIEW RELATED GALLERY »

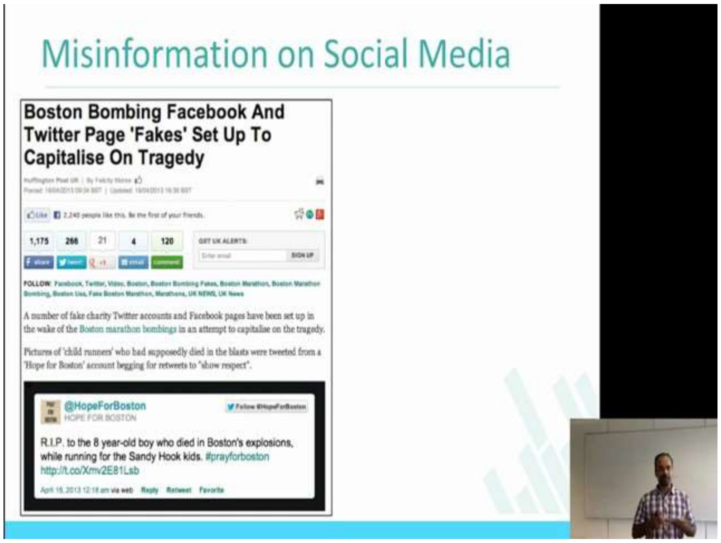
Social Media Ebola Hoax Causes Deaths

A social media message claiming that salt water can cure or prevent Ebola may have began as an exercise in black humor but went viral causing illness and deaths in West Africa.

As ABC News reported, "A social media hoax has resulted in the deaths of at least two people and sickened dozens more. A message spread throughout Nigeria last month offered bogus advice about preventing the spread of the dread disease: "Please ensure that you and your family and all your neighbors bath with hot water and salt before daybreak today because of Ebola virus which is spreading through the air," the text said in part. The message also urged people to drink as much salt water as possible as protection against catching the deadly virus."

Sonow,wewilllookatmisinformationonsocialmedia,whicharesomeexamplesofthe misinformation that was on Twitter. Here is one example, which actually took a lot of effect in social media. When Ebola was going on there were a lot of messages saying Ebolahoax which causes deaths and therewas also discussion on the post abouthow salt water could be used to actually reduce Ebola and things like that.

(ReferSlideTime:04:36)



Misinformation on Social Media

Boston Bombing Facebook And Twitter Page 'Fakes' Set Up To Capitalise On Tragedy

Huffington Post UK | By Felicity Morse

Posted: 16/04/2013 09:34 BST | Updated: 16/04/2013 16:38 BST

FOLLOW: Facebook, Twitter, Video, Boston, Boston Bombing Fakes, Boston Marathon, Boston Marathon Bombing, Boston Usa, Fake Boston Marathon, Marathons, UK NEWS, UK News

A number of fake charity Twitter accounts and Facebook pages have been set up in the wake of the Boston marathon bombings in an attempt to capitalise on the tragedy.

Pictures of 'child runners' who had supposedly died in the blasts were tweeted from a 'Hope for Boston' account begging for retweets to "show respect".

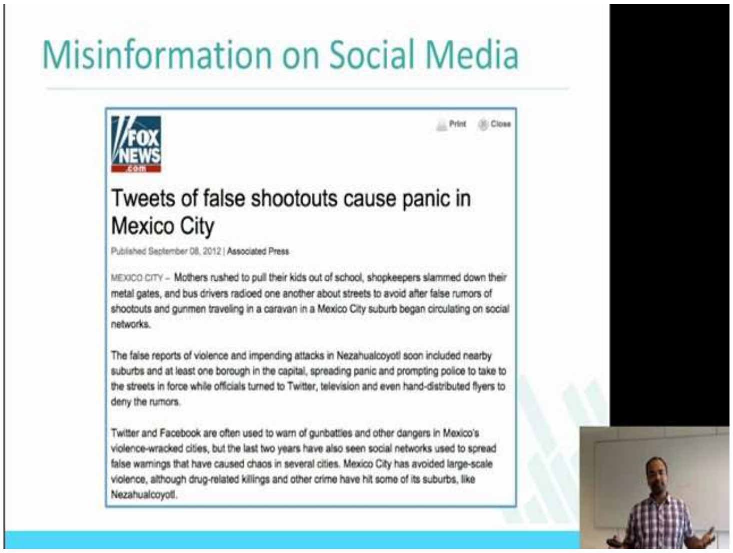
@HopeForBoston
HOPE FOR BOSTON

R.I.P. to the 8 year-old boy who died in Boston's explosions, while running for the Sandy Hook kids. #prayforboston http://t.co/Xmv2E81Lsb

April 16, 2013 12:18 am via web Reply Retweet Favorite

Boston marathon, hereis a specifictweet thatIjustnow mentioned, which is R.I.Pto the 8 year boy, who died in Boston explosions while running for the Sandy Hook kids and that was not true at all.

(ReferSlideTime:04:50)



Misinformation on Social Media

Tweets of false shootouts cause panic in Mexico City

Published September 08, 2012 | Associated Press

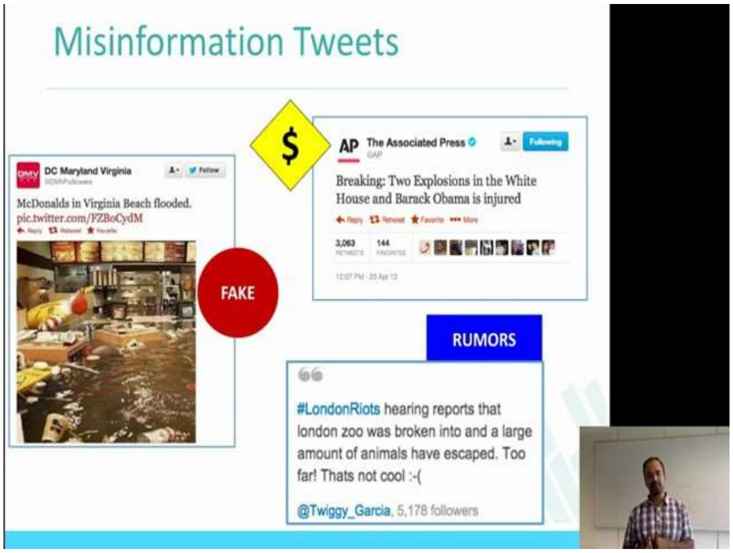
MEXICO CITY – Mothers rushed to pull their kids out of school, shopkeepers slammed down their metal gates, and bus drivers radioed one another about streets to avoid after false rumors of shootouts and gunmen traveling in a caravan in a Mexico City suburb began circulating on social networks.

The false reports of violence and impending attacks in Nezahualcoyotl soon included nearby suburbs and at least one borough in the capital, spreading panic and prompting police to take to the streets in force while officials turned to Twitter, television and even hand-distributed flyers to deny the rumors.

Twitter and Facebook are often used to warn of gunbattles and other dangers in Mexico's violence-wracked cities, but the last two years have also seen social networks used to spread false warnings that have caused chaos in several cities. Mexico City has avoided large-scale violence, although drug-related killings and other crime have hit some of its suburbs, like Nezahualcoyotl.

There's been many, many examples I am just going to give you some examples as motivationforthissection ofthiscourse,tweetsoffalseshootoutscausepanicinMexico city, this is one of the incident.

(ReferSlideTime:05:03)



Misinformation Tweets

DC Maryland Virginia

McDonalds in Virginia Beach flooded.
pic.twitter.com/FZBoCydM

FAKE

$

The Associated Press

Breaking: Two Explosions in the White House and Barack Obama is injured

3,063 RETWEETS 144 FAVORITES

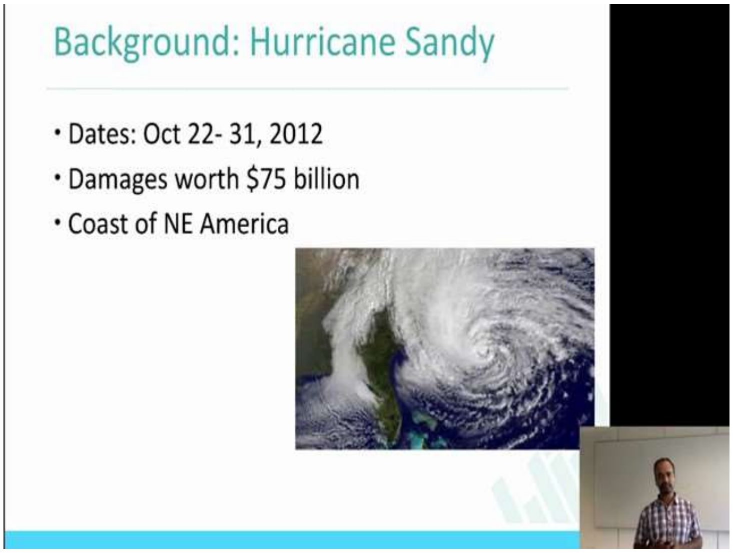
RUMORS

#LondonRiots hearing reports that london zoo was broken into and a large amount of animals have escaped. Too far! Thats not cool :-(

@Twiggy_Garcia, 5,178 followers

And some tweets, some images that I actually talked about even in my first lecture,which is McDonalds in Virginia Beach flooded, the image of the left where the image was actually the real image, but it was not taken during Virginia Beach flood, but it was actually taken many years before and they associated first and the past also. Here is a rumor in the right hand bottom, which is London riots, here it reports the London zoo was broken into and large amount of animals have escaped that is again a rumor. There have been many rumors like this in many events that have happened in the past.

(ReferSlideTime:05:37)



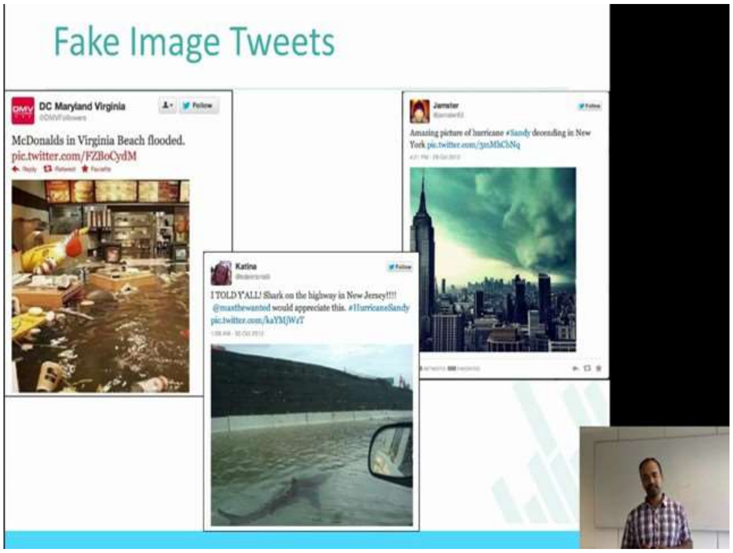
Background: Hurricane Sandy

- Dates: Oct 22- 31, 2012
- Damages worth $75 billion
- Coast of NE America

What we are going to do is we are going to actually take one specific example. We will actually do multiple examples over the course, over the entire course. Take up this event and look at the actuallythe topic ofmisinformation in this case and other topics in future to study how we can actually analyze this data and make some inferences out of it in the context of trust and credibility. We will do the similar way in the future also, for any topicwe'lltakeanevent,wewill takesomedatathathasbeencollecteddosome analysis on the content and make some inferences of the topic that we are interested in.

InthiscasewearegoingtotakeHurricane Sandy.Hurricane Sandyhappened inOctober 22-31,2012andtheideaforusinganeventisthatyouwillabletorelatetoitandmostof the times analysis is done looking at the particular event, for example, now many people are interested in studying elections in the US and I know there are people also interested in studying elections in India when it happens. So, the damages that were totally worth for Hurricane Sandy was about 75 billion and the Hurricane Sandy basically in the north eastern part of the US.

(ReferSlideTime:06:53)



Fake Image Tweets

DC Maryland Virginia

McDonalds in Virginia Beach flooded.
pic.twitter.com/FZBoCydM

Katina

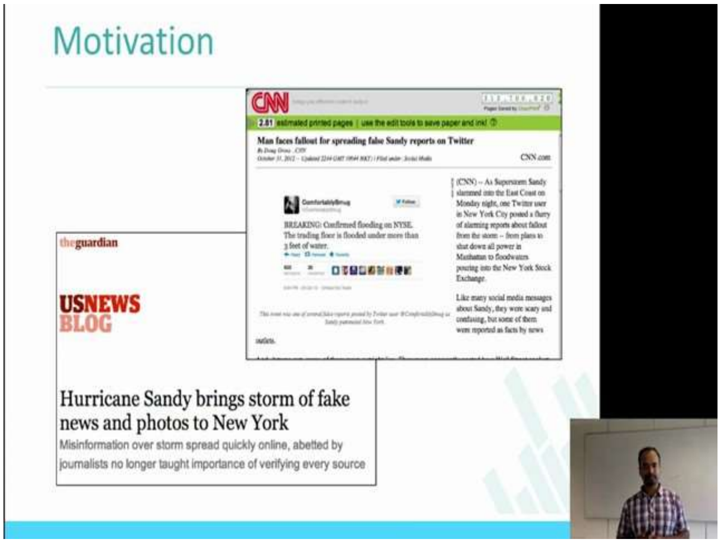
I TOLD Y'ALL! Shark on the highway in New Jersey!!!!
@maxthewanted would appreciate this. #HurricaneSandy
pic.twitter.com/kaYMjW2T

Jamster

Amazing picture of hurricane #Sandy decending in New York pic.twitter.com/3mMhChNq

There has been many, many fake images that was floating during the Hurricane Sandy. The one McDonalds as I said before and one middle has shark in the water and people were actually,there was panic among a topic and the right hand topic which is also from Hurricane Sandyand which there was an image froma picture and it was actuallyposted on Twitter saying that is how it is looking in the US now.

(ReferSlideTime:07:24)



Motivation

CNN

2.81 estimated printed pages | use the edit tools to save paper and ink!

Man faces fallout for spreading false Sandy reports on Twitter

CNN.com

ComfortablySmug

BREAKING: Confirmed flooding on NYSE. The trading floor is flooded under more than 3 feet of water.

(CNN) -- As Superstorm Sandy slammed into the East Coast on Monday night, one Twitter user in New York City posted a flurry of alarming reports about fallout from the storm -- from plans to shut down all power in Manhattan to floodwaters pouring into the New York Stock Exchange.

Like many social media messages about Sandy, they were scary and confusing, but some of them were reported as facts by news outlets.

theguardian

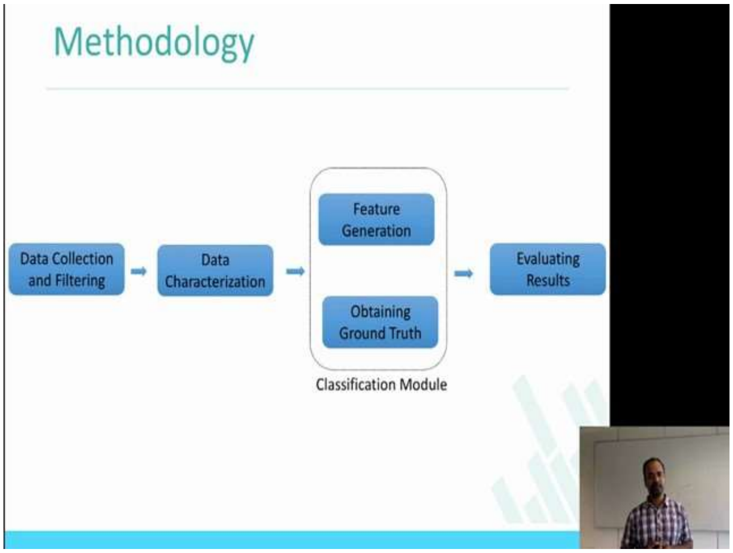
USNEWS BLOG

Hurricane Sandy brings storm of fake news and photos to New York

Misinformation over storm spread quickly online, abetted by journalists no longer taught importance of verifying every source

So,particularlyforHurricaneSandy,ifyou seethatthereis,wealsoknoweffectofthese fake information that was going on Twitter.‘Hurricane Sandy brings storm of fake news and photos to New York’, ‘Man faces fallout for spreading a fake Sandy reports on Twitter’. These are some incidence which is happening around the event HurricaneSandy.

(ReferSlideTime:07:46)

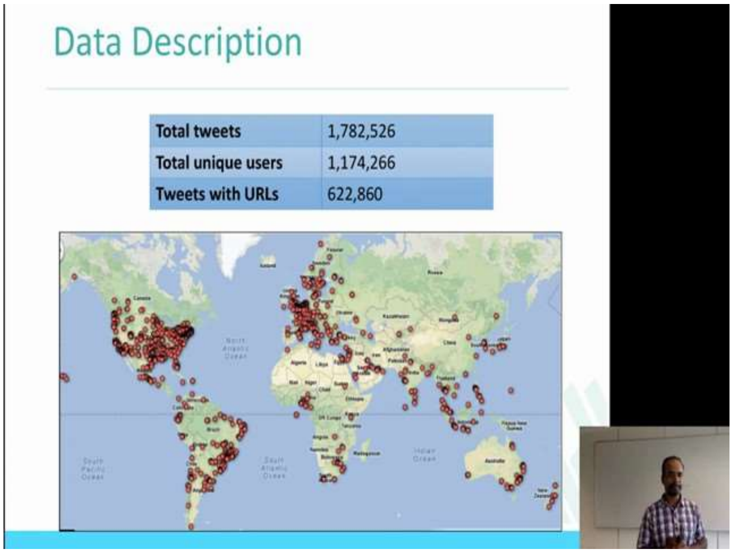


Methodology

So, what we are going to look at is again some methodology things that I will be talking about in this course is generic. I will try to actually emphasize on this methodology. So, that we would takes this methodology and apply it in any scenario that we wereinterested in, sometimes even in the homework and assignments that we will do as partof this course which is first we start collecting data from Twitter about the Hurricane Sandyand then some kind of data characterization which is understanding, how much of data is come? What data is come and things like? That future generation obtaining the ground truth and then evaluating the results.

Thisisaveryhighlevelprobably30-40000fthighlevelviewofwhatthe majorityofthe analysis on social media data would be going on. We ourselves in the course will look at different levels of view of this slide, which is later in that course we will also look at somethingmoredetailintermsofactuallythisthewholeprocess.Thesimpleprocessis collect some data and do some characterization, understand some features, use those features to create a model, use that modelto actually study the larger amount of data and evaluate the results that is the general. If you have taken any machine learning or aninformation retrieval course that is the kind of a simple process that people follow.

(ReferSlideTime:09:06)



Data Description

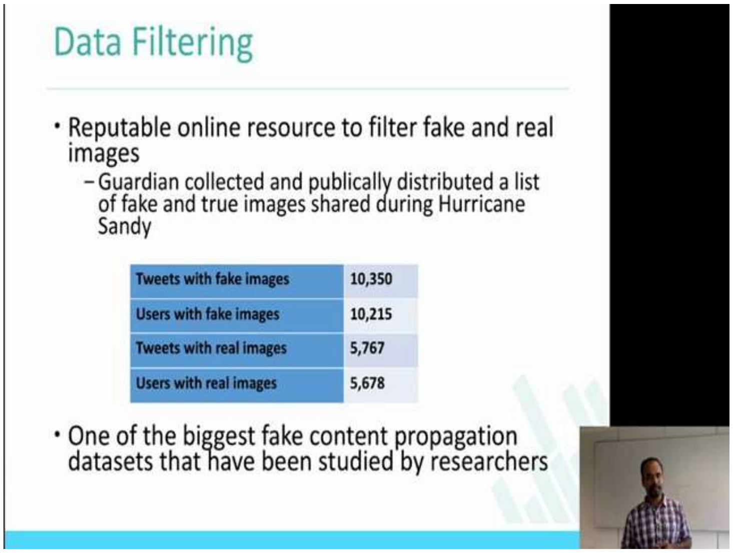
Total tweets	1,782,526
Total unique users	1,174,266
Tweets with URLs	622,860

So, the data we are talking about and this is one the most exciting thing that I feel about thestudying and researching inthearea ofonlinesocialmedia, isthesizeofthedatathat we are talking about. In this case, in the Hurricane Sandy thing we are talking about 1,782,526 tweets that were collected, while Hurricane Sandy happened. Total unique users were about 1 million users and tweets with URLs was about 622,000. So, thatgives you a sense of how much of data was collected in terms of the hurricane sandy.

So, again please keep this as a template when you are doing some kind of analysis of events. These kinds of attempt, these kinds of analysis, and these kinds of data description will help actually to look at the data to understand what the data is and in other sense it will also help for somebody who's going to do it again. These kinds of the sameanalysisorsomething similar theywould beable toactuallytakeawaysome points from the data that you would describe.

Also, in this case the map in the bottom also shows that where the tweet is come from of course, these tweets have geotagged the information therefore we are able to actually mark it on the whole map on where the tweet has come from.

(ReferSlideTime:10:31)



Data Filtering

- Reputable online resource to filter fake and real images
  - Guardian collected and publically distributed a list of fake and true images shared during Hurricane Sandy

Tweets with fake images	10,350
Users with fake images	10,215
Tweets with real images	5,767
Users with real images	5,678

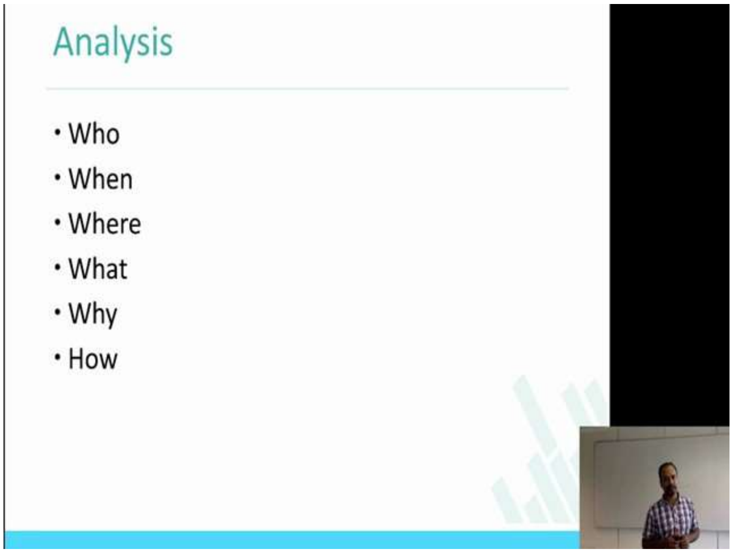
- One of the biggest fake content propagation datasets that have been studied by researchers

Of course the big question is that how will you get the ground truth because now, if we were to look at this tweets and say that which are fake, which are legitimate you need to know what the fake tweets are. So, the multiple ways that people have tried, multiple techniques that people try which is, we look at it some in the course in some probably I will just mention it as we look at the slides.

So, in this particular Hurricane Sandy analysis that was done in the way it was done was the Guardian, which is actually a media house they collected actually fake information, thenmanuallyannotatedthattheseimages arefakebecausetheyarearepositoryoflotof content that gets generated on social media, they were able to actually annotate and produce the dataset which is, which say that in Hurricane Sandy these are the fake post right. So, the reputableonlineresource tofilterfake andreal images, ‘Guardiancollected and publicly distributed a list of fake and true images’ what did they distribute, tweets with fake images 10,350 tweets; users with fake images 10,215; tweets with real images and users with real images.

So,usingthe reputableguardiandata,weactuallylooked atthedatathatwas collected in this and then form how many tweets have these posts, how many images that was posted that were actually fake and how many unique users and things like that, that is what that was done.

(ReferSlideTime:12:06)

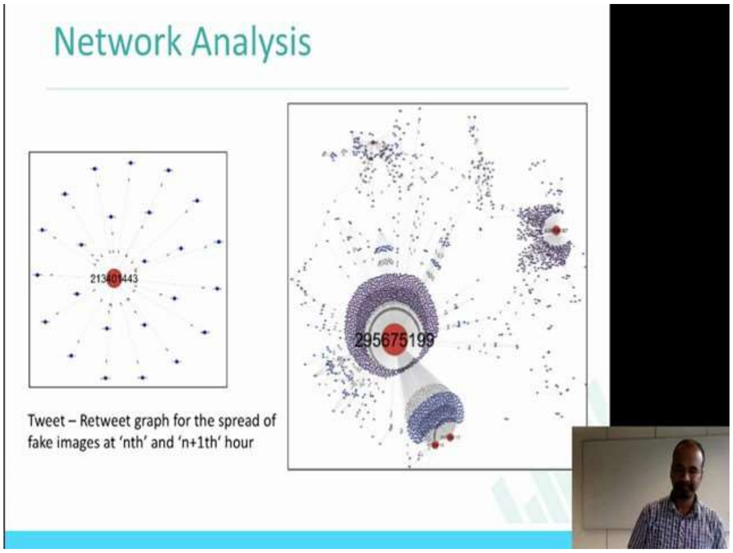


Analysis

- Who
- When
- Where
- What
- Why
- How

So, when you do these kinds of analysis when generally look at online social media analysis, it is always best to look at analysis like this; who, when, where, what, why and how.Thesekindsofanalysiswill actuallyhelpyoutoanswersomeinterestingquestions; who posted it? When did they post? Where did they post from? What did they post about? Why did they post and how did they post? So, why and how was slightly trickier here, it is hard to get, why did the person post a rumor? It is hard to tell unless the user, unless the person who posted and itself actually confesses, how do you, how did the course is probably is I mean probably possible to get which is to look at, what they why did they use how did they post, into the social media.

(ReferSlideTime:12:56)


Network Analysis

Tweet – Retweet graph for the spread of fake images at 'nth' and 'n+1th' hour

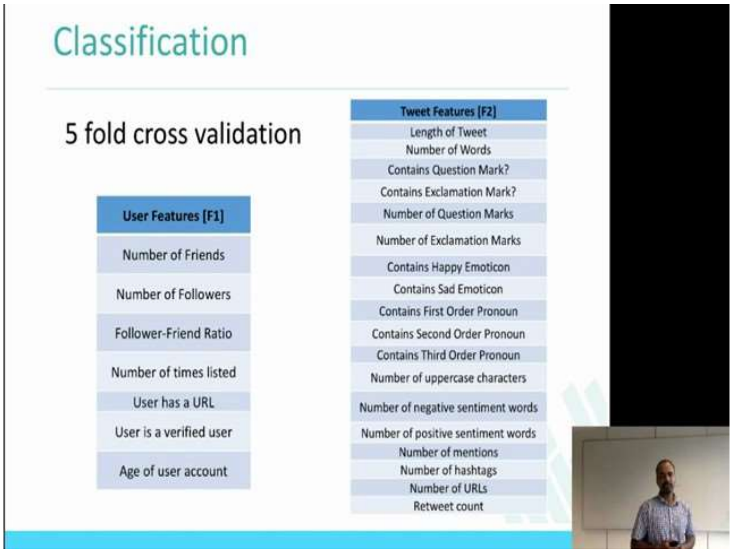
Now, let us look at more specifically the analysis on who, when, where and what and how and what. So,in this graph what I am showing you is a network analysis which is to show you, who is the person who posted the tweet and how the information is getting diffused. The one of the left is the user who posted content on this particular event and when the user actually post this content obviously, in Twitter there is going to be this retweets, favorites and mentions of the user. So, in this case the blue the red dot is theone of a particular user who posted this content and the blue dots are the ones where the users are actually retweeted this content. So, if we just look at this the users' content is actually spreading among the other users in the network.

So,theoneintheleftis givingyouthenthhour,theonetherightisgivingyouthenplus 1th hour what is the difference here. So, in the first one there is only one user where as content is getting spread, whereas in the one on the right if you see the post is actually diffused so heavily in the network, within one hour. But there is also other observations that you can actually have, if you look at the number in the left of the user id. So, this basicallytheuser id is the onethrough which wecan actuallycollectdatafromTwitter it is a unique for a every particular user. So, the number in the left and then one that numbers that are prominent in the right are actually going different.

What does this show? This shows that the content, that some body starts, let us take ifPK starts content and his content gets diffused in the network, he may or may not be the one who is actually more popular at after given point in time. This basically shows that the information is, we can actually draw multiple inferences from this analysis which is, who is posting the content? How the information is getting diffused, for example, if you say 3 plus see, on the left last 3 digits are 443 which is the prominent user, whereas ifyou look at in the right it is 199 the user which is in the center of the network.

So,thatisoneimportantanalysisthatyoucandoinference,thatyoucandrawfromthese kind of analysis, this is called network analysis. We later in the course will actually see some tools where you can actually draw these graphs with the data that you collect from Twitter or a facebook. So, in this graph you can also see in one of the user on the right hand top corner which also has more number retweets. There are some users which is bottom ofthe graph, onthe right which is, whoare also more the tweets are getting more retweeted.

(ReferSlideTime:15:43)



Classification

5 fold cross validation

User Features [F1]
Number of Friends
Number of Followers
Follower-Friend Ratio
Number of times listed
User has a URL
User is a verified user
Age of user account

Tweet Features [F2]
Length of Tweet
Number of Words
Contains Question Mark?
Contains Exclamation Mark?
Number of Question Marks
Number of Exclamation Marks
Contains Happy Emoticon
Contains Sad Emoticon
Contains First Order Pronoun
Contains Second Order Pronoun
Contains Third Order Pronoun
Number of uppercase characters
Number of negative sentiment words
Number of positive sentiment words
Number of mentions
Number of hashtags
Number of URLs
Retweet count

So, now let us look at a different kind of analysis from the data that we collect from Twitter,oneoftheproblemsthatwecanactuallysolvethis istoactuallyclassifywhether thepostispostthatisgiventous,butthatcomesTwitterisactuallyfakeorreal.So,in that context, we will actually apply a technique called classification given that this is not a machinelearning or an information retrival class. Wewillnot go into detail about what the classification is? What are different techniques? I will just only look at techniques that we applied with the data that we collect from Twitter.

So, if you look at this the different kinds of features that we actually get from the post that we get from Twitter or user features and tweet features. There are actually three kinds of content that you can actually look at from Twitter, one is the user profile which is, who am I in my case the faculty at IIIT, Delhi got my PhD from Carnegie Mellon university and things like that. Second, the people who I am connected with, that is my network mynetwork would befacultythatare around theworld, students thatarearound the world who are doing cyber security, people who are, people who studied on social media and things like that is my network.

Third is actually the content that I post itself what I am talking about I am talking about my students I am talking about PSOSM course, I am talking about some random things on social media right. So, those of the three broad categories of content that you can actuallydrawfromthesocialmedia datawhichisuserprofile,thecontentthatsomebody post and the third one being the network that somebody is connected to.

So,inthiscaseweareactuallylookingattheuserfeatures whichismorelikefeel,which ismoreliketheprofiletweetfeatures.Tweetfeaturesarefromthetweetitself;letmejust go through few of them that I have listed in this slide, which is in terms of user features number of friends. Somebody has number of followers, follower-friend ratio is also one of the important things that we can actually use while making the decision on whether this user isactuallylegitimateorfake, for example, if somebodyis verypopular user,the number of followings that they would have is actually much lower that is the people that PK will follow is actually lower than the number of people who would actually follow PK.

So, that ratio can be actually used to make a judgment on whether the network, whether the user is actually legitimate or not, number of times listed list is another feature in TwitterwhereletustakeifIwanttocreatealistofallthestudentswhoaretaking

PSOSM course I will create a list as PSOSM course on NPTEL and I will actually addall the Twitter users on to this list.

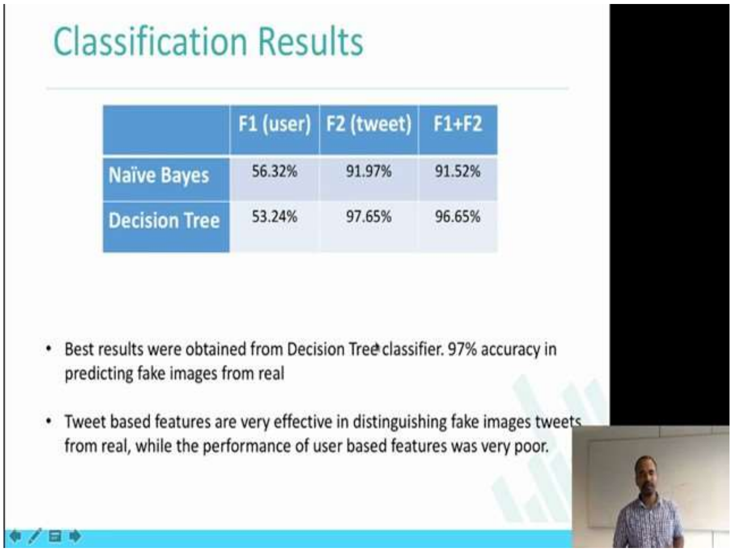
So,thatiscalledalistandparticularlythislisthasbeenusedindifferentinterestingways ifIweretofindtheexpertsaroundtheworldonparticulartopiclistcouldbeactuallyone of the good ways to find out which is if I were to create a list or somebody else creates a list on cyber security, and if they add PK onto it there is high probability that the person believes the PK is actually an expert and therefore, he is adding him or her into that list.

So, number of times listed is one feature that you can use, please go through the Twitter network, do play around with the list and other features that I am taking about, user has the url which is in my profile I will actually say that I am faculty at IIIT, Delhi and this URL called precog dot iiitd dot edu dot in, that URL is there how could the you can actually use that feature to predict user is a verified himself, verified user is another important feature that you could use because of their total number of users on Twitter there is only few a hundreds and thousands of users, who are actually verified, verified takes some process and you have to be you are to have a larger followers and things like that.

So, verified user can be a good feature to decide whether the user is legitimate or a malicioususer.Ageoftheuseraccount, andthishasbeenafeaturethatpeoplehaveused in traditional internet security methodologies, where they have actually used age of aweb site, age of the domain registration to actually find out whether the domain is legitimate or not. It is same feature which is PK created an account 5 years back it is more legitimate and there is a PK account which means or there is Amitabh Bachchan account or Rajinikanth account which is created recently which may not be actually legitimateaccount.So,thatistheintuitionbehindusingageofauseraccount.So,nextis tweet features let us just look at little bit about tweet features itself, in tweet features lengthofthetweetisagoodinformationthatyoucanactuallyusetofindoutwhetherthe post is legitimate or not.

So, these are the features that you can use in general from many different analysis I am onlyusingitfortheproblemoftrustandcredibilitythatwearetalkingabout.Also length of the tweet number of words in the tweet, contains question mark, contains exclamation marks, number of question marks, number of exclamation marks, contains happy emoticon, contains sad emoticon, and things like that. So, this these are the different features that you can actually draw from tweets and the features that I told earlier which were actually user features. So, five fold cross validation is a technique which is used to make sure that the confidence on the classification accuracy that we are building is higher that is the reason why we use actually five fold cross validation and there are many other techniques, I am not going to into details of different other techniques that are available.

(ReferSlideTime:21:22)



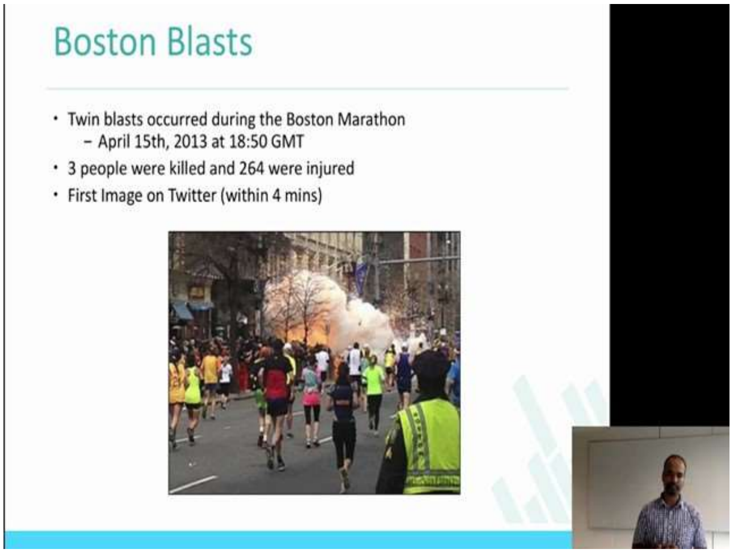
Classification Results

	F1 (user)	F2 (tweet)	F1+F2
Naïve Bayes	56.32%	91.97%	91.52%
Decision Tree	53.24%	97.65%	96.65%

- Best results were obtained from Decision Tree classifier. 97% accuracy in predicting fake images from real
- Tweet based features are very effective in distinguishing fake images tweets from real, while the performance of user based features was very poor.

Now, let us look at the results from the classification that we did. So, as I saidF1 is a user feature F2 is a tweet features, and in classification techniques we can actually useF1 F2 separately and also create a features of F1 plus F2 the two techniques that were applied one is Naive bayes, which I actually usesbayes theorem to find out whether particular post is legitimate or fake and which features actually influence a lot in making the decision. That is also another technique which is a graph based technique which is decision tree, where all the outcomes and the probabilities are actually layed down on as in the form of a graph and it is a very popular machine learning technique which is appliedtomakedecisions.Andthisparticularcasedecisiontreeactuallyseemtoawork better the people while using the tweet features, while the efficiency was about 97.65 percent, in predicting whether the post is fake or real. The tweet features I have choosed seems to a you have played well in both a naive bayes and decision tree where as use the features did not that play that much well in making the decision.

(ReferSlideTime:22:32)

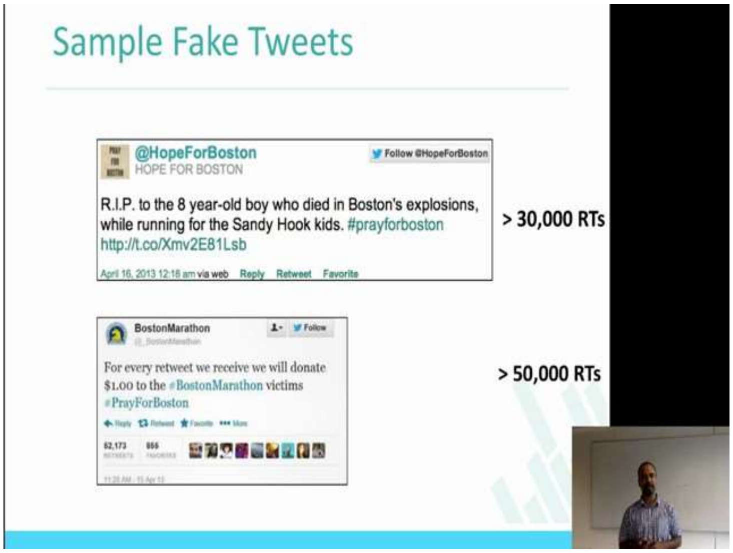


Boston Blasts

- Twin blasts occurred during the Boston Marathon
  - April 15th, 2013 at 18:50 GMT
- 3 people were killed and 264 were injured
- First Image on Twitter (within 4 mins)

Now, let us look at an event Boston blast again, we will use this technique could be looking at events through the events I will actually inject a lot of techniques that we will actually study in this class, and terminologies also that we will see. It is a twin blast that happened in 2013 in April and 3 people were killed and 264 were injured, first imagethat come on Twitter was within 4 minutes. It is basically a Boston marathon that was going on and the blast of the finish line was the event that happened.

(ReferSlideTime:23:01)



Sample Fake Tweets

@HopeForBoston
HOPE FOR BOSTON

R.I.P. to the 8 year-old boy who died in Boston's explosions, while running for the Sandy Hook kids. #prayforboston http://t.co/Xmv2E81Lsb

April 16, 2013 12:18 am via web Reply Retweet Favorite

> 30,000 RTs

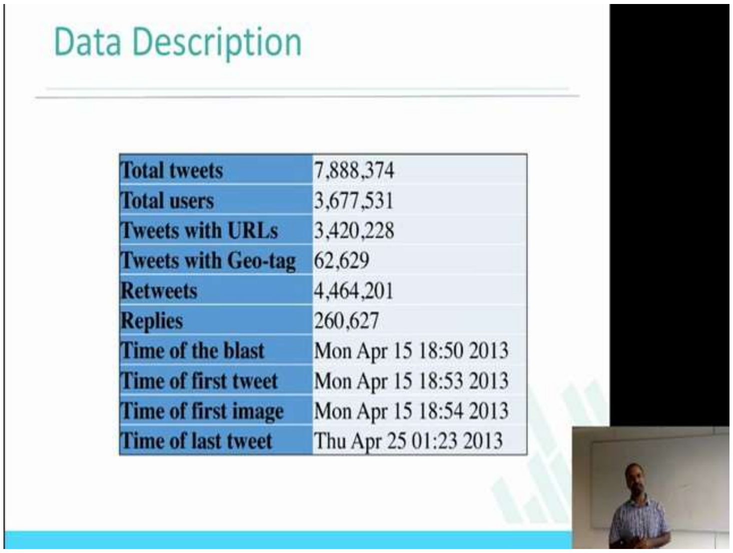
BostonMarathon

For every retweet we receive we will donate $1.00 to the #BostonMarathon victims #PrayForBoston

> 50,000 RTs

And there were actually multiple fake tweets here I am just showing you two popular fake tweets that were actually floating around, the first one I have showed this tweet in thepastalso ‘R.I.P.to the 8year old boywho died in Boston’sexplosions,whilerunning for the Sandy Hook kids’. There was no kid who has participated in the marathon and then other post the at the bottom you see, for every retweet we will donate one dollar to the Boston marathon victims, and it is posted by an account called underscore Boston marathon something that you want to keep in mind which was not a legitimate account and this post was retweeted for about 50,000 times and these are the two popular tweets that were floating around during the event which were fake

(ReferSlideTime:23:47)



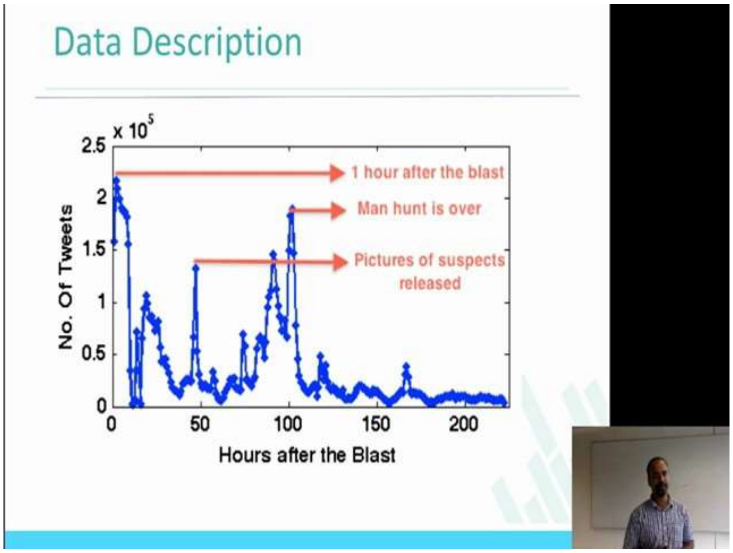
Data Description

Total tweets	7,888,374
Total users	3,677,531
Tweets with URLs	3,420,228
Tweets with Geo-tag	62,629
Retweets	4,464,201
Replies	260,627
Time of the blast	Mon Apr 15 18:50 2013
Time of first tweet	Mon Apr 15 18:53 2013
Time of first image	Mon Apr 15 18:54 2013
Time of last tweet	Thu Apr 25 01:23 2013

Datathatwascollectedduringthiseventswasactuallyabout 7.8milliontweetsthatwere collected, 3.6 million users posted this tweet and if you look at the advantage of actually working in this space of online social media is actually this large numbers that we look at, tweets with URLs is about 3.4 million, 62,000 people are posts had geo tag and about 1 percent is what Twitterclaims that the tweets that are posted on Twitter are geo tagged tweets, about 4 point 4 million replies 260,000 in the timeline of the blast. First tweet, first image, and last tweet, all of that is capture in this slide I will show it you because when you actually present and an analyze events. analyze a particular topic and I think studying this analyzing unit is only one way, but we are actually, you can actually adapt this to studying any topic.

For example in this case how do we collect the data we take hashtag Boston marathon which is actually trending and start collecting tweets, which has hashtag Boston marathon, look at other words that are in the post that has hashtag Boston marathon and use those key words to start collecting other tweets like query expansion concept and thereby we collect that post from Twitter. And this methodology can be used for collecting any data, data could be hashtag macbook pro, hashtag apple hashtag india and things like that. Not necessarily it has to be hashtag also, it could be any other words.

(ReferSlideTime:25:19)

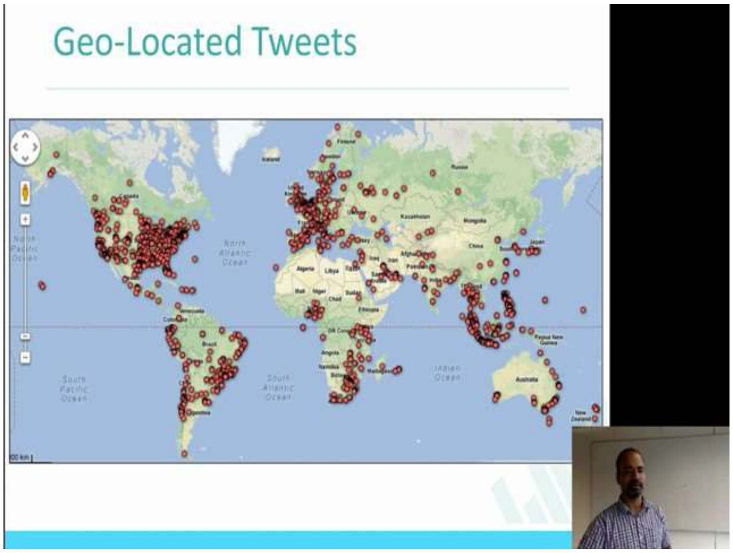


Data Description

HereIamactuallyshowingyou a graphwhichis onthe x-axis is the hours afterthe blast in the y axis is number of tweets. So, the the crux of this slide is to show you that the spikes that happen on social media is actually very very connected or correlated to the events that happened in real world, for example, here are the first spike that happened is one hour after the blast then there was a spike in Twitter tweets which is pictures ofsuspects released and man hunt is over.

So, if you reallylook at it, that is the waythat the content gets generated on social media isactuallybehavingand we've tried lookingatthesekindofblastformany,manyevents and itlooksverysimilar,interms ofactuallyspiking correlated totherealworldincident or an event that happens. And here is another slide which actually shows you the geo tagged coordinates of the tweets that were posted on Twitter.

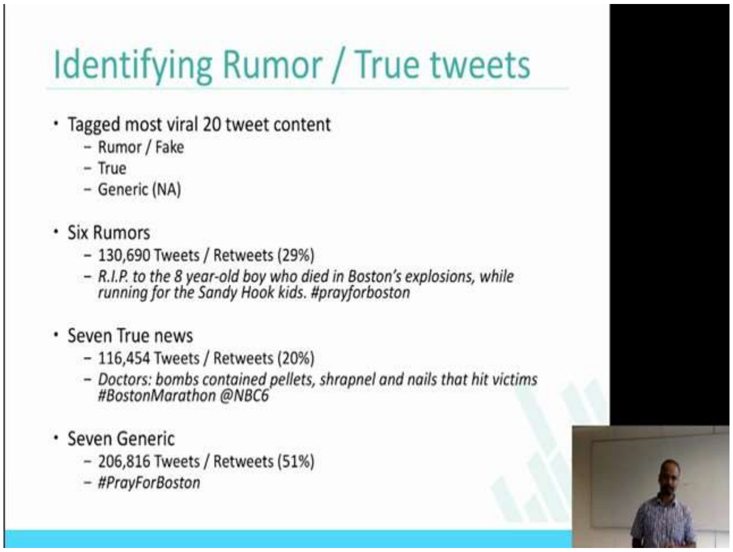
(ReferSlideTime:26:16)



Geo-Located Tweets

Particularly for this Boston marathon blast, and every dot in this slide show you the tweets that have come from that particular location, understandably the posts have come mostly from the US.

(ReferSlideTime:26:39)



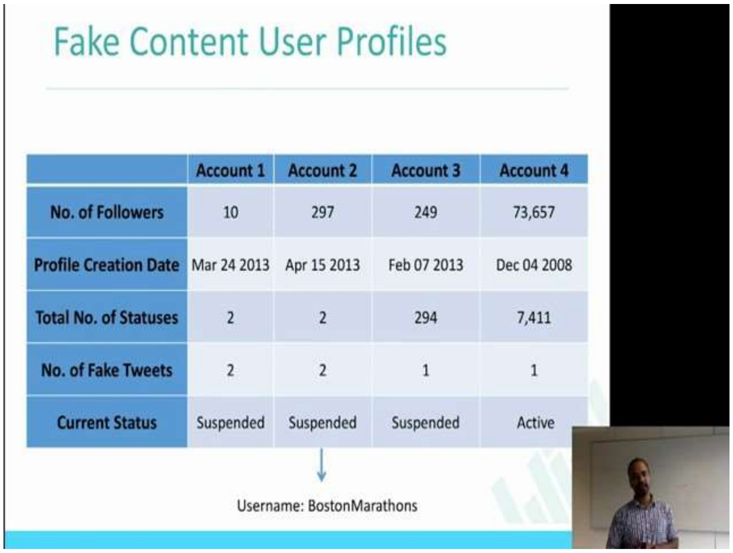
Identifying Rumor / True tweets

- Tagged most viral 20 tweet content
  - Rumor / Fake
  - True
  - Generic (NA)
- Six Rumors
  - 130,690 Tweets / Retweets (29%)
  - *R.I.P. to the 8 year-old boy who died in Boston's explosions, while running for the Sandy Hook kids. #prayforboston*
- Seven True news
  - 116,454 Tweets / Retweets (20%)
  - *Doctors: bombs contained pellets, shrapnel and nails that hit victims #BostonMarathon @NBC6*
- Seven Generic
  - 206,816 Tweets / Retweets (51%)
  - *#PrayForBoston*

So,nowinthiseventonBostonmarathonthetechniquethatwasappliedforfindingout whether a post is actually rumor, true or fake, first tagged most viral 20 tweet content which is whether it is rumor fake true or generic rumors, 6 rumors were actuallycollected from the posts that were talking about Boston marathon and seven true news was collected, which is doctored bomb contained pellets, shrapnel and nails that hit victims Boston marathon hashtag NBC6. So, those kinds of tweets were collected which is about true news that that was getting posted during the Boston marathon and six rumors werecollected and seven generic posts that had prayfor that the Boston. Prayfor Boston also was actually trending during that time.

So, essentially what we were trying to do is we were trying to study, look at the rumors thatwereposted,truenewsthatwerecoming andsome genericinformation,genericpost that happened during in the Boston marathon event. In this kind of you see a generic sense of what are the different post that happened in an event like this and rumors about 29 percent were actually retweeted, true news about 20 percent were retweeted and 51 percent of the generic content was actually fit for retweeted.

(ReferSlideTime:28:06)



Fake Content User Profiles

	Account 1	Account 2	Account 3	Account 4
No. of Followers	10	297	249	73,657
Profile Creation Date	Mar 24 2013	Apr 15 2013	Feb 07 2013	Dec 04 2008
Total No. of Statuses	2	2	294	7,411
No. of Fake Tweets	2	2	1	1
Current Status	Suspended	Suspended	Suspended	Active

Username: BostonMarathons

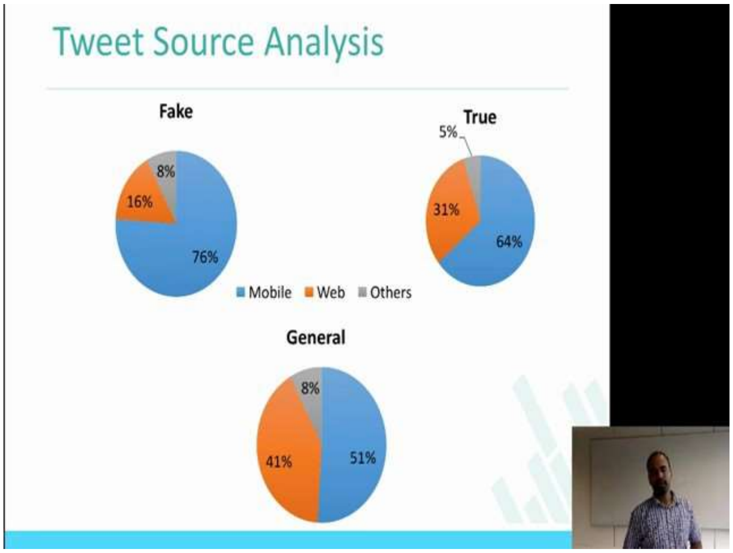
Here is another view of the data which is to show you the fake content user profiles. So, every time such event happens incidentally what happens is many of the fake user profilesgetgeneratedduringtheevent,fakeaccountsgetsgeneratedtoactuallyusethe eventtopropagatemaliciouscontent.

So, in this example I am showing you account 1, account 2, account 3, account 4 different accounts were actually created, you will actually see that the accounts has number of followers which is account 4 being very high, account 1 being very low. So, this is and if you see the account 1 which is actually created on March 24, 2013 which waspostingfakecontent,account2whichwascreatedonOctober15,2013veryclose to the Boston marathon and account 3 February 2013 and account 4 2008. Total number of status this is the updates that they created and number of fake tweets that they posted is2, 2, 1 and 1 respectively.

And if you seesome of theaccounts where getactuallysuspended and these suspensions happen because people report about this handle to Twitter in a multiple ways to actually keep down a particular account, while one is large number of people actually report a particular handle to Twitter and it gets suspended and there are through government processes you can actually apply for suspending an account.

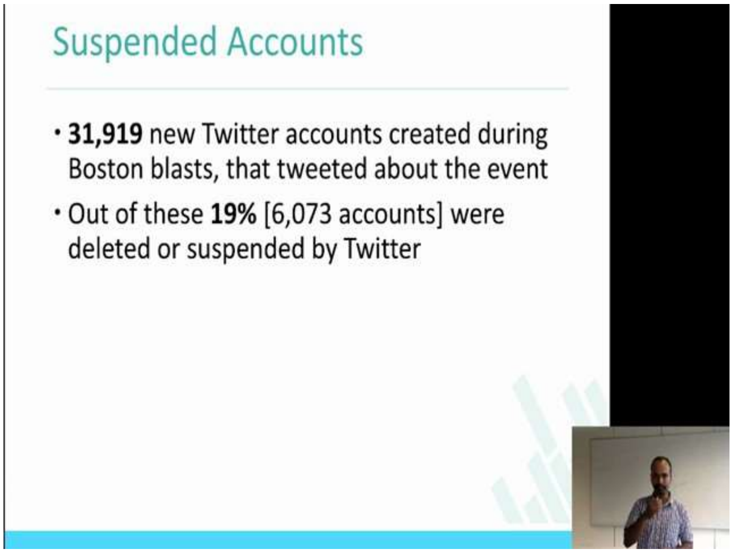
Some if you see the last column, it is interesting that some of the user handles which are posting fake content on events like Boston marathon actually are active, even when we were actually collecting the data. This shows you that fake content propagated by fake user handles and these user handles created just after the event or just before the event, just after the event happens.

(ReferSlideTime:30:04)



Now, let us look at different view of the analysis which is tweet source analysis thisgives you an insight about what devices do people use while posting the content, 76 percentofthe postthat was identifiedas fake was posted throughmobile,whereas the 64 in true and 51 in general content that were posted. This insight about what device isbeing used, while posting this content can be very useful in making decisions, for example, if you wanted totargeted advertisement, what kind of devices are being used can be very useful in making the decision. So, the device this information is available in theJSON thatyou collectfromTwitter foreverytweet. So, you can usethatto makethis judgement.

(ReferSlideTime:30:47)

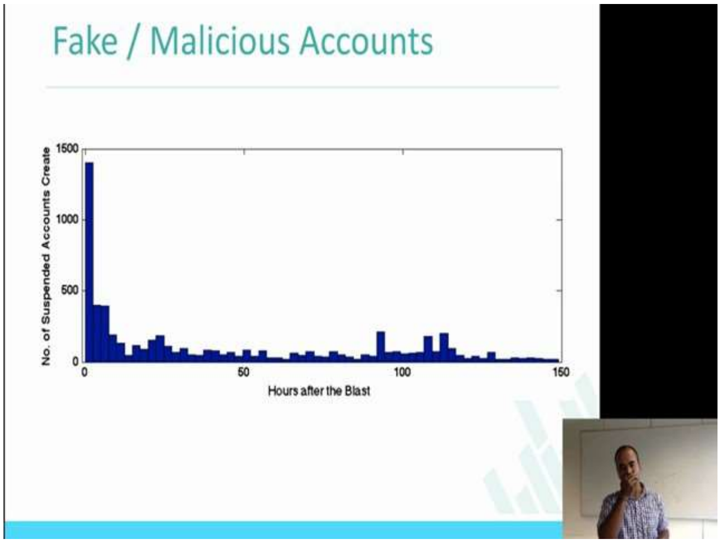


Suspended Accounts

- **31,919** new Twitter accounts created during Boston blasts, that tweeted about the event
- Out of these **19%** [6,073 accounts] were deleted or suspended by Twitter

So, if you look at the number of accounts that were created during this event, it wasabout 32,000 new Twitter accounts were created during this event, which were actually talking about this particular event. Out of this 19 percent were deleted or suspended by Twitter which again could have happen for multiple reasons and 19 percent of accounts that were created were actually suspended.

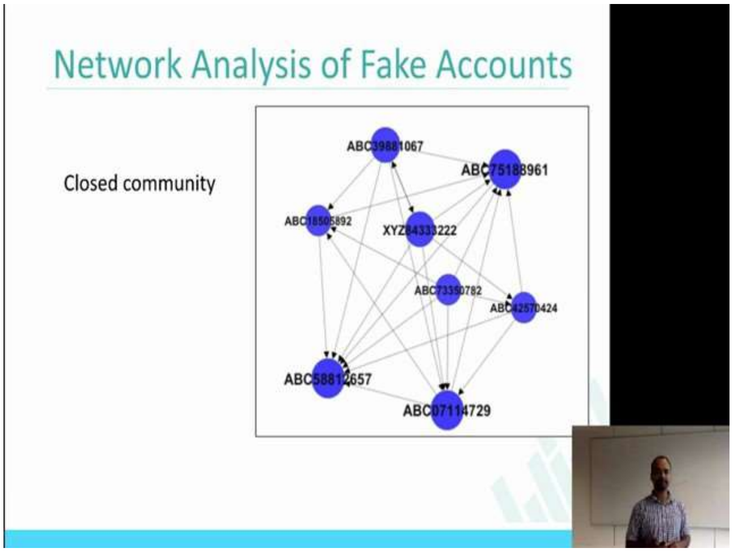
(ReferSlideTime:31:12)



Fake / Malicious Accounts

So, this is graph to show you hours after the blast; x-axis being the hours after the blast and y-axis being number of suspended accounts that were created, which is to just to show you that the number of accounts that gets created immediately after the account is also high, in addition to that number of accounts getting deleted also high the fake or maliciousaccounts thatweresuspended,thatwerecreated andsuspended wereveryhigh immediately after the event and after the event it kind of reduces little bit.

(ReferSlideTime:31:45)



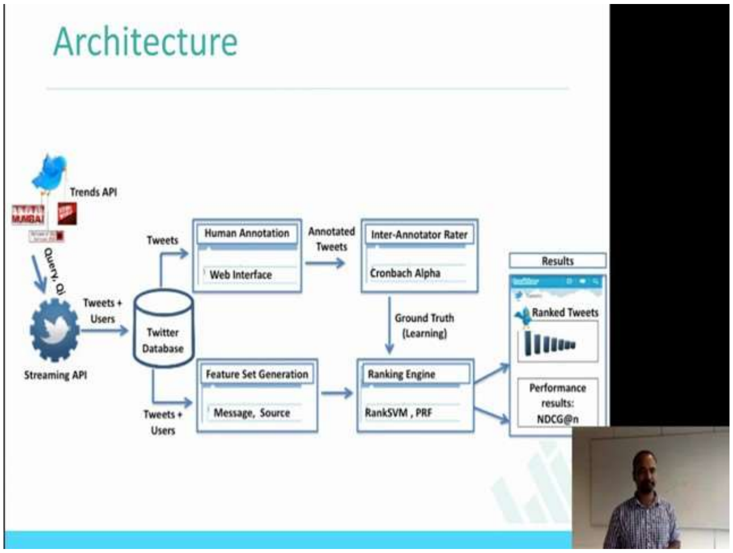
Network Analysis of Fake Accounts

Closed community

Let us look at the connection between the fake accounts itself. So, again just keep in mind the kinds of analysis that were doing is who, when, where and what, whyand how. So, again that one insight into the analysis is this how were the people who were posting this content which is fake are connected. So, one insight is that they are actually pretty closed and this is not only in this domain you can actually seen this kind of analysis in many other domains also, for example, in classical security problem like phishing.

The number of groups, number of accounts, number of sets of people who do this actually by it is small and they are all very well connected, similar kind of inferences is derived from this particular analysis also where for the Boston marathon if you look at people,thenodehereis theuserandtheedgebetweenthenodesaretheactionofretweet followingandfollowers.So,therefore,thereisathereisaclosedcommunitythatis actuallyoperatingin termsofpostingthisfake content.

(ReferSlideTime:32:51)



Architecture

Trends API

Query, QI

Tweets + Users

Streaming API

Twitter Database

Tweets

Human Annotation

Web Interface

Annotated Tweets

Inter-Annotator Rater

Cronbach Alpha

Ground Truth (Learning)

Tweets + Users

Feature Set Generation

Message, Source

Ranking Engine

RankSVM , PRF

Results

Ranked Tweets

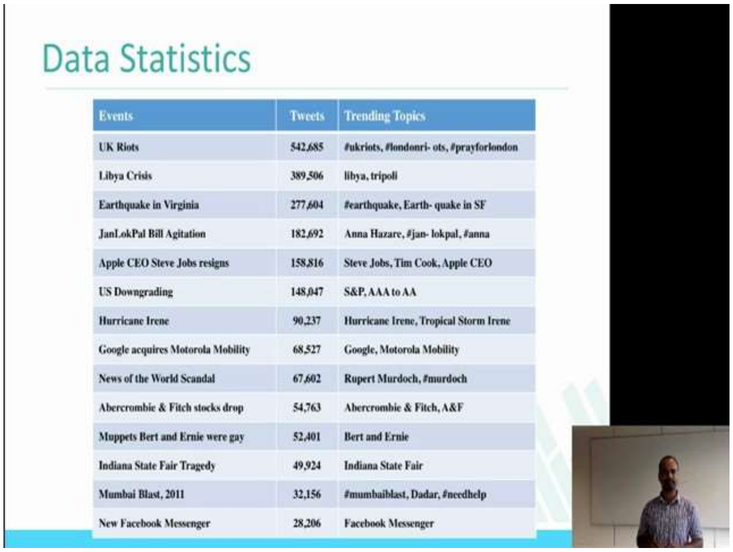
Performance results: NDCG@n

So, while doing this you could actually think about. So, earlier I think in week one I actually showed you some very high level slide about how machine learning and about how these kind of approaches of identifying fake and legitimate can be done, this is just to slightly zoomed in view of the same slide, which is to show you that the data is comingfromTwitter throughstreamingAPI's.Assumeyouknow,everybodyknowsnow what is streaming API is, which is to collect data from the social networks, tweets are dumped to the database, human annotations which we saw earlier also in terms of annotating the post even now we saw about fake, generic and true our post those are all sometimes human annotated, sometimes you could actually use some simple techniques to do the annotation, one of the important thing that you also want to do in this annotations are done are inter annotator agreement which is if I say something that is legitimate and if you say something that is legitimate then probably more people saying post is legitimate, then the post must be legitimate, that is the kind of intuition that the Cronbach's Alpha, which is value that you may get for finding out inter annotator agreement and at the Cronbach's Alpha is generally about 0.7.

Itis actuallyunderstoodthat the data has youcan have more confidence intheinferences that you are drawing from this particular data. Cronbach's Alpha is the value that youwill calculate while finding out inter annotator agreement and so as we discussed earlier also feature extraction, feature extraction is a technique by which we took F1, F2 and those things you will use that to find the model here.

I will just describe a little bit in the later slides about what model can be generated and you use that to find out whether this particular post is legitimate or not and then you can actuallyshow thatto theuseralso,thatis the architecturethatis presented hereagainis a very simple machine learning approach which is take the posts, use the post, do some feature extraction, use the feature extraction to create a model, use the model to actually predict whether the post is legitimate or not.

(ReferSlideTime:35:14)



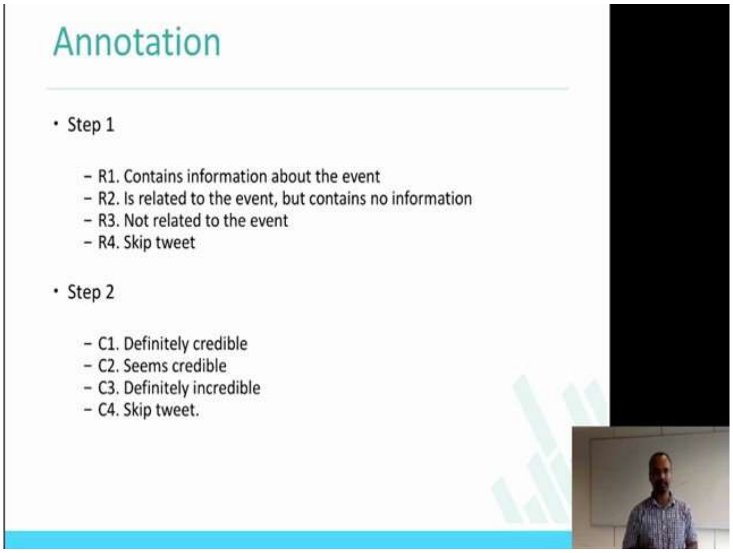
Data Statistics

Events	Tweets	Trending Topics
UK Riots	542,685	#ukriots, #londonri- ots, #prayforlondon
Libya Crisis	389,506	libya, tripoli
Earthquake in Virginia	277,604	#earthquake, Earth- quake in SF
JanLokPal Bill Agitation	182,692	Anna Hazare, #jan- lokpal, #anna
Apple CEO Steve Jobs resigns	158,816	Steve Jobs, Tim Cook, Apple CEO
US Downgrading	148,047	S&P, AAA to AA
Hurricane Irene	90,237	Hurricane Irene, Tropical Storm Irene
Google acquires Motorola Mobility	68,527	Google, Motorola Mobility
News of the World Scandal	67,602	Rupert Murdoch, #murdoch
Abercrombie & Fitch stocks drop	54,763	Abercrombie & Fitch, A&F
Muppets Bert and Ernie were gay	52,401	Bert and Ernie
Indiana State Fair Tragedy	49,924	Indiana State Fair
Mumbai Blast, 2011	32,156	#mumbaiblast, Dadar, #needhelp
New Facebook Messenger	28,206	Facebook Messenger

So,inusingthis architecture justtaking,insteadofjust doingoneortwoevents, multiple events data were collected and used to find out whether a particular technique, technology can be identified where this post is legitimate or not and here are the events UKriots,Libyacrisis,earthquakeinVirginiaandUSdowngradingtherearemany,many events data were collected and in as I said before the column in the third column here which is trending topics.

These were the topics that were trending using which the data was collected; the column 2 shows you the number of tweets, again alarge number of datawas used in while doing this analysis.

(ReferSlideTime:35:58)



Annotation

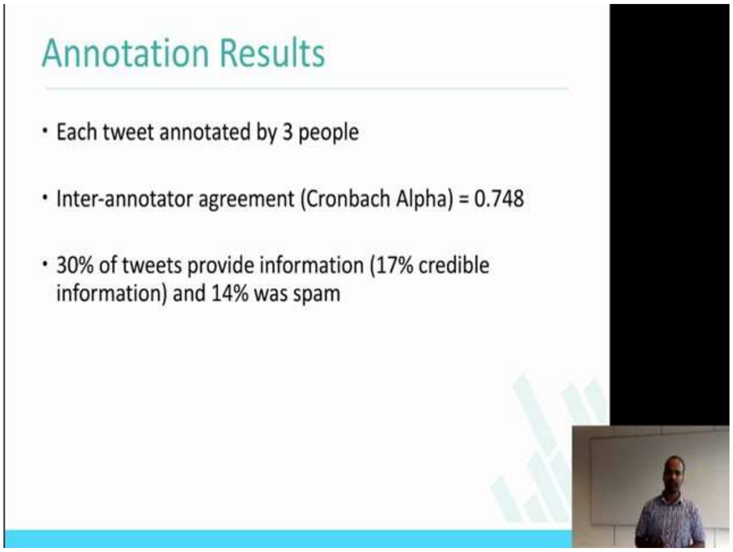
- Step 1
  - R1. Contains information about the event
  - R2. Is related to the event, but contains no information
  - R3. Not related to the event
  - R4. Skip tweet
- Step 2
  - C1. Definitely credible
  - C2. Seems credible
  - C3. Definitely incredible
  - C4. Skip tweet.

As discussed before, one of the methods used to find out whether particular post is legitimateornot,annotationwasdonetherearemultipleways todothisannotation.Also you could get 3 of your friends to sit down together and I will tell you whether this every post is legitimate or fake, you can do through mechanical turk, mechanical turk is a crowd source mechanism by which you could actually show these posts in theplatform of mechanical turk. The turkers also to look at it, turkers are basically people all around the world who are doing this task for a small money, step one in this case contains information about the tweet, postis shown to the user and in the user actually decides on one of these four characteristics which is contains information,is related to the event not only related to the events, skip.

Ifin thesteponeandsays thatcontainstheuserdecides thatthereisainformationinthis post,thentheuserisaskedaboutdefinitelycredible,seemscredible,definitelyincredible and skip tweet, again here, I am only going through the methodology which is post is taken.Itisannotatedyoucouldannotateitforanyparticulartopicthatyouwouldwhere you want to study, in this case it is credibility, but you could also think about it whether this post has phising URL or not, if this post is talking about a particularly event or not, this post is sensitive or not you could do many, many things in terms of annotations and in the topics that you are interested in studying from the post that is being collected.

So, from step one you take the data and then you ask the users to classify, it as definitely credible, seems credible, definitely credible and if there is nothing the user cannot makea decision, skip the tweet.

(ReferSlideTime:37:41)

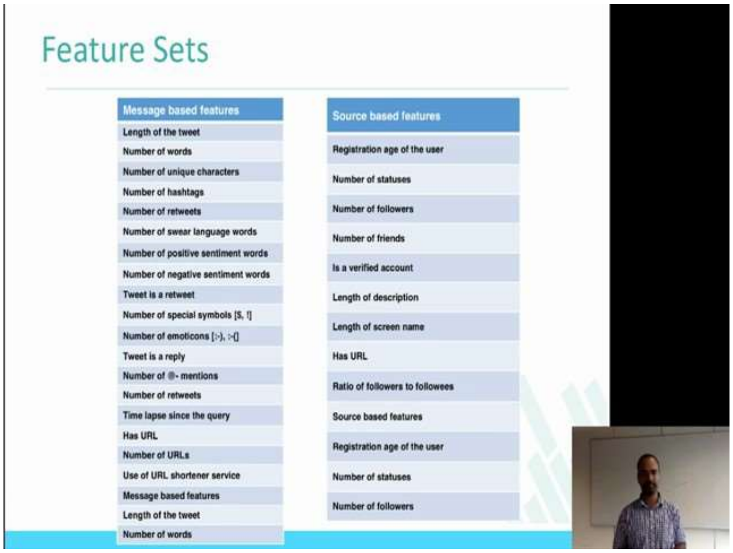


Annotation Results

- Each tweet annotated by 3 people
- Inter-annotator agreement (Cronbach Alpha) = 0.748
- 30% of tweets provide information (17% credible information) and 14% was spam

AndthatiswhyIsaidabout Cronbach'sAlphawhichissomethingIwillemphasizehere, each tweet should be annotated by at least three people because that will give you more confidence in the data and when you do this it is called inter annotator agreement or Cronbach's Alpha. If you calculate that and if it is more than about 0.7 it is generally accepted that the data has more value or confidence in it, 30 percent of the tweetsprovide information which is in the step one users agree that 30 percent of the tweets are shown to them add information, only 17 percent have credible information and 14 percent was spam.

(ReferSlideTime:38:21)



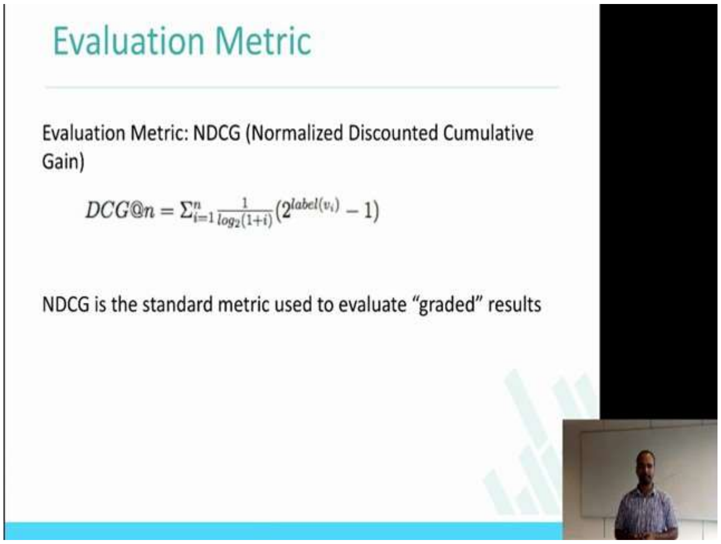
Feature Sets

Message based features
Length of the tweet
Number of words
Number of unique characters
Number of hashtags
Number of retweets
Number of swear language words
Number of positive sentiment words
Number of negative sentiment words
Tweet is a retweet
Number of special symbols [$, !]
Number of emoticons [:-), :-(]
Tweet is a reply
Number of @- mentions
Number of retweets
Time lapse since the query
Has URL
Number of URLs
Use of URL shortener service
Message based features
Length of the tweet
Number of words

Source based features
Registration age of the user
Number of statuses
Number of followers
Number of friends
Is a verified account
Length of description
Length of screen name
Has URL
Ratio of followers to followees
Source based features
Registration age of the user
Number of statuses
Number of followers

So, feature sets, we now discussed just now some time back about different features F1 and F2 in that slide here message based features and source base features which is again, if you lookatitItold you aboutfeaturesfromthepostswhich is messagebased features. So, features from the profile which is a source based features.

(ReferSlideTime:38:46)



Evaluation Metric

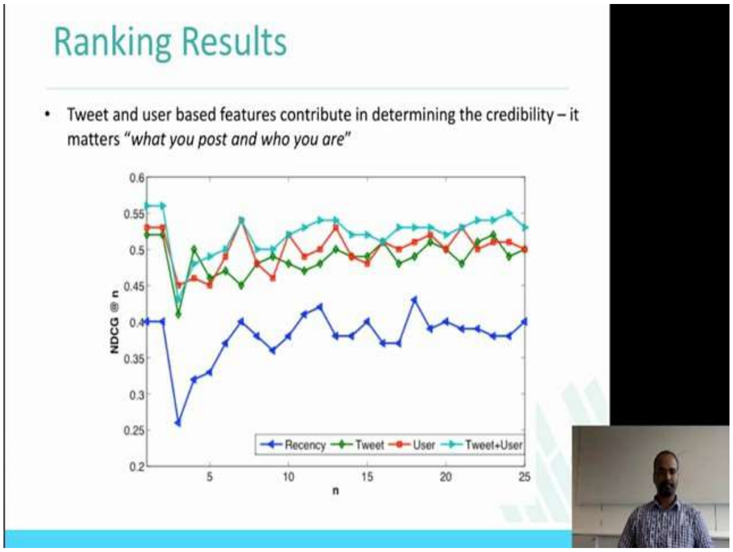
Evaluation Metric: NDCG (Normalized Discounted Cumulative Gain)

$$DCG@n = \Sigma_{i=1}^{n} \frac{1}{log_2(1+i)}(2^{label(v_i)} - 1)$$

NDCG is the standard metric used to evaluate "graded" results

Using these features we used a metric called the NDCG, which is normalized discounted cumulative gain it is nothing, but way by which you can actually mention the efficiency of the search, NDCG is being commonly used in finding, how good a search engine is performing in this case, we are using it to find out how what is the quality of the classification that we make whether it is legitimate or fake in this metric called NDCG.

(ReferSlideTime:39:20)



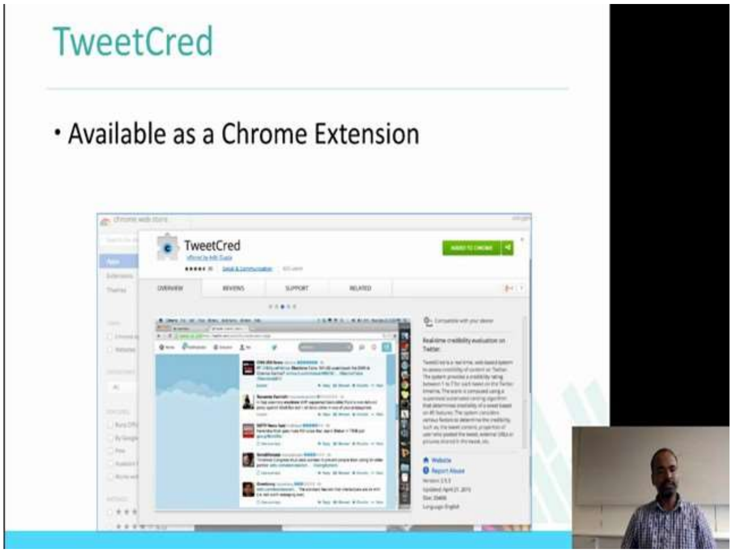
Ranking Results

- Tweet and user based features contribute in determining the credibility – it matters *"what you post and who you are"*

Also here is a graph for looking at the content from the tweets that we collected looking at the post it that we collected, recency, tweet, user, Twitter plus user, these are the features that we used. If you remember to find out whether a post is legitimate or not we are here we are drawinggraph of n,which is on the x-axis and NDCG value, which is on the y-axis, you can clearly see that the tweet plus user which is at the top of the graph doing well in terms of the NDCG values. This basically helps to understand that what you post and who you are a good features to make judgment on whether the contentis legitimate or not, that is the kind of inferences that you should be chasing while you are analyzing the content from social media which helps in some actionable information also.

For example, here what you post and who you are helps to find out whether the post is credible or not which helps in making lot of decisions.

(ReferSlideTime:40:34)

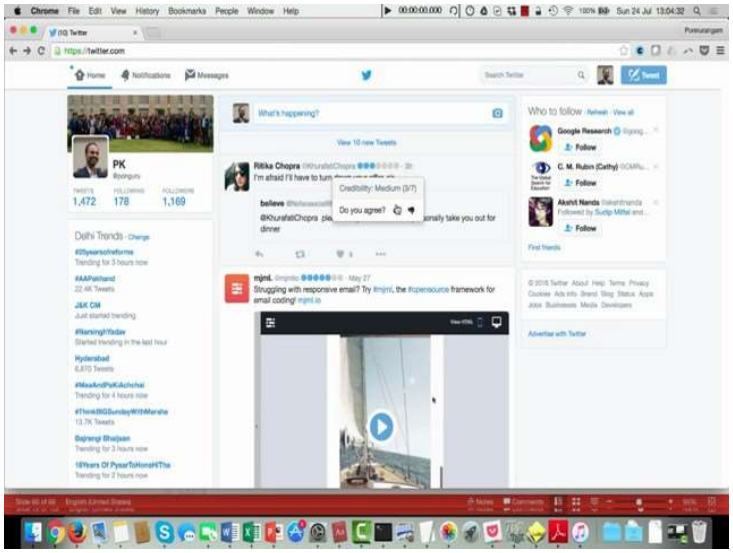


TweetCred

- Available as a Chrome Extension

Using this understanding of what feature works and what features do not work and what feature actually influences in finding out whether the post is credible or not, the TweetCred, a chrome browser extension was built and this extension it helps you to identify whether this particular post is credible or not. I'll just show you a light demo of the TweetCred extension and then I will walk you through, what this available in the chrome extension.

(ReferSlideTime:41:02)



Ritika Chopra
I'm afraid I'll have to turn

Credibility: Medium (3/7)

Do you agree?

believe
@KhurafatiChopra ple... onally take you out for dinner

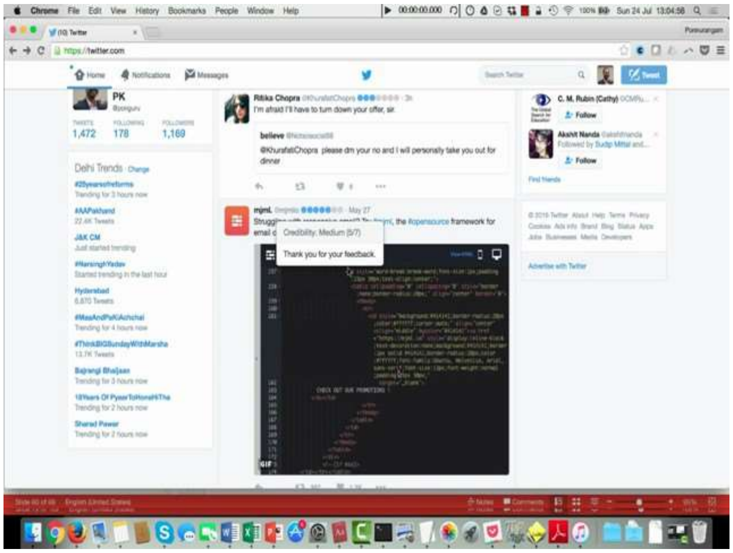
mjml. · May 27
Struggling with responsive email? Try #mjml, the #opensource framework for email coding! mjml.io

What I am showing you here is the Tweetcred browser plugin, chrome browser plugin which actually helps you in making a decision whether a particular tweet is credible or not. It just gives you; it basically uses all the features that we discussed until now,where it is being bundled with this chrome extension.

So, look at this tweet, if you go to Twitter dot com in your timeline, this information about whether a post is how credible is it will not be there this is coming from TweetCred. If you look at this it actually gives you a value of 3 on 7, it is calculating the value of credibility on a scale of 1 to 7 and in this case it is showing that this is my timeline in this case it is showing that this post is 3 on 7, this post is5 on 7 and values like that. So, it is going to work for all the post that are in a timeline, it is going to work for what in your search its going to work for post in dm and things like that.

So, let us look at the values that it is presenting also. So, if you see here it is actually showing you a value of 3 on 7 and then when I hover it. It actuallygives me information called credibility medium 3 on 7, do youagree? So, this is the way if you remember,the machinelearningmodelthatpeopleusingthefeatureyoutakethatmoduleandwhenever we get feedback like this, we can actually go and update that model to make it more efficient.

(ReferSlideTime:42:55)



Ritika Chopra @KhurafatiChopra · 3h
I'm afraid I'll have to turn down your offer, sir.

believe
@KhurafatiChopra please dm your no and I will personally take you out for dinner

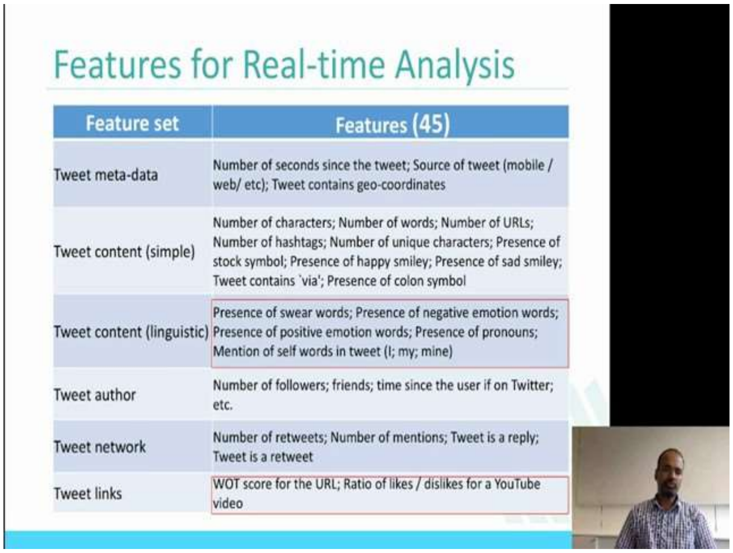
mjml. · May 27
Struggling … #opensource framework for email …

Credibility: Medium (5/7)

Thank you for your feedback.

So, inthiscase you could actuallysaythat no,I actuallyagree with the value of3 on7, it gives you message saying thank you for the feedback. In this case, let us take it if I were to say that the value I do not agree with value then it actually asking what you are agreeing with. So, I say no this is actually more credible it should be actually 7, when it says thank you for feedback. So, what it basically does is, it is capturing these details from you and it is going to make use of it when we end up updating the model that was built at the back end for the TweetCred. This information can be used in making the judgment. So, that is the chrome browser extension of TweetCred, which basically takes the features in real time and makes the judgments and presents it to the user with the values of 1 to 7.

(ReferSlideTime:43:27)

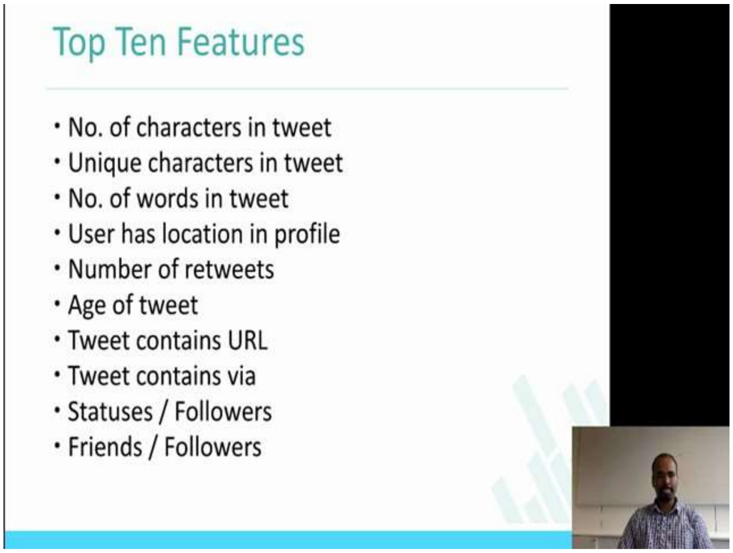


Features for Real-time Analysis

Feature set	Features (45)
Tweet meta-data	Number of seconds since the tweet; Source of tweet (mobile / web/ etc); Tweet contains geo-coordinates
Tweet content (simple)	Number of characters; Number of words; Number of URLs; Number of hashtags; Number of unique characters; Presence of stock symbol; Presence of happy smiley; Presence of sad smiley; Tweet contains `via'; Presence of colon symbol
Tweet content (linguistic)	Presence of swear words; Presence of negative emotion words; Presence of positive emotion words; Presence of pronouns; Mention of self words in tweet (I; my; mine)
Tweet author	Number of followers; friends; time since the user if on Twitter; etc.
Tweet network	Number of retweets; Number of mentions; Tweet is a reply; Tweet is a retweet
Tweet links	WOT score for the URL; Ratio of likes / dislikes for a YouTube video

So, you may remember the features that we discussed in this lecture, but unfortunatelyall features cannot be actually used while doing it in real time, for example, finding out all the followers that you have and using some scores on the followers is actually hard. So, here are the 45 features that were actually used to while doing the real time analysis itself, specially I wanted to highlight on the presence of swear words, presence of negative emotion words, presence of positive emotion words, web of trust score, whichis WOT scorefor the URL and ratio of likes and dislikes fromthe YouTubevideo which has links to the YouTube. So, these are the features that we did not discuss before. So, I kind of thought we'll highlight them when I'm presenting the slide. So, these features were used in building tweetcred demo that I showed you.

(ReferSlideTime:44:24)

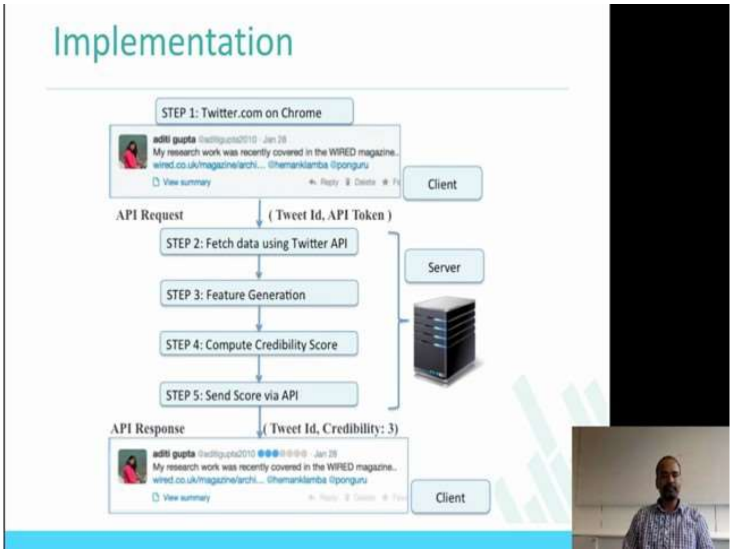


Top Ten Features

- No. of characters in tweet
- Unique characters in tweet
- No. of words in tweet
- User has location in profile
- Number of retweets
- Age of tweet
- Tweet contains URL
- Tweet contains via
- Statuses / Followers
- Friends / Followers

Of course, the common question is what are the top 10 features that actually makes the decision or which influences in identifying whether post is a credible or not more efficiently.Itisnumberofcharactersinthetweet,uniquecharactersinthetweet,number of words in the tweet, user has location in the profile, number of retweets, age of tweet, tweet contains URL, tweet contains via which is through, how the post was done, status and followers, friends and followers, thoseare thetop 10 features of fromTwitter, which can be actually used to make a judgment on whether a post is legitimate or not. Please keepinmindthisisonlyforTwitter,thefeaturesthatyoulookforFacebook,thefeatures that you may look for Instagram, in other social network may be very different.

(ReferSlideTime:45:16)



Implementation

STEP 1: Twitter.com on Chrome

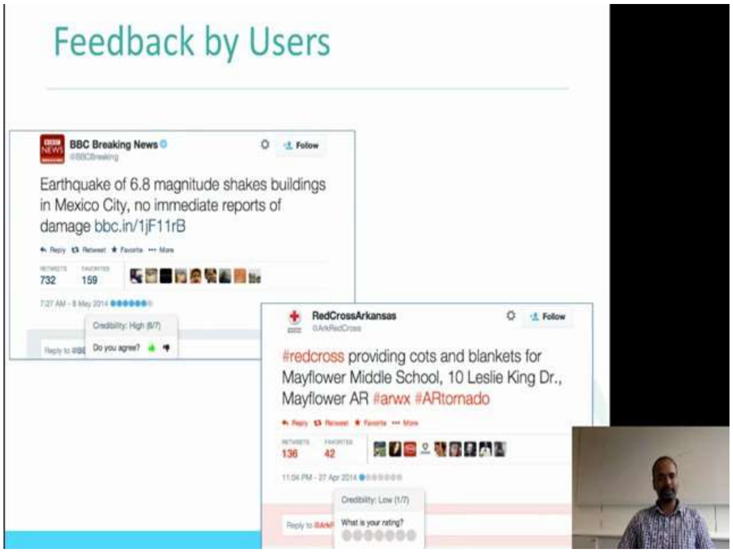
aditi gupta @aditigupta2010 · Jan 28
My research work was recently covered in the WIRED magazine..
wired.co.uk/magazine/archi... @hemanklamba @ponguru

View summary
Reply
Delete
Client
API Request
(Tweet Id, API Token)
STEP 2: Fetch data using Twitter API
STEP 3: Feature Generation
STEP 4: Compute Credibility Score
STEP 5: Send Score via API
Server
API Response
(Tweet Id, Credibility: 3)

aditi gupta @aditigupta2010 ●●●○○○○ · Jan 28
My research work was recently covered in the WIRED magazine..
wired.co.uk/magazine/archi... @hemanklamba @ponguru

Here is just a slide to show you how the implementation for the tweetcred was done which is chrome browser extension. There is a post that is on your timeline, it takes the post fetches data using Twitter API, which is the architecture that I showed you earlier, where feature extractions were done. Then, the model which is built, tweet is taken through API feature extracted the credibility score is computed with the techniques that we discussed until now and the values assigned back to the API, and then tweet ID and the credibility value comes back, it is presented in your timeline saying this value is actually 3 on 7, the demo I showed you. It is simple chrome extension that was about to show these values.

(ReferSlideTime:46:05)



Feedback by Users

BBC Breaking News
@BBCBreaking

Earthquake of 6.8 magnitude shakes buildings in Mexico City, no immediate reports of damage bbc.in/1jF11rB

RETWEETS 732 FAVORITES 159

7:27 AM - 8 May 2014

Credibility: High (6/7)

Do you agree?

RedCrossArkansas
@ArkRedCross

#redcross providing cots and blankets for Mayflower Middle School, 10 Leslie King Dr., Mayflower AR #arwx #ARtornado

RETWEETS 136 FAVORITES 42

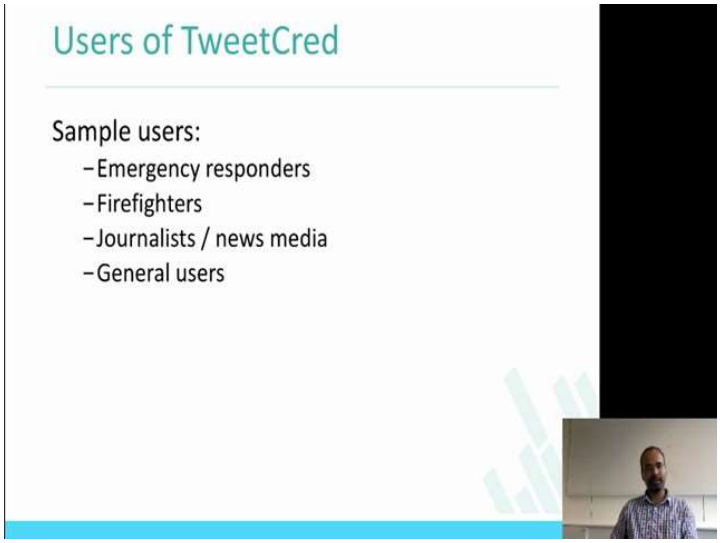
11:04 PM - 27 Apr 2014

Credibility: Low (1/7)

What is your rating?

So, users can also give feedback to the system and that is showed in your demo TweetCred actually ask user to say agree or disagree with the values that is presented. If you agree that is okay, if you don't agree please provide the information, please providea value that it should be what you think it should be, that is what is presented in the top left which is when you agree , bottom right is actually saying if you disagree.

(ReferSlideTime:46:35)

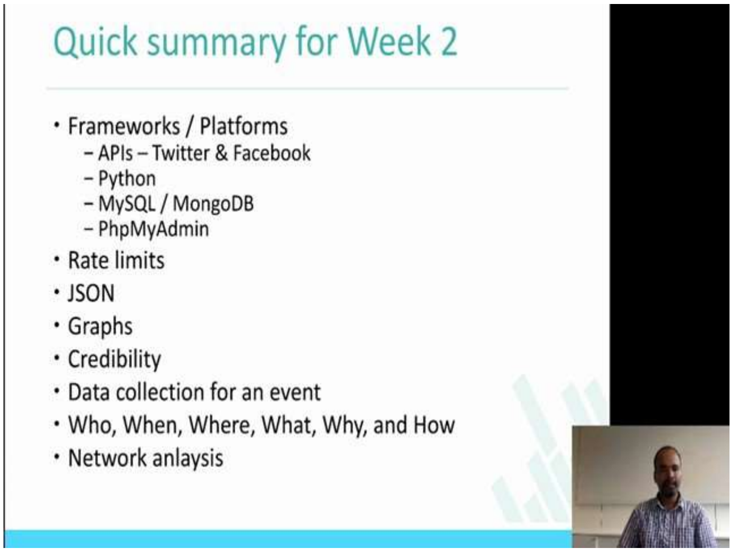


Users of TweetCred

Sample users:
- Emergency responders
- Firefighters
- Journalists / news media
- General users

Different types of users that can be foreseen using TweetCred type of tools. You can at leastremember TweetCredisonlyoneexamplethatIampresentinghere,therearemany other tools that one to think of while analyzing social media content and information presented to the user. In this case emergency responders, fire fighters, journalists and news media and general users also have started using tools like this.

(ReferSlideTime:47:01)



Quick summary for Week 2

- Frameworks / Platforms
  - APIs – Twitter & Facebook
  - Python
  - MySQL / MongoDB
  - PhpMyAdmin
- Rate limits
- JSON
- Graphs
- Credibility
- Data collection for an event
- Who, When, Where, What, Why, and How
- Network anlaysis

Let us do a quick summary of week 2, when we started we actually looked at API's, programming interfaces and we you also have a tutorial for Facebook in this week. Then we looked at very, very briefly what Python programming language, MySQL, Mongo DB, PhpMyAdmin and when you collect the data you are going to actually get the rate limits. Please remember that there is always going to be rate limits when you are collecting the data from these social media services and we talked a little bit about the format in which the social media service is store the data, which is JSON, when you collect the data, you are going to get JSONs which you have to analyze through your scripts and Facebook stores all the data in terms of a graph. We looked at that brieflythen westarted digging deeper intotrustincredibilityasafocusarea.Welookedatthese concepts of trust and credibility through events being Boston marathon was one of the events, Hurricane Sandy is another event that we looked at we looked at these events. ThroughtheseeventsthatIwastryingtotellyouhowdataisbeinganalyzed,whatkind oftechniquesare beingapplied onthisdata.

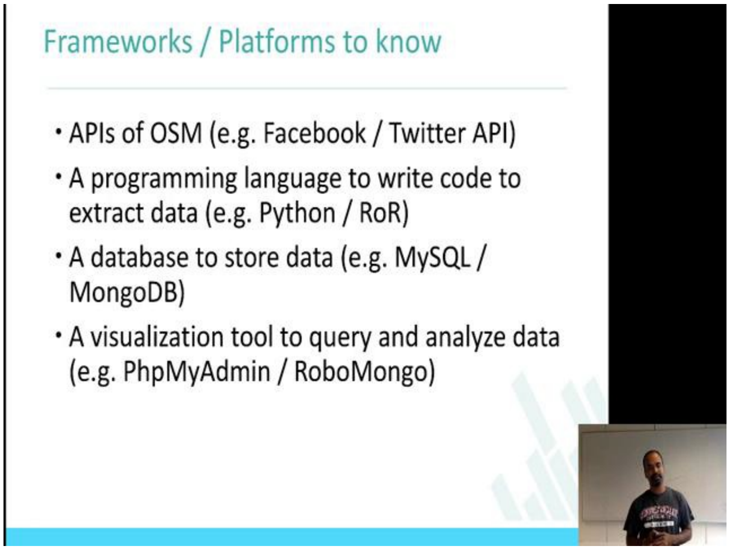
We looked at classification is one of the major technique that is used while designing whether a post is credible or not and during this analysis, I also told you about who, when, where and what, why and how are the basic questions that you can actually analyze using in the social media content and specifically we have also looked at some social network analysis techniques inputs. So, that is the week 2, hopefully you will go through the content and if you have any questions please go and ask in the forum, we’ll be happy to actually answer there.

Unit-2

MisinformationonSocialMedia

Welcomeback to thecoursePrivacyandSecurityin OnlineSocialMedia, this isweek 3. Let me put you go over what we covered in week 2.

(ReferSlideTime:00:20)

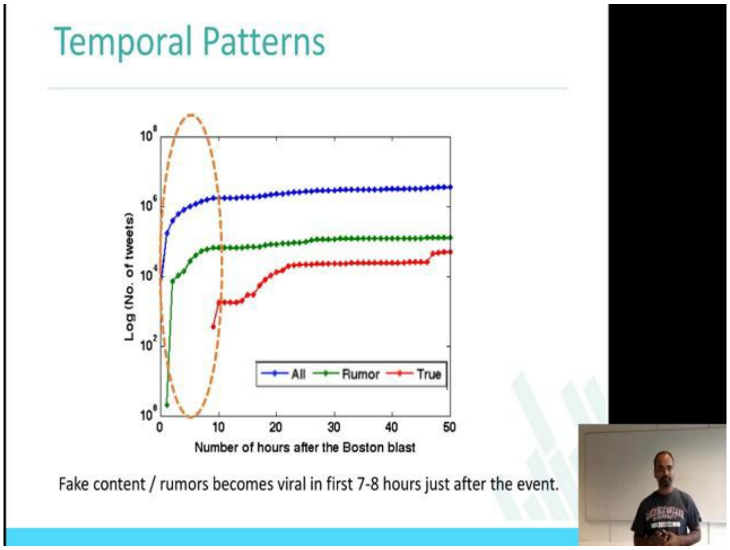


Frameworks / Platforms to know

- APIs of OSM (e.g. Facebook / Twitter API)
- A programming language to write code to extract data (e.g. Python / RoR)
- A database to store data (e.g. MySQL / MongoDB)
- A visualization tool to query and analyze data (e.g. PhpMyAdmin / RoboMongo)

We started looking at what an API is, and then we looked at what Python Programming Languages, MongoDB, how the data is stored, how we can actually visualize the data using PhpMyAdmin or RoboMongo. I hope all of you have already done hands on exercises and practices with Facebook API and twitter API.

(ReferSlideTime:00:45)

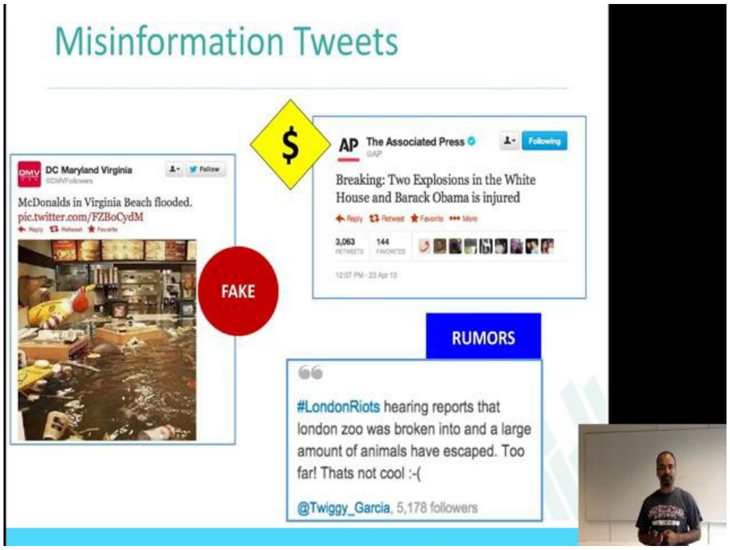


Temporal Patterns

Fake content / rumors becomes viral in first 7-8 hours just after the event.

After that we looked at the topic Trust and Credibility, in that we looked at multiple events; through the events we looked at some concepts. Here is a slide that I used inweek 2, where we showed that the truthful information is coming into the social media slower than the rumours, and there are multiple techniques by which you can actually attack this problem.

(ReferSlideTime:01:10)



Misinformation Tweets

DC Maryland Virginia @DMVFollowers

McDonalds in Virginia Beach flooded. pic.twitter.com/FZBoCydM

The Associated Press @AP

Breaking: Two Explosions in the White House and Barack Obama is injured

3,063 RETWEETS 144 FAVORITES

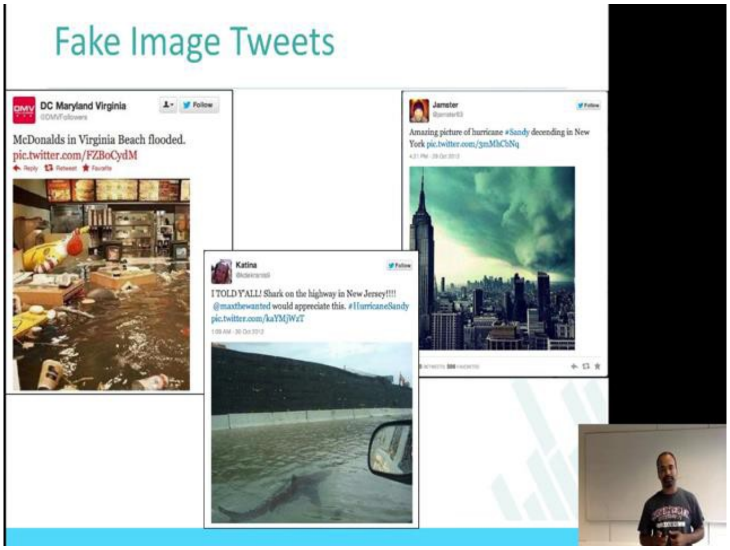
12:07 PM - 23 Apr 13

#LondonRiots hearing reports that london zoo was broken into and a large amount of animals have escaped. Too far! Thats not cool :-(

@Twiggy_Garcia, 5,178 followers

And I also showed you some examples about Misinformation; tweets that were being posted on social media and there have been multiple effects of it; fake content getting viral, some values on the stock market getting affected and of course rumours also.

(ReferSlideTime:01:30)



Fake Image Tweets

DC Maryland Virginia
@DMVFollowers

McDonalds in Virginia Beach flooded. pic.twitter.com/FZBoCydM

Katina

I TOLD Y'ALL! Shark on the highway in New Jersey!!!! @maxthewanted would appreciate this. #HurricaneSandy pic.twitter.com/kaYMjWzT

1:09 AM - 30 Oct 2012

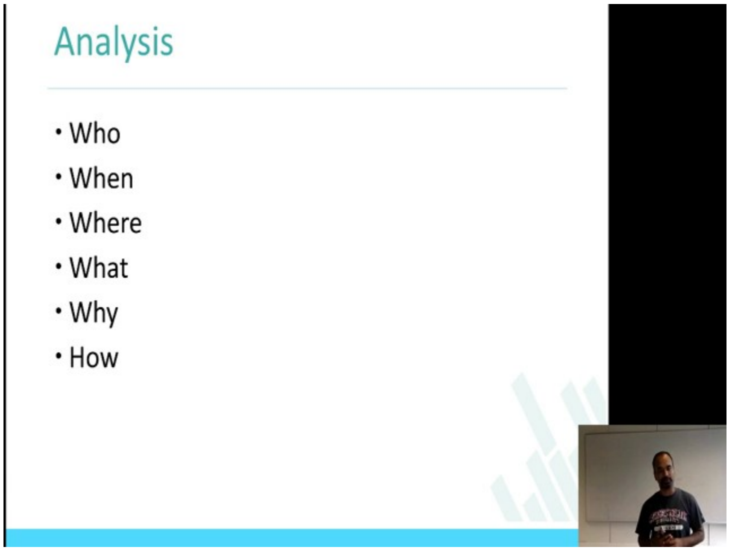
Jamster

Amazing picture of hurricane #Sandy decending in New York pic.twitter.com/3mMhCbNq

4:21 PM - 29 Oct 2012

More examples that I showed you in week 2, showing that there is a shock in the hurricane sandywhere this picture got viral and it had effects on the public also.

(ReferSlideTime:01:45)

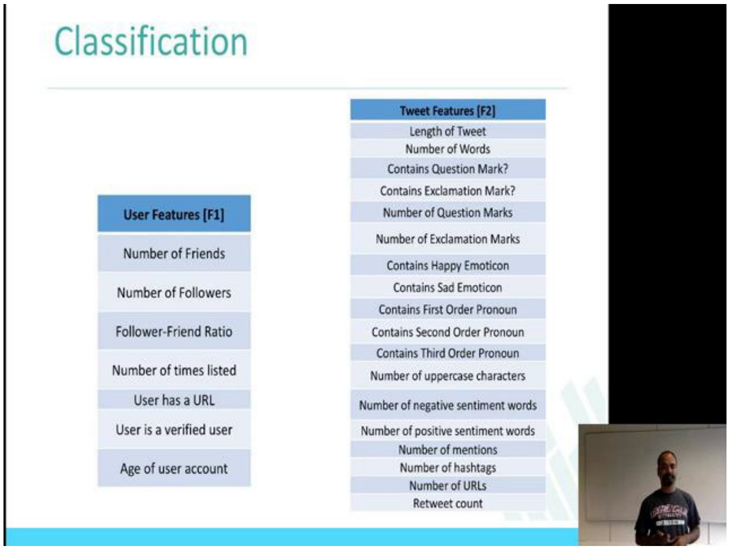


Analysis

- Who
- When
- Where
- What
- Why
- How

Most specifically I was trying to give you an intuition about how, what analysis can be done. Particularly; who, when, where, what, why, and how. These are the kind ofanalysis that you should be interested indoing while looking at the socialmedia content.

(ReferSlideTime:02:04)



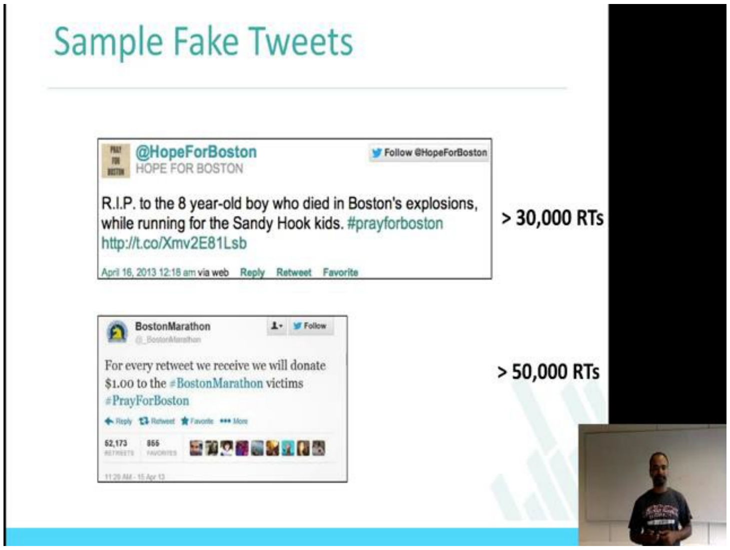
Classification

User Features [F1]
Number of Friends
Number of Followers
Follower-Friend Ratio
Number of times listed
User has a URL
User is a verified user
Age of user account

Tweet Features [F2]
Length of Tweet
Number of Words
Contains Question Mark?
Contains Exclamation Mark?
Number of Question Marks
Number of Exclamation Marks
Contains Happy Emoticon
Contains Sad Emoticon
Contains First Order Pronoun
Contains Second Order Pronoun
Contains Third Order Pronoun
Number of uppercase characters
Number of negative sentiment words
Number of positive sentiment words
Number of mentions
Number of hashtags
Number of URLs
Retweet count

And then, we later looked at different features that are available in tweets particularly user features and that tweet features and we tell detail about what these features mean, how these features can be put together to create a classifier which can look at tweet and then say that whether it is legitimate or fake tweet.

(ReferSlideTime:02:24)



Sample Fake Tweets

@HopeForBoston
HOPE FOR BOSTON

Follow @HopeForBoston

R.I.P. to the 8 year-old boy who died in Boston's explosions, while running for the Sandy Hook kids. #prayforboston http://t.co/Xmv2E81Lsb

April 16, 2013 12:18 am via web Reply Retweet Favorite

> 30,000 RTs

BostonMarathon
@_BostonMarathon

Follow

For every retweet we receive we will donate $1.00 to the #BostonMarathon victims #PrayForBoston

Reply Retweet Favorite More

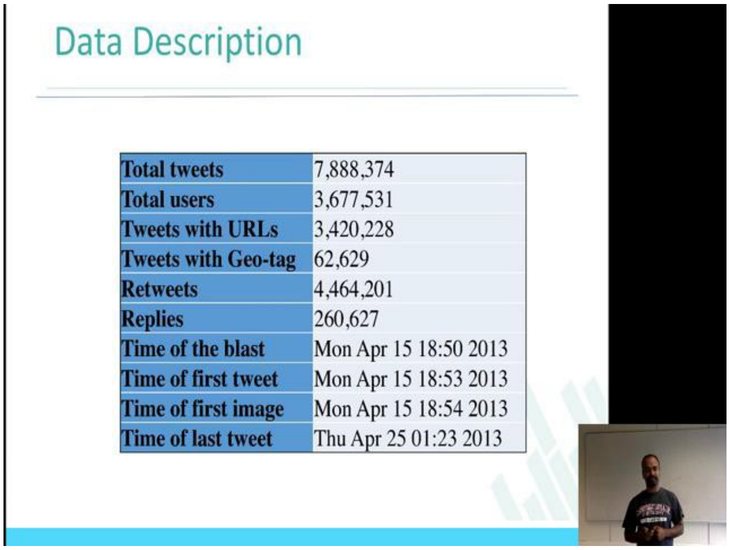
52,173 RETWEETS 865 FAVORITES

11:29 AM - 15 Apr 13

> 50,000 RTs

Some more examples, particularlythis is fromthe BostonMarathonwhere I showed you that a tweet which said, RIP to the 8 year-old who died in Boston explosion was retweeted more than 30000 times and malicious user used this spread or occurence ofan events to actually spread the content and get victims to go to malicious URL's which share malicious information.

(ReferSlideTime:02:51)

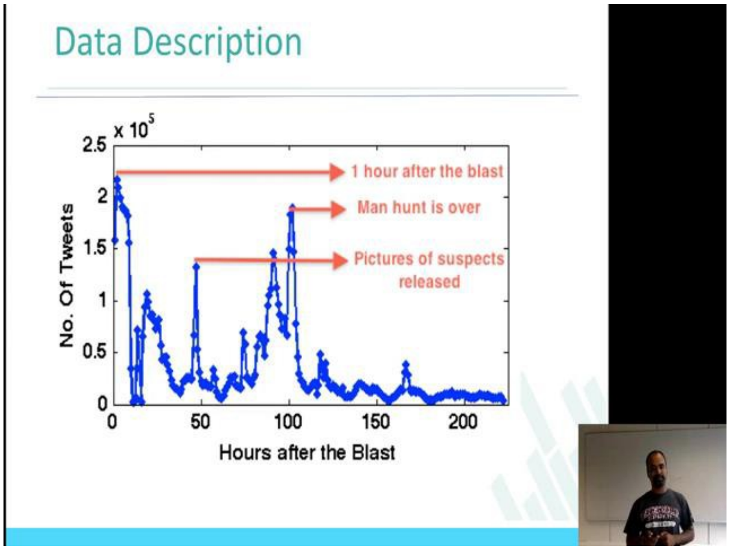


Data Description

Total tweets	7,888,374
Total users	3,677,531
Tweets with URLs	3,420,228
Tweets with Geo-tag	62,629
Retweets	4,464,201
Replies	260,627
Time of the blast	Mon Apr 15 18:50 2013
Time of first tweet	Mon Apr 15 18:53 2013
Time of first image	Mon Apr 15 18:54 2013
Time of last tweet	Thu Apr 25 01:23 2013

This is one slide when I talked about, how the data can be represented, what data has been collected for doing these kinds of analysis.

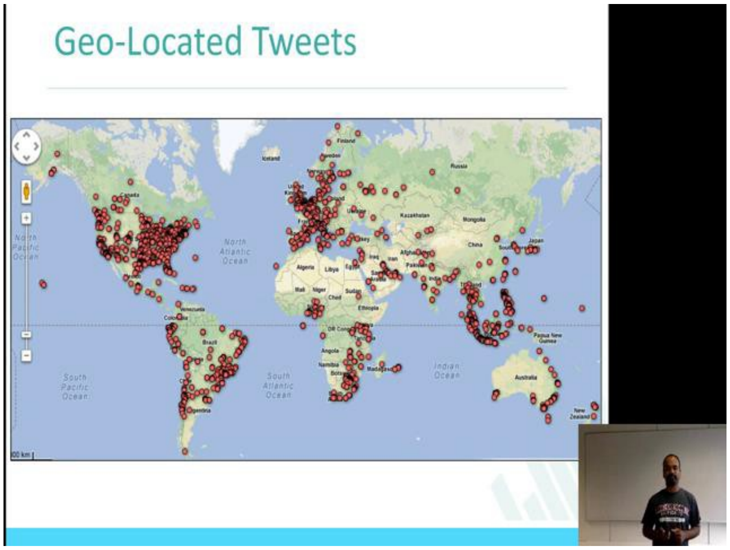
(ReferSlideTime:03:02)



Data Description

I also mentioned about the spikes in the social media data content that is generated on social media is very, very populated with the actually event that happens in the real world. This is one example where man hunt is over, a lot of people are talking about it and therefore there is spike in the tweets that are showing up.

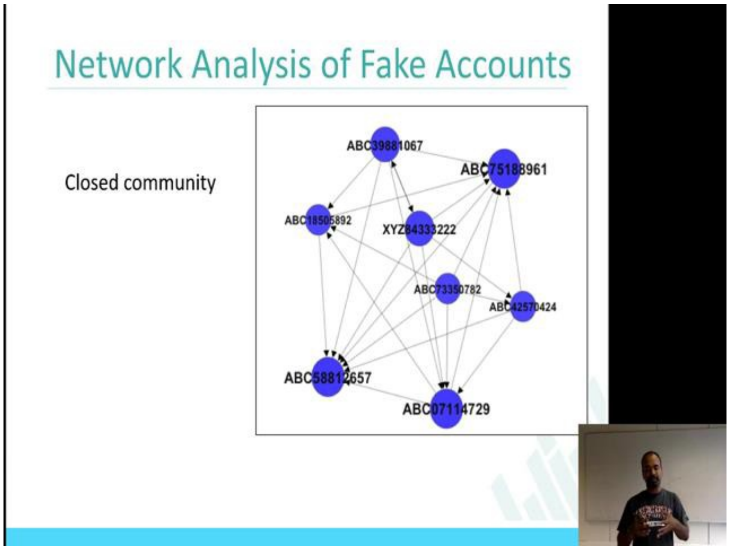
(ReferSlideTime:03:23)



Geo-Located Tweets

This is Geo-Located tweets where each dotis a tweet which has a geo tag attached withit and such kind ofgraphs can be helpful in saying where these tweets are coming from.

(ReferSlideTime:03:36)

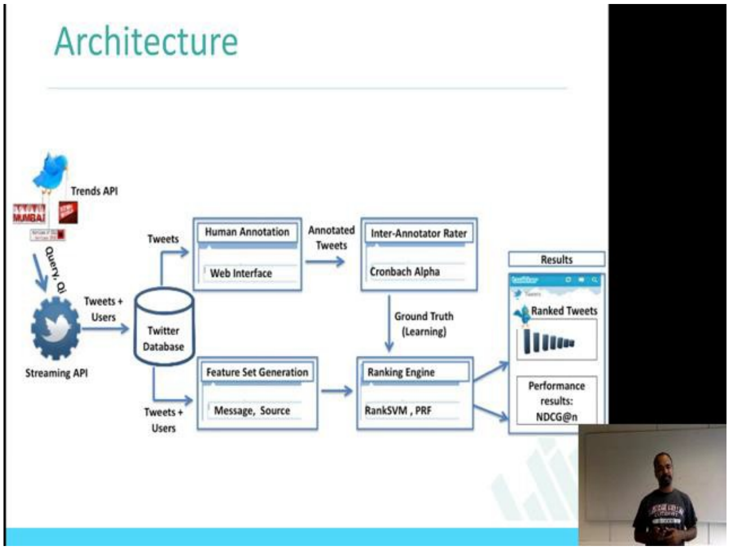


Network Analysis of Fake Accounts

Closed community

Wealso talkedabout howthecommunityofusers who arepostingthis fakecontent, who created fake accounts, how they are connected. Interestingly, they are all connected very closely and it is a closed community.

(ReferSlideTime:03:53)

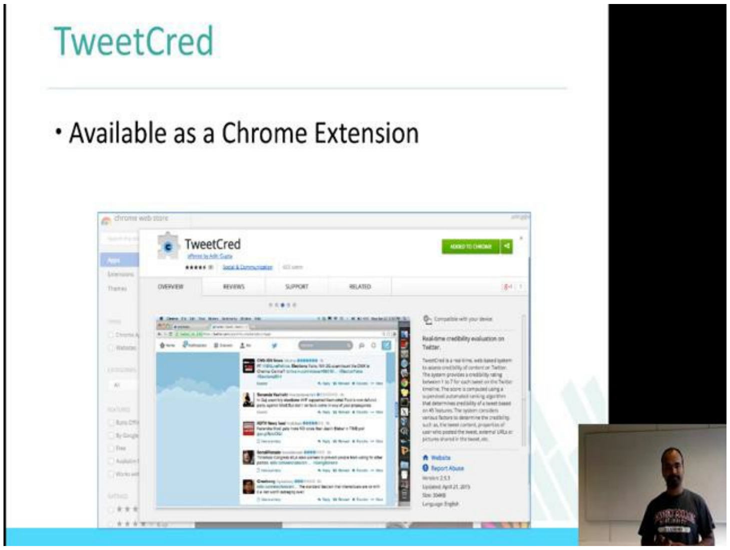


Architecture

And I also walked you through a multiple architecture diagrams mentioning about how data is collected from social media, what kind of annotations and how do you actually verify with the data that you annotated is actually of high quality and, what kind of feature generations can be done, what is the model that we developed and what is the model that one can develop, and now what are the evaluation matrix to actually find out whether the technique that we have applied and the model that we have created isactually good.

So, this is an architecture that I tell in detail talking about each block and explaining all this block helps in creating some interesting solutions for the problems in the trust and credibility space.

(ReferSlideTime:04:39)

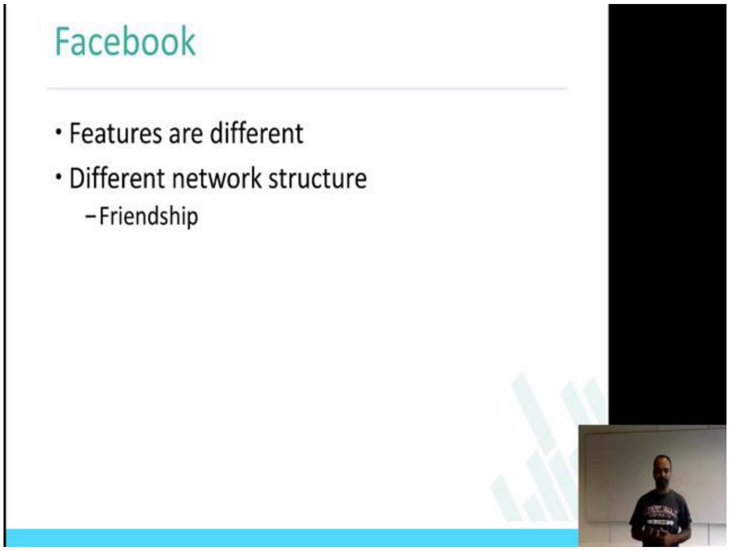


TweetCred

- Available as a Chrome Extension

And then I showed you about plugin which is called TweetCred, I hope some of youhave played around with the tweetcred plugin to find out how the tweets are evaluated and the value of x on 7 is presented to the users.

(ReferSlideTime:04:58)



Facebook

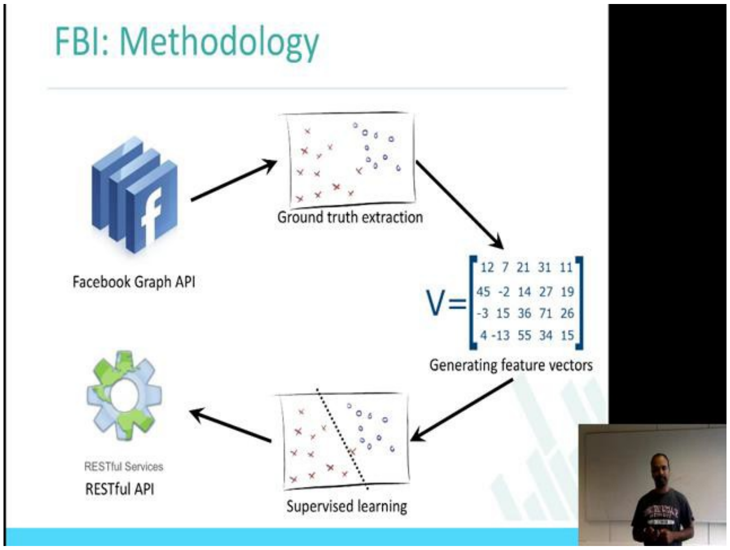
- Features are different
- Different network structure
  - –Friendship

So, now what I want to actually cover is little bit about how one can actually take this understanding of twitter and then apply into other social networks, because this is a privacy and security in online social media course so I thought that it will be interesting tofind outhowthesekindoftechniquesthatwelearntfromtwitter canbeacquiringinto other social networks particularly I will talk about Facebook. If you think about it initially I talked about how Facebook and twitter are different in my week 1 lecture, where we said that Facebook is a bi-directional network and twitter is a unidirectional network and the structure itself is very different.

And, the features that are available in these two social networks to study are also different. In twitter it is followers and following and Facebook, it is actually friends and the informationthat these networks provide through API are also verydifferent. And the structureofthe networks have different, particularlyI wanted to highlight this friendship thing in Facebook;theconnections are morepersonal and ifthere is apost that showsup by your friend, there is some tendency that it is more likely to be truthful and then weyou believe that your friend's post is actually more truthful than a random person's post.

So, that is the one of the differences between the Facebook and twitter network, particularly keeping this trust and credibility as the space of discussion. I wanted to highlight this difference, and now given this difference we should also look at how we can actually use the modelthat we have understood in twitter to apply it into Facebook.

(ReferSlideTime:06:37)



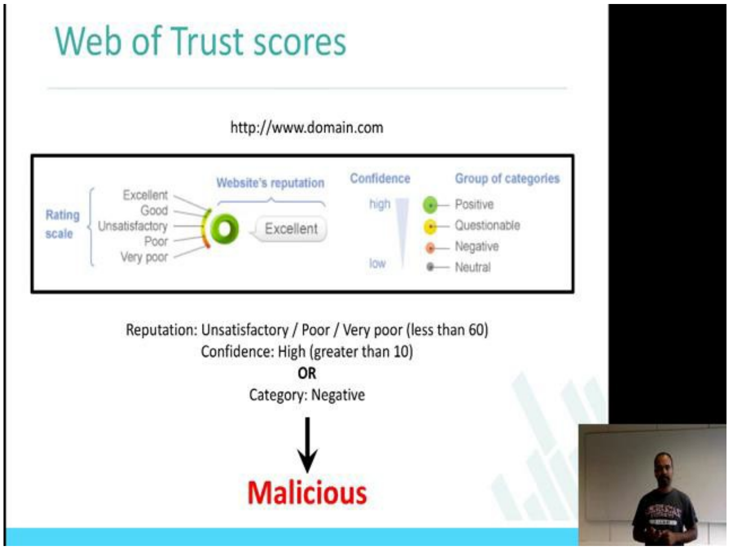
FBI: Methodology

$$V=\begin{bmatrix} 12 & 7 & 21 & 31 & 11 \\ 45 & -2 & 14 & 27 & 19 \\ -3 & 15 & 36 & 71 & 26 \\ 4 & -13 & 55 & 34 & 15 \end{bmatrix}$$

The architecture if you see it is almost the same, in this case it is just presented slightly differently. FBI: stands for Facebook Inspector, a similar tool that is like tweetcredwhich takesthe Facebook post fromFacebook graph API and then looksat the postsand make some judgement on how whether these post are malicious or not, credible or not, trust worthy are not.

In this case itis the same architecture which takes the post, do some feature extraction, do some ground truth understanding ofthe post, then creates some feature vectors out of it, create a model out of it, in this case supervised learning model because we actually have data from the posts that we are collecting and then create a RESTful API through which you can actually find out whether this post is malicious or not. Same architecture, very similar to tweetcred so I do not think we should actually spend a lot of time in understanding more details of this. If there is any question please feel free to ask in theforum for sure.

(ReferSlideTime:07:43)



Web of Trust scores

http://www.domain.com

Rating scale: Excellent, Good, Unsatisfactory, Poor, Very poor

Website's reputation: Excellent

Confidence: high, low

Group of categories: Positive, Questionable, Negative, Neutral

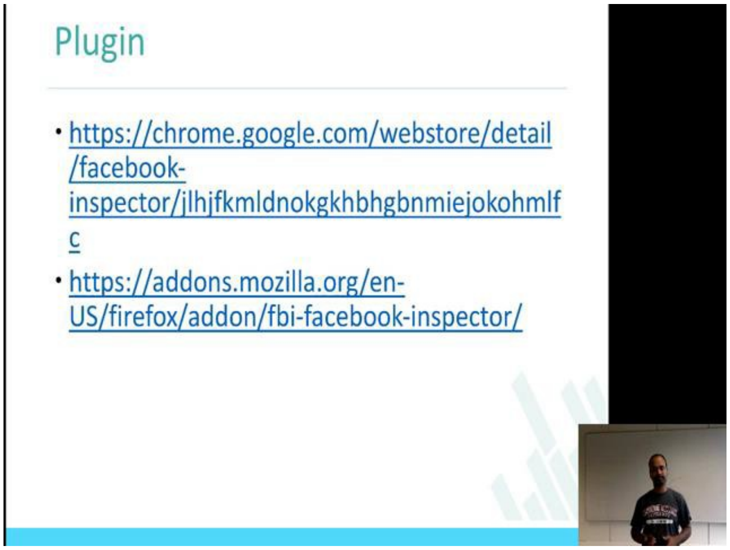
Reputation: Unsatisfactory / Poor / Very poor (less than 60)
Confidence: High (greater than 10)
OR
Category: Negative

Malicious

So, one thing that was also mentioned in the tweetcred or the twitter trust and credibility slides is a Web of Trust that is called WOT. Then I thought I would just mention it briefly what does it mean. It basicallytakes adomain and producesanoutput which says that a value similar to tweetcred, similar to other services that you may have seen where input is a domain and the output is score, which you can use to say that whether it is a malicious domain or not. Then in the past also I mentioned about how long the domain has been registered, who registered the domain and things like that. These features canbe used to make the judgement.

So web of trust basically gives you value of excellent, good, satisfactory, poor and very poor. If you give domain saying iiit dot edu dot in, it will actually come back with the rating scale and a confidence scale. We use this in Facebook inspector because in Facebook inspector it is also going to look at URL as the feature or particularly the domain as a feature from the post that we are analyzing.

(ReferSlideTime:08:57)

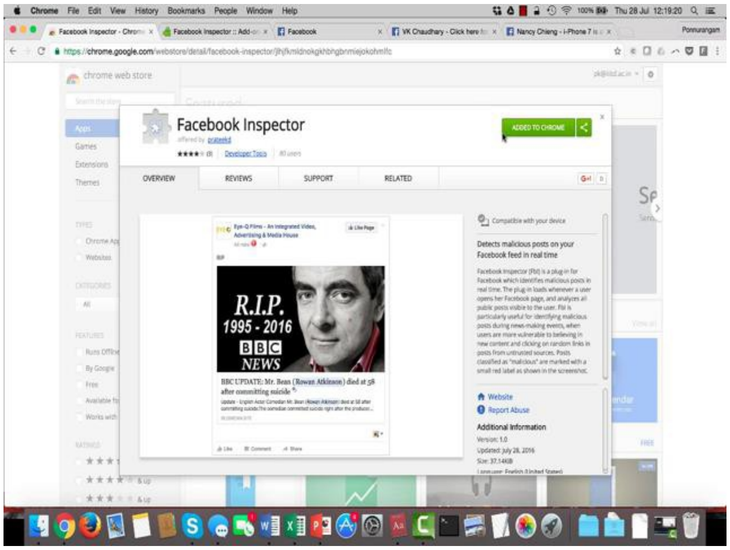


Plugin

- https://chrome.google.com/webstore/detail/facebook-inspector/jlhjfkmldnokgkhbhgbnmiejokohmlfc
- https://addons.mozilla.org/en-US/firefox/addon/fbi-facebook-inspector/

So, here is the pointer to the plugin. It will be interesting if you canactuallydownload it and play around with it. These are links to the Chrome extension and to the Firefox. Let me just walk you through how this plugin works, what does it do, how is it different or how is it very similar to tweetcred.

(ReferSlideTime:09:21)



Facebook Inspector

ADDED TO CHROME

OVERVIEW REVIEWS SUPPORT RELATED

Compatible with your device

Detects malicious posts on your Facebook feed in real time

Facebook Inspector (FbI) is a plug-in for Facebook which identifies malicious posts in real time. The plug-in loads whenever a user opens her Facebook page, and analyzes all public posts visible to the user. FbI is particularly useful for identifying malicious posts during news-making events, when users are more vulnerable to believing in new content and clicking on random links in posts from untrusted sources. Posts classified as "malicious" are marked with a small red label as shown in the screenshot.

Website

Report Abuse

Additional Information

Version: 1.0

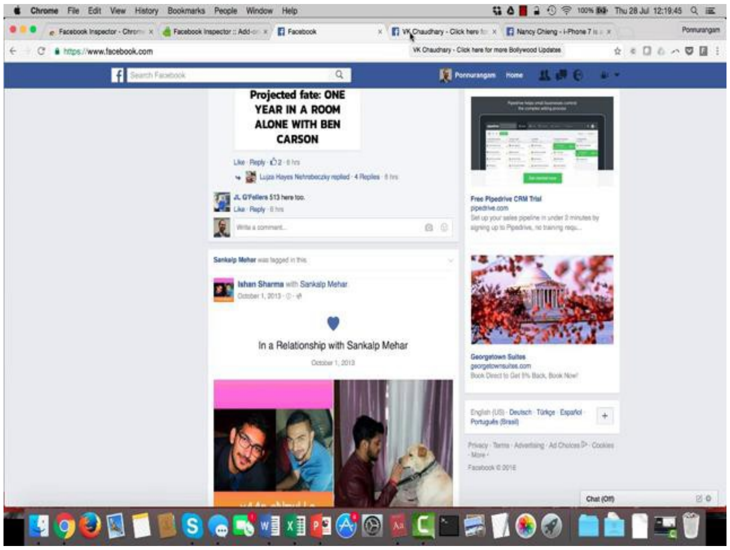
Updated: July 28, 2016

Size: 37.14KiB

R.I.P.
1995 - 2016
BBC NEWS
BBC UPDATE: Mr. Bean (Rowan Atkinson) died at 58 after committing suicide

This is Facebook inspector on the Chrome store, and basically you can add this to your browser, I meanI already have it on my Chrome otherwise it should say add to Chrome. When you add it, when you go to your Facebook timeline, you should be able to see some difference in the post that you are seeing.

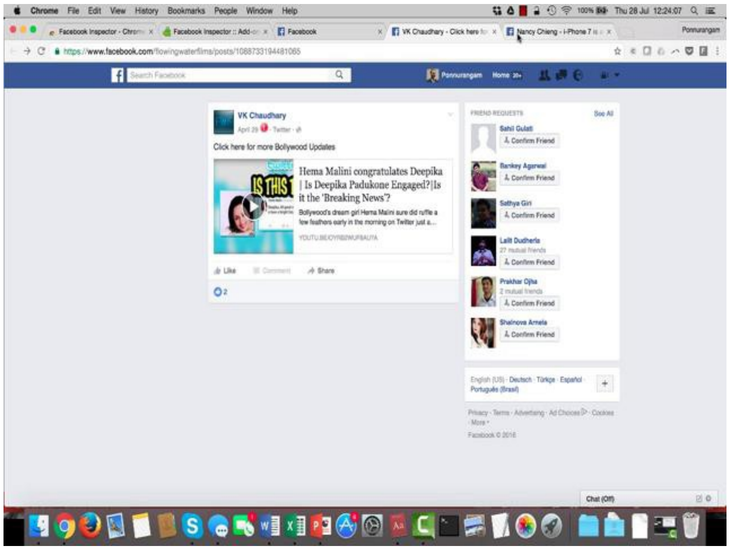
(ReferSlideTime:09:47)



Projected fate: ONE YEAR IN A ROOM ALONE WITH BEN CARSON
In a Relationship with Sankalp Mehar
Free Pipedrive CRM Trial
Georgetown Suites

For example here is my Facebook timeline for now and newsfeed, if you see there is no many,ifyoulook atthepost thereisnoannotationsdoneinthe poststhatyou canseeon my timeline. Whereas let me show you some examples where the Facebook inspector is actually showing you some information.

(ReferSlideTime:10:09)



VK Chaudhary

April 29 · Twitter ·

Click here for more Bollywood Updates

Hema Malini congratulates Deepika | Is Deepika Padukone Engaged?|Is it the 'Breaking News'?

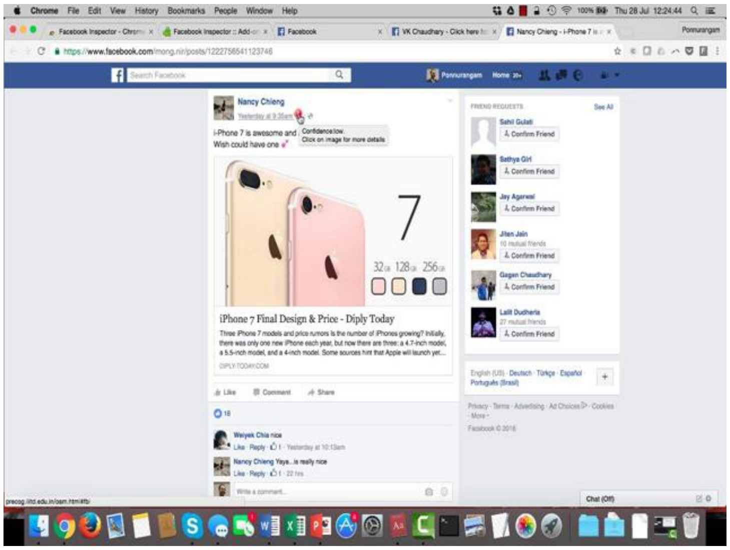
Bollywood's dream girl Hema Malini sure did ruffle a few feathers early in the morning on Twitter just a...

YOUTU.BE/OYRB2WUFBAUYA

Chrome File Edit View History Bookmarks People Window Help
Thu 28 Jul 12:24:07
Facebook Inspector - Chrome
Facebook Inspector :: Add-on
Facebook
VK Chaudhary - Click here for
Nancy Chieng - i-Phone 7 is
https://www.facebook.com/flowingwaterfilms/posts/1088733194481065
Search Facebook
Ponnurangam
Home 20+
IS THIS
Like
Comment
Share
2
FRIEND REQUESTS
See All
Sahil Gulati
Confirm Friend
Bankey Agarwal
Confirm Friend
Sathya Giri
Confirm Friend
Lalit Dudheria
27 mutual friends
Confirm Friend
Prakhar Ojha
2 mutual friends
Confirm Friend
Shalnova Arnela
Confirm Friend
English (US) · Deutsch · Türkçe · Español · Português (Brasil)
Privacy · Terms · Advertising · Ad Choices · Cookies · More
Facebook © 2016
Chat (Off)

Here is an example of a post which is done by VK Choudhary and the post we just click here for Bollywood updates. Hema Malini congratulates Deepika. Is Deepika Padukone engaged? In this post if you see, there is some annotation done by the Facebook inspector, it says confidence is low, the decision that was made with the model that was generated is low and it is using features, click on the image for more details. If you are interested, you can actually click on the image, see for more details. Here is another example that I will show you.

(ReferSlideTime:10:42)


Chrome File Edit View History Bookmarks People Window Help
Thu 28 Jul 12:24:44
Facebook Inspector - Chrome
Facebook Inspector :: Add-on
Facebook
VK Chaudhary - Click here
Nancy Chieng - i-Phone 7 is
https://www.facebook.com/mong.nir/posts/1222758541123746
Search Facebook
Ponnurangam
Home
Like
Comment
Share
FRIEND REQUESTS
See All
Confirm Friend
Chat (Off)

Nancy Chieng
Yesterday at 9:35am

i-Phone 7 is awesome and … Wish could have one

Confidence:low.
Click on image for more details

iPhone 7 Final Design & Price - Diply Today

Three iPhone 7 models and price rumors Is the number of iPhones growing? Initially, there was only one new iPhone each year, but now there are three: a 4.7-inch model, a 5.5-inch model, and a 4-inch model. Some sources hint that Apple will launch yet...

DIPLY-TODAY.COM

Weiyek Chia nice
Like · Reply · 1 · Yesterday at 10:13am

Nancy Chieng Yaya...is really nice
Like · Reply · 1 · 22 hrs

Write a comment...

Sahil Gulati
Sathya Giri
Jay Agarwal
Jiten Jain
10 mutual friends
Gagan Chaudhary
Lalit Dudheria
27 mutual friends

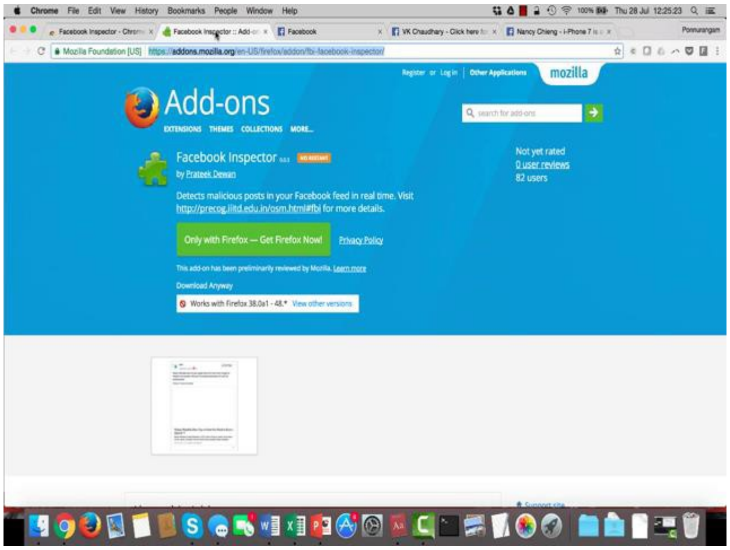
English (US) · Deutsch · Türkçe · Español
Português (Brasil)

Privacy · Terms · Advertising · Ad Choices · Cookies · More ·
Facebook © 2016

precog.iiitd.edu.in/osm.html#fbi

In the first case it is probably a rumour, that is why it is actually finding out and saying Facebook inspector is producing this result with red mark. Here is an example which could beaspam, whichisasofnowwereallydo notknowiPhone7designsand features that are available, but this post says about iPhone 7 is awesome and amazing whichcould have one. And this post is being annotated by Facebook inspector saying it is a malicious post. That is how Facebook inspector works.

(ReferSlideTime:11:20)


Chrome File Edit View History Bookmarks People Window Help
Thu 28 Jul 12:25:23
Facebook Inspector - Chrome
Facebook Inspector :: Add-on
Facebook
VK Chaudhary - Click here
Nancy Chieng - i-Phone 7 is
Ponnurangam
Mozilla Foundation [US]
https://addons.mozilla.org/en-US/firefox/addon/fbi-facebook-inspector/
Register or Log in
Other Applications
mozilla
Add-ons
EXTENSIONS
THEMES
COLLECTIONS
MORE...
search for add-ons
Only with Firefox — Get Firefox Now!
Privacy Policy

Facebook Inspector

by Prateek Dewan

Detects malicious posts in your Facebook feed in real time. Visit http://precog.iiitd.edu.in/osm.html#fbi for more details.

This add-on has been preliminarily reviewed by Mozilla. Learn more

Download Anyway

Works with Firefox 38.0a1 - 48.* View other versions

Not yet rated
0 user reviews
82 users

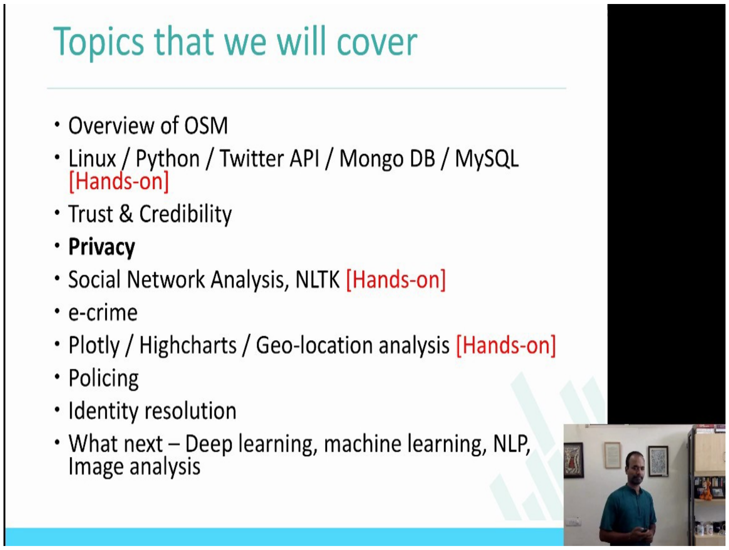
And here is also a plugin that you can use if you are a Firefox user. Therefore, if a Facebook inspector is available as a Chrome browser plugin and as a Firefox plugin add on which you can use on your browser.

So, that is the way you could think about taking way and understandings from twitter, where we studied about howto build techniques using the features from twitter to create an understanding of whether the post is a credible or not, here I showed you about Facebook.

PrivacyandPictureonOnlineSocialMedia

Welcome back to the course. I hope you are enjoying the course in terms of studying some new concepts, new ideas, and new solutions. This is the week 4 of the course Privacyand Securityin Online Social Media, what I will do now is continue the topic on privacy that we were talking last time.

(ReferSlideTime:00:32)

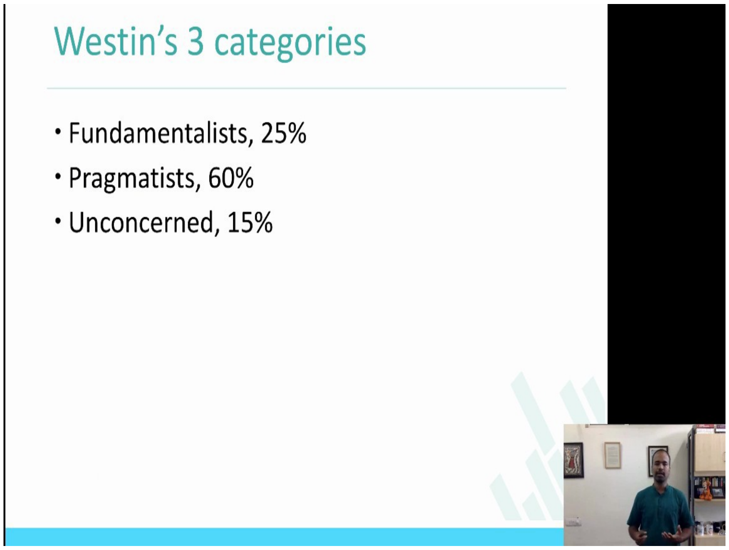


Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- Trust & Credibility
- **Privacy**
- Social Network Analysis, NLTK [Hands-on]
- e-crime
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Policing
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

Now,just let to let you know we are in the topic of privacy for now,we just covered the trust and credibility, and I assume by now you are all very well versed with little bit of Linux little bit of a Python, how to collect data from twitter, how to store the data, what kind of MySQL queries you should write and collecting data and all that.

(ReferSlideTime:00:54)

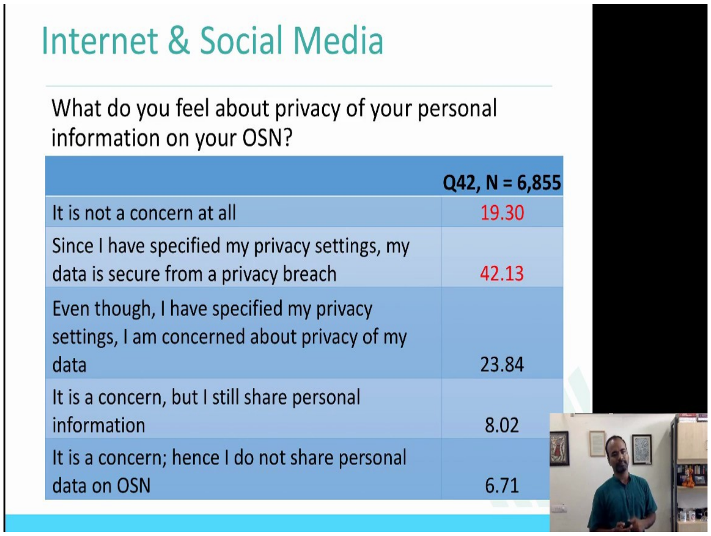


Westin's 3 categories

- Fundamentalists, 25%
- Pragmatists, 60%
- Unconcerned, 15%

In the last week we saw about how Westin categorized all the US citizens into 3 categories; Fundamentalist, Pragmatists and Unconcerned. Fundamentalist is being 25 percent, pragmatists is being 60 percent and unconcerned being 15 percent. Fundamentalist are the people who actually do not give away any personal information. Pragmatists makedecision aboutprivacykeeping the situation in mind. Unconcerned are the set of people who gave away personal information and be part of revealing personal information is about 15 percent in the US.

(ReferSlideTime:01:27)



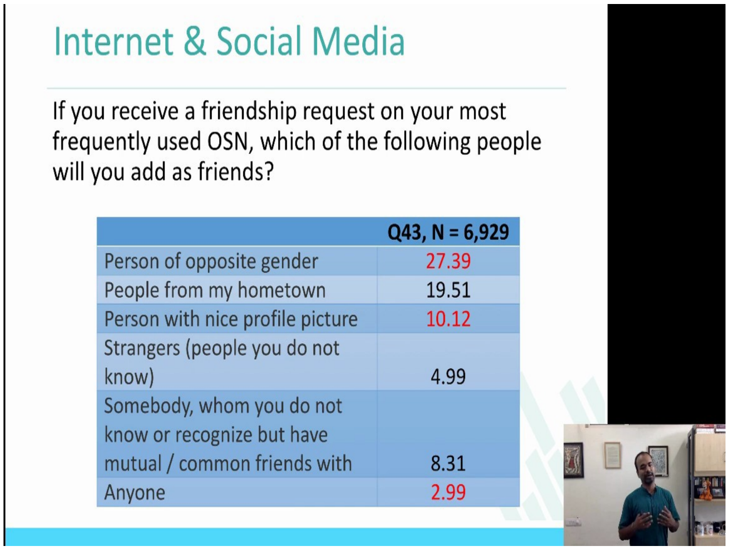
Internet & Social Media

What do you feel about privacy of your personal information on your OSN?

	Q42, N = 6,855
It is not a concern at all	19.30
Since I have specified my privacy settings, my data is secure from a privacy breach	42.13
Even though, I have specified my privacy settings, I am concerned about privacy of my data	23.84
It is a concern, but I still share personal information	8.02
It is a concern; hence I do not share personal data on OSN	6.71

I kind of asked you couple of questions last time about some data that was collected among large set of population in India. So this is one of the questions that I asked which is what you feel about privacy of your personal information on your online social network, which is about Facebook. About 42 percent, the highest was about 42 percent who said that specified my privacy settings my data is secured from a privacy breach.

(ReferSlideTime:02:00)



Internet & Social Media

If you receive a friendship request on your most frequently used OSN, which of the following people will you add as friends?

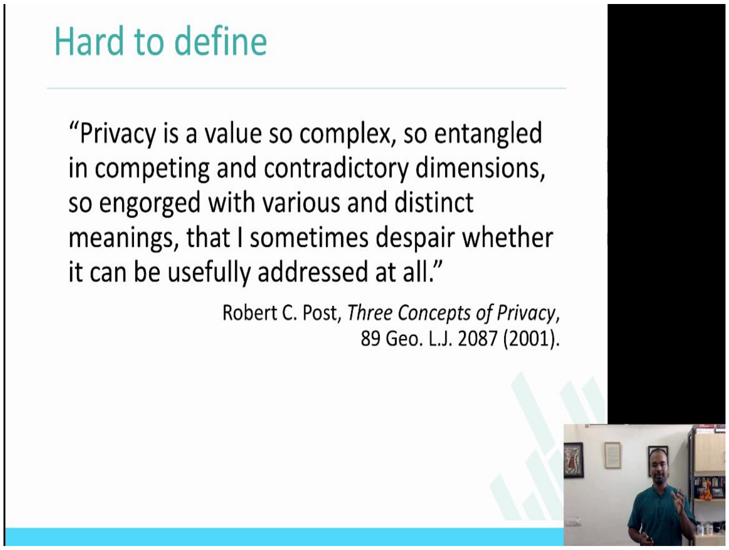
	Q43, N = 6,929
Person of opposite gender	27.39
People from my hometown	19.51
Person with nice profile picture	10.12
Strangers (people you do not know)	4.99
Somebody, whom you do not know or recognize but have mutual / common friends with	8.31
Anyone	2.99

Another question that I asked you also is about if you receive a friendship request on yourmostfrequentlyusedonlinesocialnetwork,whichisFacebookinthiscasewhichof the following people will you add as friends. And the highest was actually person of opposite gender.I am pretty sure in the last couple of weeks going through the class that youaretakingonthesocialnetworknow,even yourownbehaviormaybe changing,youmay be looking at some of these requests more closely, you may be devising your mechanism by which any friend request that you get, how you are going to accept it or how you are going to deny it.

(ReferSlideTime:02:35)

http://precog.iiitd.edu.in/research/privacyindia/

Now,thedataispubliclyavailablepleasefeelfreetoactuallyplayaroundwiththedata. (Refer Slide Time: 02:40)



Hard to define

"Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all."

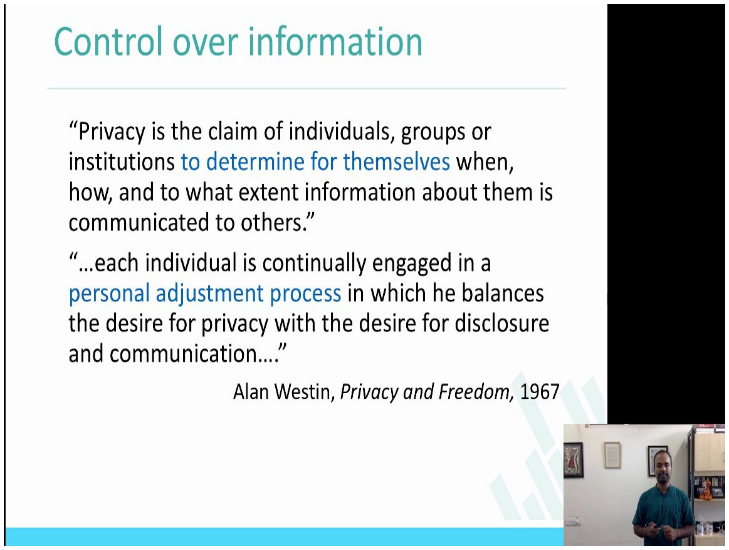
Robert C. Post, *Three Concepts of Privacy*, 89 Geo. L.J. 2087 (2001).

Last time I left you withthe question saying; what are the kind of privacyissues that you haveonFacebook,Twitter?Howyoudefineprivacy?Ithinkitisnicetoseesomeofyou posting information about various Facebook privacy issues or your own questions about Facebook privacy issues on the forum. We should actually make the forum more active because I think there are some very repeated questions that comes up, we're tying to answerassuchaspossiblebutwhentheyareveryrepeatedwecanavoidactually answering also. I strongly recommend you to ask, check the forum before posting the questions.

So, let us look at what privacy is a little bit and then give a little detail about some research that was goes down in terms of analyzing the privacy status on Facebook. One of the definitions that was given earlier about privacy was that "Privacy is a value so complex, so entangle in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all." So that was Robert talking about privacy in his book 'Three Concepts of Privacy.'

But I think the privacy by definitionsis actually thought. I mean, if you were to look at what privacy is for you, why are you sitting and listening to this lecture, versus privacyin your school, privacy at home, privacy at work is very different. It is very hard todefine what privacy is for a particular individual across various situations, that is what this definition is actually trying to capture. Contradictory dimensions, so entangled and competing and contradictory dimensions.

(ReferSlideTime:04:31)



Control over information

"Privacy is the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."

"…each individual is continually engaged in a personal adjustment process in which he balances the desire for privacy with the desire for disclosure and communication…."

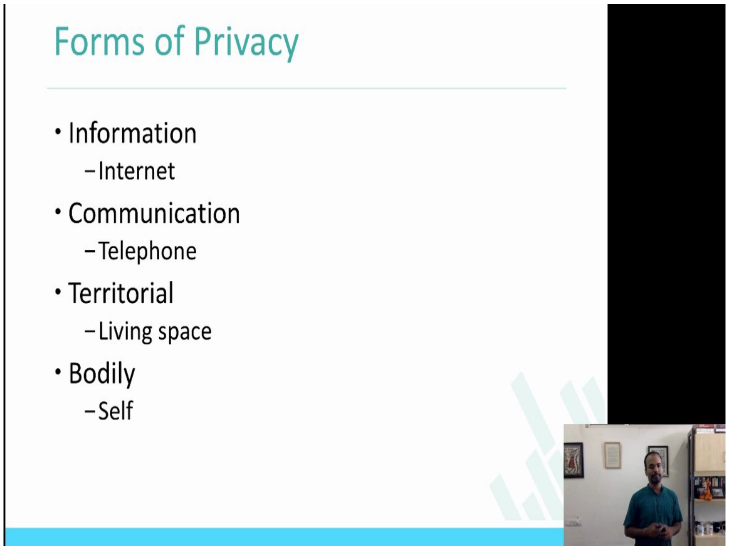
Alan Westin, *Privacy and Freedom*, 1967

Fundamentally privacy is been always talked about control over information, here are two definitions of Alan Westin actually tried defining in his book in a 'Privacy and Freedom'in1967."Privacyistheclaimofindividuals,groupsorinstitutionsto determine themselves when, how and what extent information about them is communicated to others."

So it is basically about to determine for themselves, how much of my information I can actually share with others. "Each individual is continually engaged in personal adjustment process in which he balances the desire for privacy with the desire for disclosure and communication." How much do I want to reveal about myself, how much doIwanttoactuallyanonymizeinformationaboutmyself,howmuchdoIwanttoreveal about myself, is the way that the word privacy is defined and is the way by which youare controlling the information that you are actually spreading.

So, I am sure you kind of get the definition privacy which is very hard to define and also it is very difficult to actually come up with the list of privacy expectations for any individual in all given contexts. They strictly convey privacy is about control over information. It sometimes could be actually a group information also, given that idea is more or collective society we generally talk about a privacy of a group, instead of individual privacy, that the society is where its individualistic society where the privacy information of the individuals are more protected than the privacy information of the group.

(ReferSlideTime:06:11)



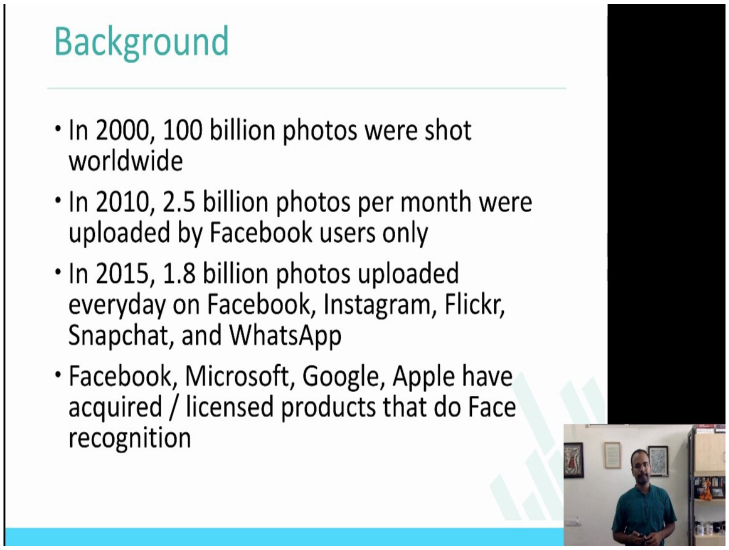
Forms of Privacy

- Information
  - Internet
- Communication
  - Telephone
- Territorial
  - Living space
- Bodily
  - Self

Some forms of privacy that people have come up with; information privacy, communicationprivacy,territorialprivacyandbodilyprivacy.Majorityofthetimes when we talk about privacy particularly in courses like these it is always referred to as information privacy and particularly the internet privacy.

There is also communication privacy which is telephones and other forms of communication. Territorial privacy is about my living space, my home, my city, my country and, the topics around that. Bodily privacy is about self. So, information about my own physical presence is actually also discussed in the concept of privacy. For example, a CCTVcamera is one example where bodily privacy can be actually attacked.

(ReferSlideTime:07:04)



Background

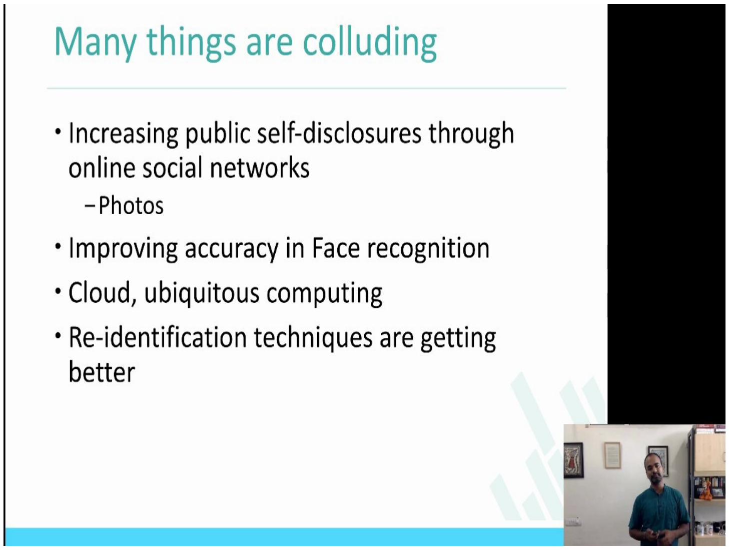
- In 2000, 100 billion photos were shot worldwide
- In 2010, 2.5 billion photos per month were uploaded by Facebook users only
- In 2015, 1.8 billion photos uploaded everyday on Facebook, Instagram, Flickr, Snapchat, and WhatsApp
- Facebook, Microsoft, Google, Apple have acquired / licensed products that do Face recognition

Now let us look at some specific studies that are being done in terms of analyzing the privacy in online social networks. Here is the study that I will walk you through the referencetothe studyis attheendofthedayoftheslides,butwewalkyou throughwhat theydid, what they find, how revealing the information are, how good the studywas and how the privacy is being actually studied in the context of Facebook and social networks and publicly available information.

Somebackgroundaboutpicturesthatwereuploadedonsocialnetworksitself.Intheyear 2000, 100 billion photos were shot worldwide. In 2010, 2.5 billion photos per month were uploaded by Facebook users only. Whereas, if you remember the first lecture 1 where I actually showed you a infographic about what among the information is uploadedonsocialnetworksin1minute,weactuallysawthat1.8billionphotoswere uploaded everyday on Facebook, Instagram, Flickr Snapchat, and Whatsapp together.Sothere is a lot of information, lot of pictures that are actuallyuploaded on social networks.

Companies like Facebook, Microsoft, Google,Apple have actually acquired a lot of face recognition companies in the last few years, to study, to understand, to use these technologies to identify faces on pictures that are being uploaded on the all social networks or online services. It has become very, very important to apply these kind techniques like, machine learning, deep learning and concepts around that into these images to study what is happening on online social networks, I actually recently wrote also a blog about the importance of images on online social networks. I'll actuallyshared it on the forum just after this lecture.

(ReferSlideTime:09:07).



Many things are colluding

- Increasing public self-disclosures through online social networks
  - Photos
- Improving accuracy in Face recognition
- Cloud, ubiquitous computing
- Re-identification techniques are getting better

If you really look at what is going on currently in terms of these pictures that were uploaded and the privacy about individuals, increasing public self disclosures through onlinesocialnetworkshappen,whichisItakeapictures,Itakeaselfiestandingnearone of the very important spots let us take in Delhi I upload this picture you know that I amin Delhi,orletustakeapicturenexttoTaj Mahalandupload itonmyFacebookaccount you know that I am actually traveling to Taj Mahal now.

There used to be actually a site called please rob me dot com I do not think so thewebsite is active now. This website what did they did was its called please rob me dot com, what we interestingly did was let us take it if I have a twitter account and I created it from Delhi and posting about weather in Chennai or Hyderabad or California they would actually pick this tweet and post it on please rob me dot com saying that this account was originally created from Delhi and whereas now this post is actually talking about weather in California, so probably you are not at home and therefore your homes should be locked.

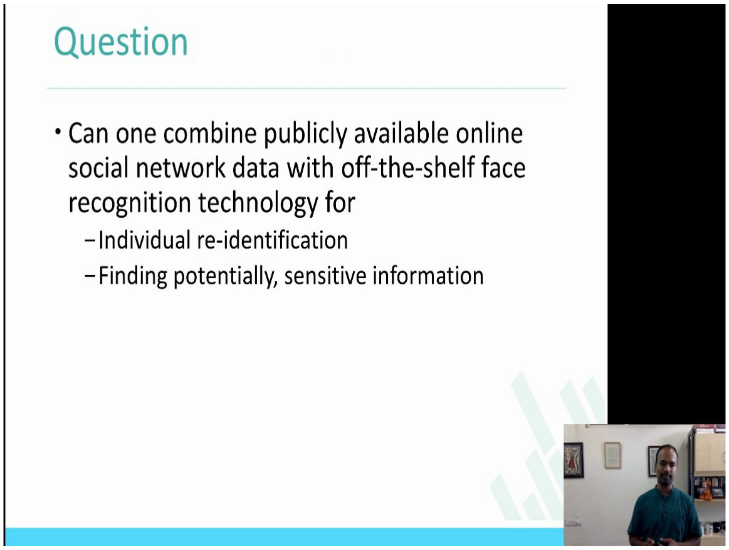
It got flacked a lot, but I think it is an interesting idea that they actually picked up to makeuseoftheinformationthattheusersofsocialnetworkaredisclosingbythemselves about their location. As a self-disclosure through online social networks and there are many manyissues that are going all around because of self-disclosure of information on Twitter, Facebook, Instagram and other networks.

Parallely in one side this increase in public information is going on. In parallel there is also increase in face recognition accuracy. In earlier the accuracy which lower now the techniques, technologies that are actually improved. In particular if you look at networks like Facebook it is actually pretty high it is because they search space that they have to search for a particular face in the picture that you are uploading is actually only your friends, majorityof the times you're going to be taking pictures with thefriends to whom you are already are connected with or probably they are in a one, and one and half hour or two hours away from here.

So, that is happening on one side. And also this is whole idea of cloud, storing information on the cloud, easily able to compute, computing cost is becoming lower and lower for doing any of these analysis. On the fourth dimension, problem is that identification of this users, who they are, what kind of information they are valuing is also getting better. Meaning, the concepts like k-anonymity came in 15 or 20 years before, but certain many further and advance techniques that have been developed to identify users, to identify faces, to identify information about users, to re-identify people on social network, people on other networks.

Those are four different things that are eluding; one, increasing self-disclosure, improving the accuracy of face recognition techniques, the whole idea of cloud and ubiquitous computing, and the techniques for re-identification of users is actually getting better and better.

(ReferSlideTime:12:33)



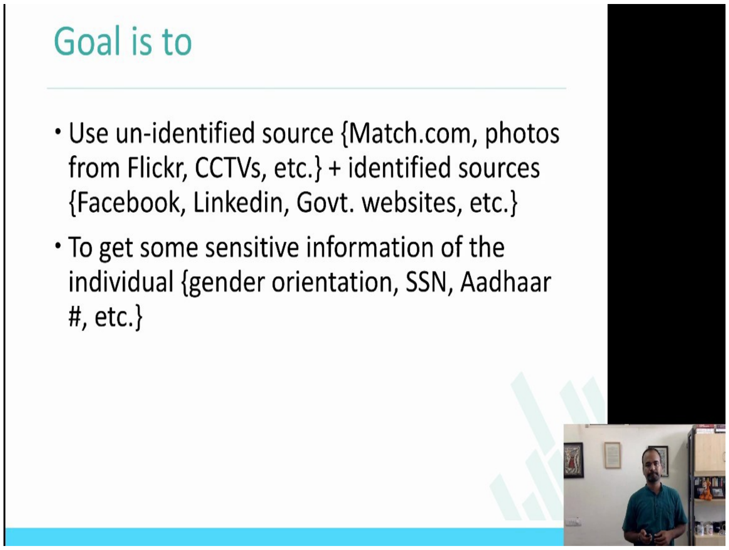
Question

- Can one combine publicly available online social network data with off-the-shelf face recognition technology for
  - Individual re-identification
  - Finding potentially, sensitive information

Theoneimportantquestionandoneinterestingquestionthatpeoplecouldaskis,canone combine publicly available online social network data with the off the shelf face recognition technology which is something that is already available, and be able to re-identifying individuals and finding potentially sensitive information. So that is the question thatwewereetalkingaboutinthe nextdeckofslides which is,canwetakesome publiclyavailableinformationwhichisthatthethingsthatIhaduploadonFacebook,the things that I had upload on Twitter.

Can you use that and connect it with the off the shelf face recognition technology which is some tools like tensorflow that I will also mention later in the slides. Use these techniques to identify just basis and be able to actually re-identify the person and or also find out sensitive information about the users themselves. That is the question that we will be talking about right now.

(ReferSlideTime:13:35)



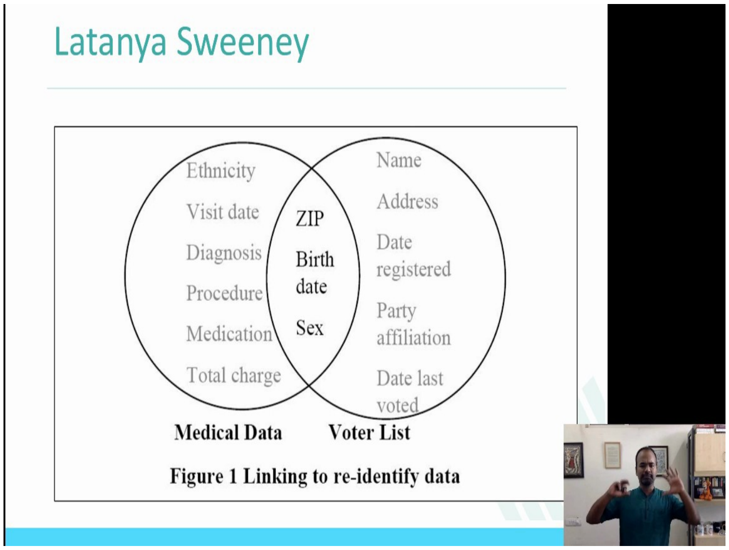
Goal is to

- Use un-identified source {Match.com, photos from Flickr, CCTVs, etc.} + identified sources {Facebook, Linkedin, Govt. websites, etc.}
- To get some sensitive information of the individual {gender orientation, SSN, Aadhaar #, etc.}

Here is a goal. Goal is to use un-identified sources which is any websites that you canthinkof,matchdotcom,shaadidotcom,photosfromFlickr,CCTVfeedsandthingslike that, which is impossible to identify or its very hard, the user themselves are not disclosing who they are in these websites. It could be either they have psuedonyms and names that you cannot identify or re-identify to that particular person. Can we actually take these sources, shaadi dot com and pictures from Flickr and Facebook, connected to identify sources which are on Facebook, I would actually reveal that I am so and so on.

On Linkedin I will put this as I am so and so, on government website and other services that are available. Which is un-identified sources like, shaadi dot com, identified sources which is where I am disclosing that I am so and so, and I upload a picture my account is actually ponnurangam.kumaraguru, can we actually put these two together to get some sensitive information of the individual. For example, gender orientation like example SocialSecurityNumber,likeexampleAdhaar card number and theinformation like that. It can be pretty nasty if you can actually put this together and the get some personal information. So that is what we will be studying in our next slots.

(ReferSlideTime:15:00)



Latanya Sweeney

Ethnicity
Visit date
Diagnosis
Procedure
Medication
Total charge

ZIP
Birth date
Sex

Name
Address
Date registered
Party affiliation
Date last voted

Medical Data

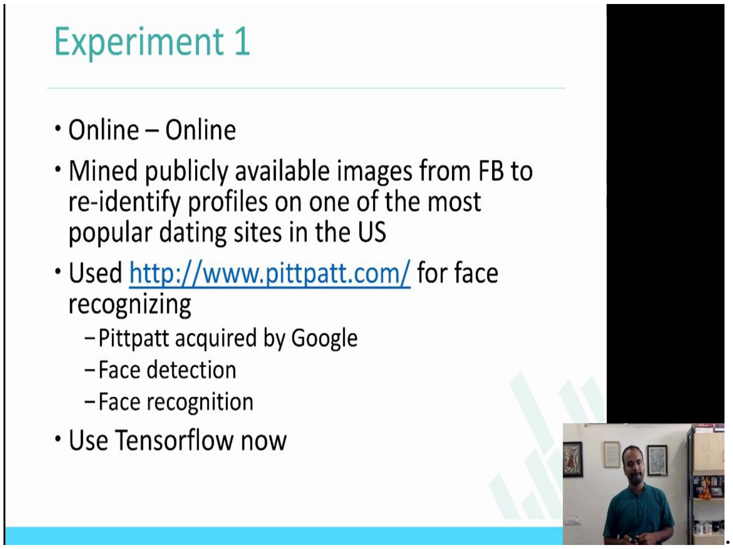
Voter List

Figure 1 Linking to re-identify data

Justto giveyou someverybroad old view ofsome phenomenonalworkthatwas donein this topic Latanya Sweeney, who did this word called k-anonymity, where she actually picked up the medical data and connected to the voter list which is publicly available. If you look at the medical data she has ethnicity, visit date, diagnosis, procedure, medication and the total charges that was paid by the patient. Name, address, date registered, party of affiliation, date last voted. Taking this information which is from voters list and from the medical data putting it together she had found actually zip code, birth date and gender was actually common among both of them.

She was able to identify if you give the system that she built birth day and gender she was able to re-identify a lot of US citizens uniquely. So that is the idea that built on to create something called as k anonymity, but the problem she highlighted was that bringing these two different sets of data which is independent medical data and voter data, you could actually re-identify users uniquely.

(ReferSlideTime:16:18)

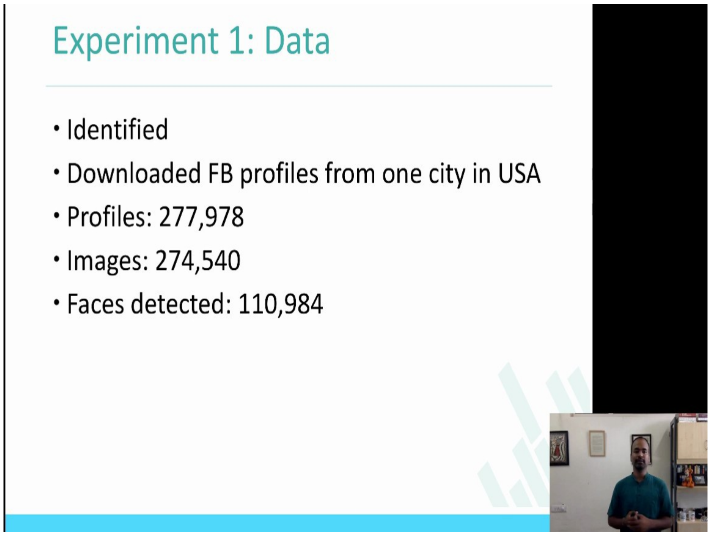


Experiment 1

- Online – Online
- Mined publicly available images from FB to re-identify profiles on one of the most popular dating sites in the US
- Used http://www.pittpatt.com/ for face recognizing
  - Pittpatt acquired by Google
  - Face detection
  - Face recognition
- Use Tensorflow now

In experiment one, they actually connected the online data to the online data. Theyinterestingly mined publicly available images from Facebook and they going to re-identify profiles just on one of the most popular dating sites in the US. They used this tool called pittpatt dot com, which was face recognizing tool. Well, after the study was done the tool was actually acquired by Google it is doing face detection and face recognition.Youcould actuallyuseTensorflownow.Tensorflowisaopen source library for machine learning techniques. Please consider exploring tensorflow little bit and how it works and what are the libraries that are available inside tensor flow.

(ReferSlideTime:17:06)

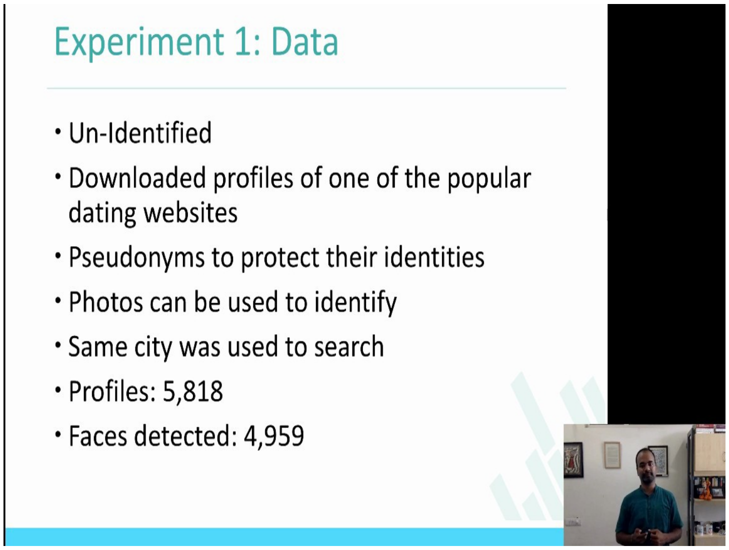


Experiment 1: Data

- Identified
- Downloaded FB profiles from one city in USA
- Profiles: 277,978
- Images: 274,540
- Faces detected: 110,984

The data that theyused was first as I said; they took the identified data, theydownloaded the Facebook profiles from one city in the US which is possible in the way that youknow about Facebook data collection now you could actually collect data from a particular city. Profiles that they collected were about 270,000, images that were collected around 274,000. The faces that are detected were about 110,000 faces. This is the data that they had for the identified data set, which is where you could actually say these are the names; these are profiles that are connected to these pictures.

(ReferSlideTime:17:50)



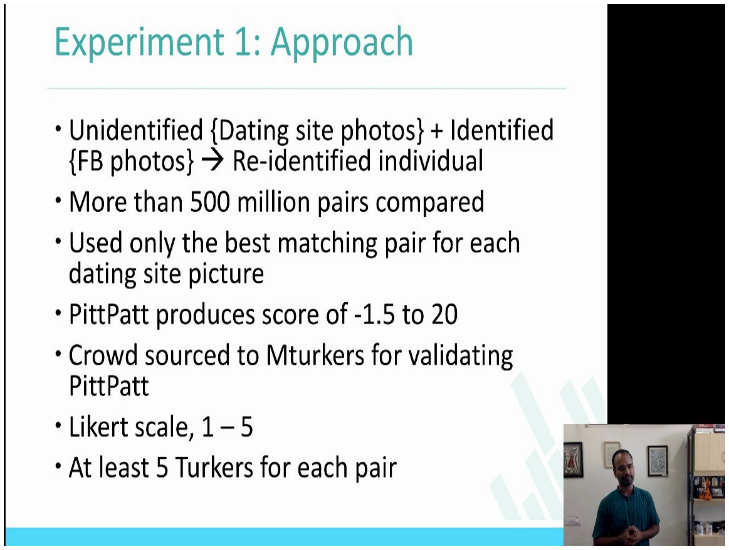
Experiment 1: Data

- Un-Identified
- Downloaded profiles of one of the popular dating websites
- Pseudonyms to protect their identities
- Photos can be used to identify
- Same city was used to search
- Profiles: 5,818
- Faces detected: 4,959

Un-identified data, they downloaded the pictures of one of the popular dating websites. So first identified, take a back; the first is the identified data, now we are talking about un-identified data, which is like the CCTV camera, publicly available information or from match dot com, shaadi dot com. Theydownloaded the profiles and the pseudonyms of their,to protect their identities, of course the names were notgoing to be revealed, the accounts may actually have pseudonyms also.

Thephotosthatweredownloadedfromthesewebsiteswhereactuallyusedtoidentifythe profile. To make the connection appropriate they actually use the same city for the search, they download data from Facebook and the city from this un-identified data set. The profiles that were collected here were about close to 6000 and the faces that were detected were about closed to 5000. So that is identified and that is un-identified data.

(ReferSlideTime:18:51)



Experiment 1: Approach

- Unidentified {Dating site photos} + Identified {FB photos} → Re-identified individual
- More than 500 million pairs compared
- Used only the best matching pair for each dating site picture
- PittPatt produces score of -1.5 to 20
- Crowd sourced to Mturkers for validating PittPatt
- Likert scale, 1 – 5
- At least 5 Turkers for each pair

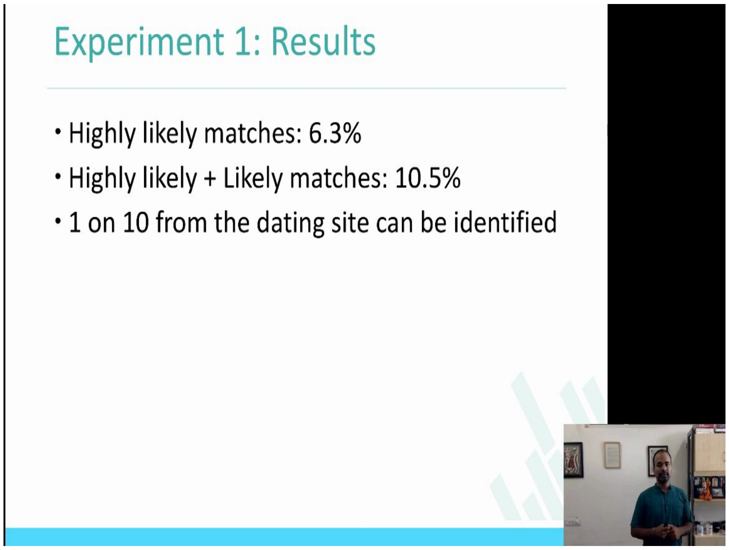
The approach that was taken was un-identified data, dating website, identified data, Facebook profiles and the re-identification was to be done. More than 500 million pairs were actually compared, because if each picture and each of the profile, each of the data set were compared with each of the pictures in the other data set, from the un-identified to the identified and the reverse also. What they did was, they did only used the best matching pairforeachofdatingsite picture, andpittpatt andIamsure intensorflow also itgivesyou inspecificvalues,itactuallyproducesvaluesinsomerangetheyusethebest value that they could get in terms of comparing two pictures.

And to confirm, to get ground truth when this pictures are just the same data sometimesif the techniques that are machine learning techniques are not going to be fool proof and they are not going to make 100 percent right prediction. Therefore, they are actually showed these pictures to Mturkers, the users who are part of mechanical turk which is a crowdsourced mechanismwhere you can actuallyputa small task of likethis identifying where these two pictures are same people and you could actually pay them small money for doing the task.

And there were asked to rate the pictures on the likert scale of 1 to 7, at least 5 Turkers for each pair.Again please tryandlook at what are MechanicalTurkers,mechanical turk is a crowdsourced mechanism. For example, if I were do a task in identifying whether a givenemailisphishingornotIwouldactuallyitshowtotheMturkers,Iwouldcreate the taskonmechanicalturk andget users toactuallylook atthe image andsaywhether it is phishing or not. Look at the profile and Twitter to say whether it is fake or not, they would actually go to the profile, they would click on the link in go to the profile in Twitter look at the profile and then make a judgment whether it is legitimate or not.

So it is the very popular and there are many many services like this, crowd flower which is mechanism in which many of these services come together, it is also very popular crowd flower is one - crowdflower, is one of the popular services like this - Mechanical turk which is from Amazon is also very popular. They took these two pictures showed to users, mechanical turkers asking to actually compare the images and make the decision. So, at least 5 Turkers for each pair because then we'll see more confidence, more and more people say that, more and more people take a image and say that this is the chair and there is high confidence that is going to be a chair.

(ReferSlideTime:21:37)



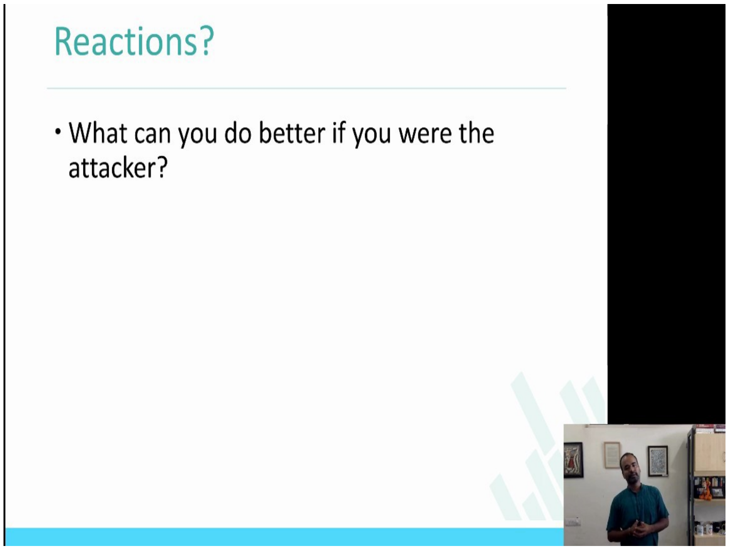
Experiment 1: Results

- Highly likely matches: 6.3%
- Highly likely + Likely matches: 10.5%
- 1 on 10 from the dating site can be identified

What they were able to find out was highly likely, which on the likert scale, is highly likely matches where about 6.3 percent that images that they took from this un-identified and identified and randomly they compared using the pittpatt tool and showed in the mechanical turkers. The comparison highly matches were about 6.3 percent and highly likely and likely matches were about 10.5 percent. Which basically says that 1 on 10 fromthedatingsitecanbeidentified,becausethedatingsite isanun-indentifieddataset, whereas Facebook is my identified.

So every time I see one of the pictures in the 10 pictures that I see, I will be able to actually clearly exactly identify who this person is, because I have the Facebook data, this is done of the same city and therefore it should be probably correct and mechanical turkers actuallyconfirmed that. So, you can seethat10 percentof thetimes theuserscan be actually identified.

(ReferSlideTime:22:40)

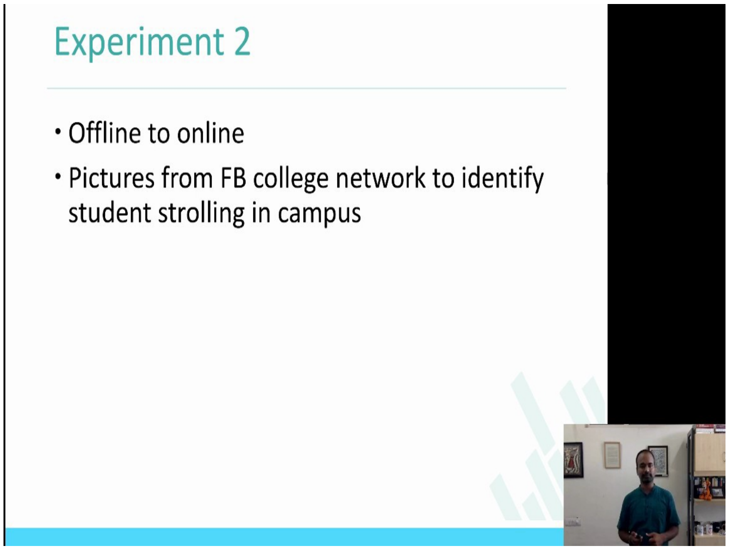


Reactions?

- What can you do better if you were the attacker?

One question to you and I hope this question since there will be some discussion inforum also is that; what can you do better if you were the attacker? And if you weremake use of this information and do something to increase the rate of the efficiency or use this information to do something against the user what kind of things would you do Because as an attacker you making one this percentage to be more right, because it is 10 percent you're getting a hit rate of only 10 percent, or 1 and 10 pictures. Whereas, if you were to have a better attack or threat mechanism you could actually do things by which you can increase this percentage to more, so more and more pictures are actually re- identified and therefore it can be actually used maliciously.

(ReferSlideTime:23:01)



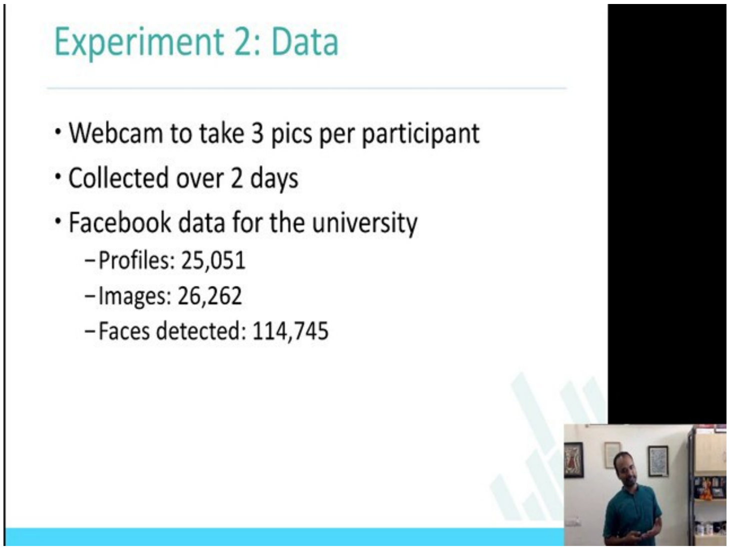
Experiment 2

- Offline to online
- Pictures from FB college network to identify student strolling in campus

Experiment 2 as I said there are 3 things. So the second one what they did was they connected the offline and the online. First one, they compare online versus online which is the dating website and Facebook, now what they did was they did the offline and the online.

Pictures from Facebook, one of the Facebook college network data was collected to identify students who are in campus and it was actually compared to the offline pictures also. What was stated when the students were actually participating in the study. So this is the experiment number 2; all connecting to the same questions which is can we actually take images, pictures from these social networks like Facebook and re-identify peoplewhoconnectedtonetworks,to datawhereuserscannot getinfrom,CCTVsource in.

(ReferSlideTime:24:21)

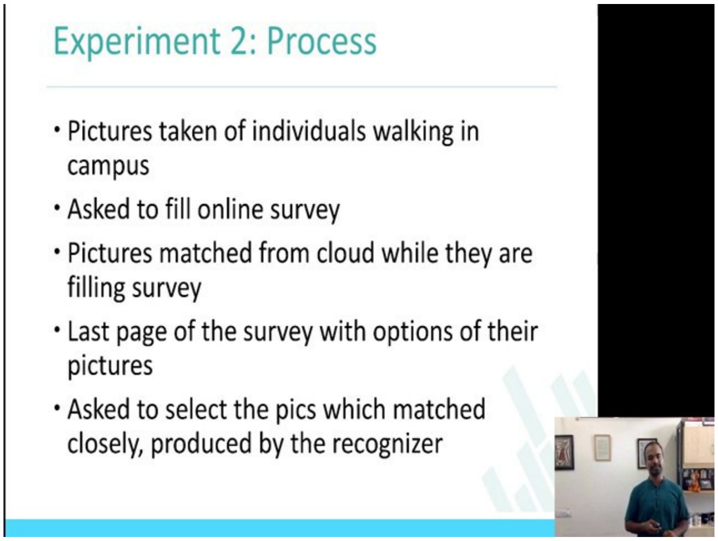


Experiment 2: Data

- Webcam to take 3 pics per participant
- Collected over 2 days
- Facebook data for the university
  - Profiles: 25,051
  - Images: 26,262
  - Faces detected: 114,745

So, what they did was they actually put a booth in the university, took 3 pictures of the participant, they basically were standing and collecting data of the college students inthis university took 3 pictures for participant, collected data over 3 days. They collected about 25 percent profiles, images were about 26,262 and the face is detected were about 114000, so Facebook data for that university. So, the data that were collected from Facebook whichisonlineisabout25000,profileswereabout25000,pictureswereabout 26000, faces were about 114000 thousand.

(ReferSlideTime:24:59)



Experiment 2: Process

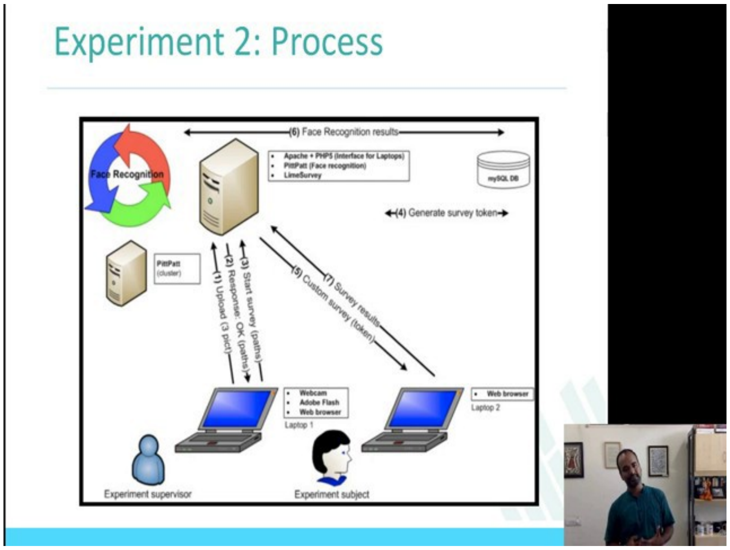
- Pictures taken of individuals walking in campus
- Asked to fill online survey
- Pictures matched from cloud while they are filling survey
- Last page of the survey with options of their pictures
- Asked to select the pics which matched closely, produced by the recognizer

Just to summarize or just to look at the whole experimental set up itself is that, pictures taken of individuals walking in campus, asked them to fill the survey. Next slide I also have a image to actually show you what was the process of the study. But now pictures were taken of the individuals walking on the campus, they were asked to fill an online survey.Pictures matched from cloud while they are filling the survey,because what they did was they ask that you want to participate in the study, ok I will take you 3 pictures, when they took the pictures then they asked into fill on online survey.

While they were actually filling the online survey, technique the system that they are acted would go compare this pictures what they are took to the Facebook pictures that they are already collected from the university itself and bring back the comparison and showedtothem.Lastpageofthesurveywithoptionsofthatpictures,sobythetimethey actually fill the survey they were actually shown the pictures, saying what this is the picture that we got from Facebook, do you actually agree to it. Asked to select the pics which matched closely, produce by the recognizer.

So, that is the process of the study,please understand how the study was done, collected pictures were taken individually walking in campus, they were asked to fill the survey, while filling the survey the data the system was comparing the pictures on Facebook, pictures were brought back to the survey showed to the user and saying tell us if these pictures are right about you.

(ReferSlideTime:26:28)

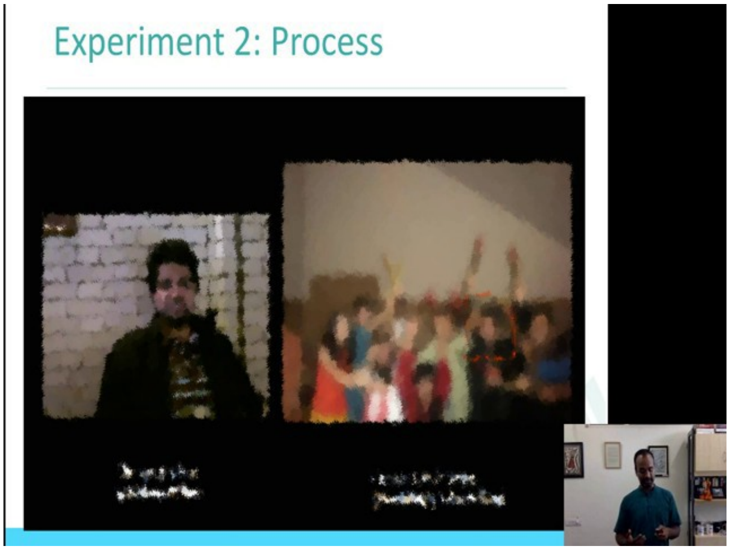


Experiment 2: Process

Same thing is captured here in the process format in the background, which is upload pictures of the users, pictures are taken which is 1 and then responses coming from the server, start survey which is 3 and then 4 is generated survey token, so that through this survey token you will actually be able to say that comparing the images and bringing it back, which is 5is looking at custom surveytokens sendto the user who can actuallyfill the survey. And then by the time of 6 is happening which is face recognition results are being produced and then survey results both the images that are actually used which is given to 7.

Sothatistheprocessofthestudy,notaverydifficult,notaverycomplicatedstudybutit is actually collecting some very interesting data.

(ReferSlideTime:27:22)

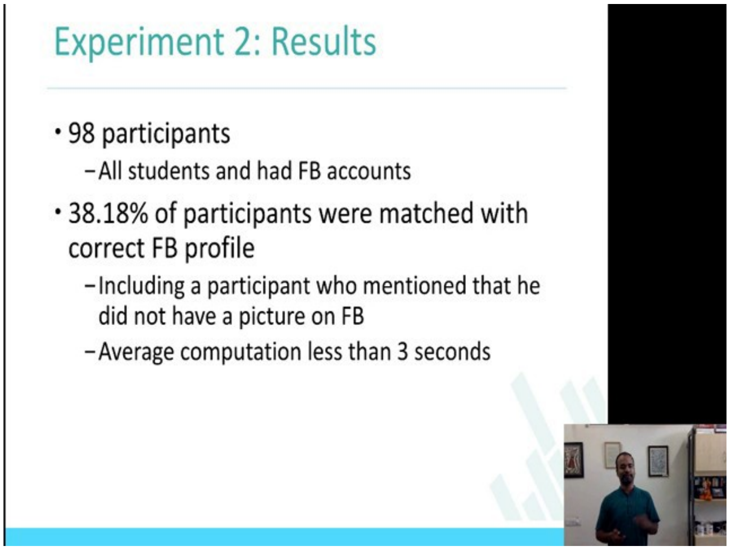


Experiment 2: Process

This is the result what they did from the data collection. The left picture is the picture autonomous to the picture for the purpose of just re-identification of the user itself. The pictureontheleftisthepicturethattheytook whiletheuser wasactuallyparticipatingin the study. So when the user logged in they took the picture that on the left.

Using that the picture they are able to actually identify the picture on the right which is the picture from Facebook where this user was actually identified. So, that is the output so to say, the input is the picture with the survey and output is the image from Facebook which is re-identified this person in particular pictures. This can be actually pretty revealing the pictures compared on Facebook.

(ReferSlideTime:28:06)



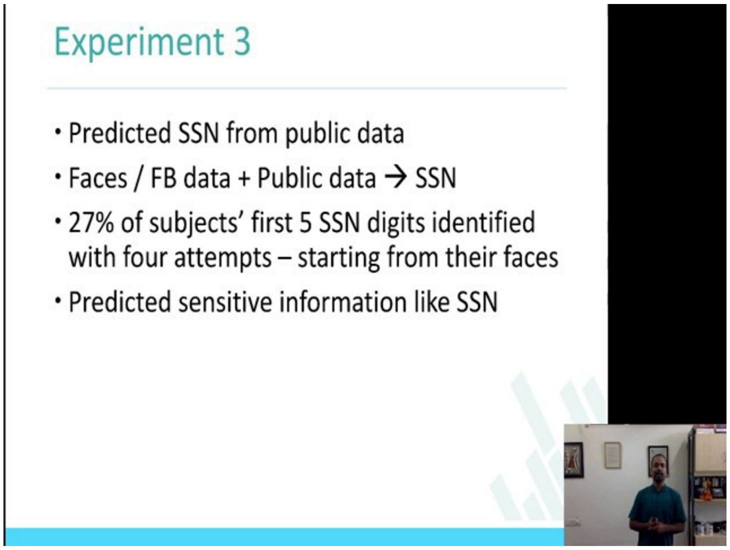
Experiment 2: Results

- 98 participants
  - All students and had FB accounts
- 38.18% of participants were matched with correct FB profile
  - Including a participant who mentioned that he did not have a picture on FB
  - Average computation less than 3 seconds

In about 98 participants all students in the study, there were about 98 participants, all students were the ones who participated they were collecting it from the university setup and they all had Facebook accounts also. The results were 38 percent of participantswere matched with correct Facebook profiles, which is the pictures that were taken, 38 percent of the people who took the pictures in the study were exactly matched with the Facebookprofileandtheiraccount,theirinformationisactuallybroughtbacktocompare to confirm it with the user.

Interestingly there was also a participant who mentioned that he did not have the picture onFacebook, actuallyinformationofthatparticular person,ofthatparticularparticipants was also brought back. Of course, it was actually taking very less time to do this comparison. I hope the study is making sense which is 38 percent of the times the users that were taken pictures from the university campus were identified from the Facebook profile.

(ReferSlideTime:29:14)



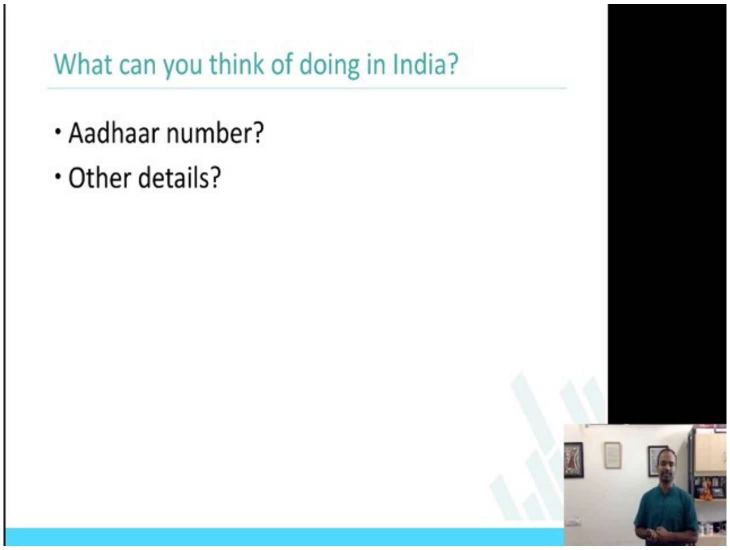
Experiment 3

- Predicted SSN from public data
- Faces / FB data + Public data → SSN
- 27% of subjects' first 5 SSN digits identified with four attempts – starting from their faces
- Predicted sensitive information like SSN

Experiment 3 is interesting because they actually tried using the experiment understandings from experiment 1 and 2 to take this personally identifiable with information likes Social Security Number. In this experiment 3 they wanted to predict Social Security Number from public data. So, they used the faces and the Facebook data that werecollected from the experiment 1 and2 with the public data to predict the Social Security Number. 27 percent of subjects' first 5 Social Security Number digits were identified with four attempts.

So essentially what is this means, this means that every time I took up a face from the database, I was able to identify the first 5 digits of the Social Security Number, 27 percent of the times. That revealing, that is not a very good sign, were 27 percent of the subjects were able to find out five SSN digits of them. So that is the third experiment. And I am keeping the third experiment little light because this is in total the interesting things were pictures, un-identified data sets, identified data set and at the end they were able actually do connected to social security member also.

(ReferSlideTime:30:37)

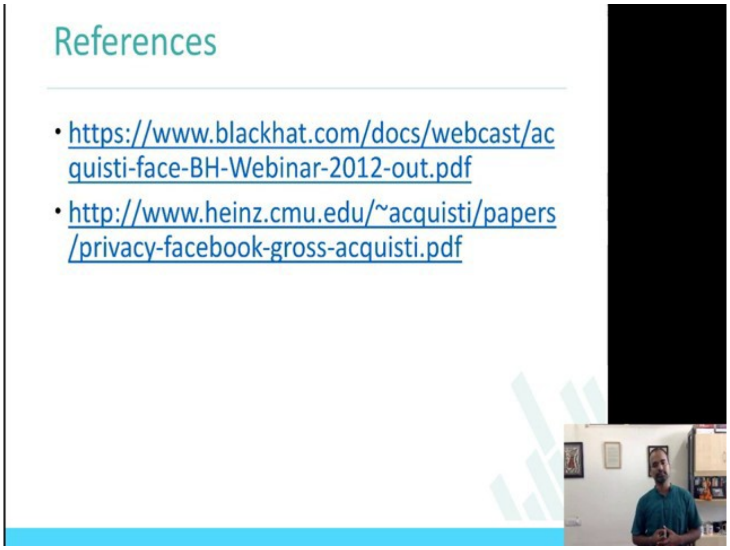


What can you think of doing in India?

- Aadhaar number?
- Other details?

Interestingly I am sure you could also think about how these kinds of techniques can be applied in terms of identifying Adhaar number in India also and other personal details. The study was done in the US and therefore if you were to repeat this study and find out Adhaar number or others details of Indian Citizens it will be actually interesting to look at that. If there is anyideas, if there is any questions that you have in terms of how study could be performed in India, it will be interesting to talk about it in the forum.

(ReferSlideTime:31:09)

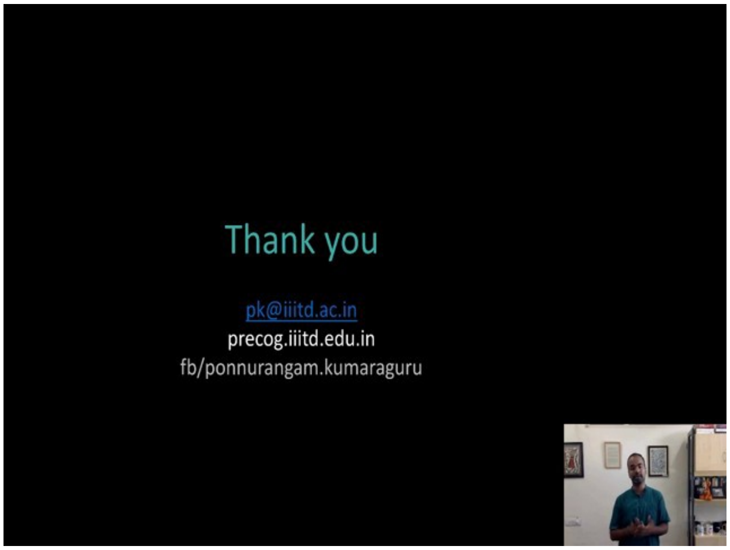


References

- https://www.blackhat.com/docs/webcast/acquisti-face-BH-Webinar-2012-out.pdf
- http://www.heinz.cmu.edu/~acquisti/papers/privacy-facebook-gross-acquisti.pdf

Hereare the pointerstostudythatIjust nowdiscussedabout.

(ReferSlideTime:31:17)



Thank you

pk@iiitd.ac.in
precog.iiitd.edu.in
fb/ponnurangam.kumaraguru

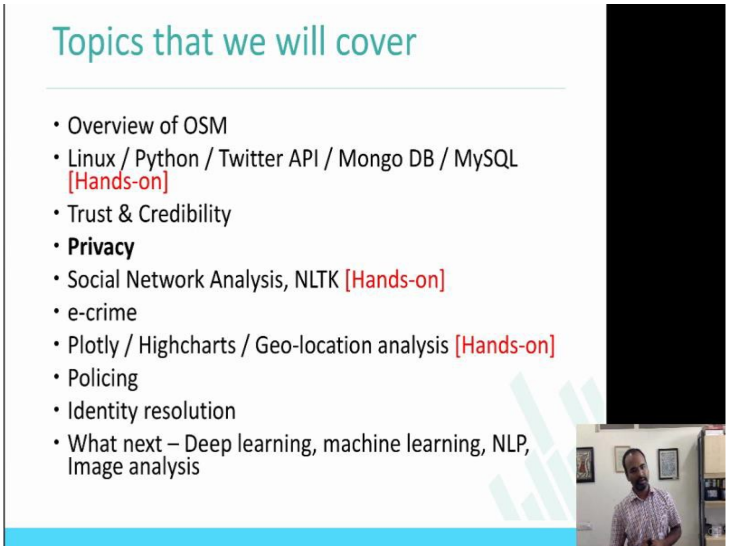
And with this I will actuallywrap-up the 4.1 week. I hope you understood what we were talking about, we just talking about the Privacy Issues in Online Social Networks particularly focused on collecting images and identifying users using the face, pictures, using the images that are uploaded on social networks.

Unit-3

Policing and Online Social Media Part-I

Welcome back tothe course Privacyand SecurityinOnline SocialMedia, this isweek 5. Hope you have enjoyed the course by doing the home works, doing some of theexercises that we are talking in the class and generally also getting excited about the course.

(ReferSlideTime:00:27)

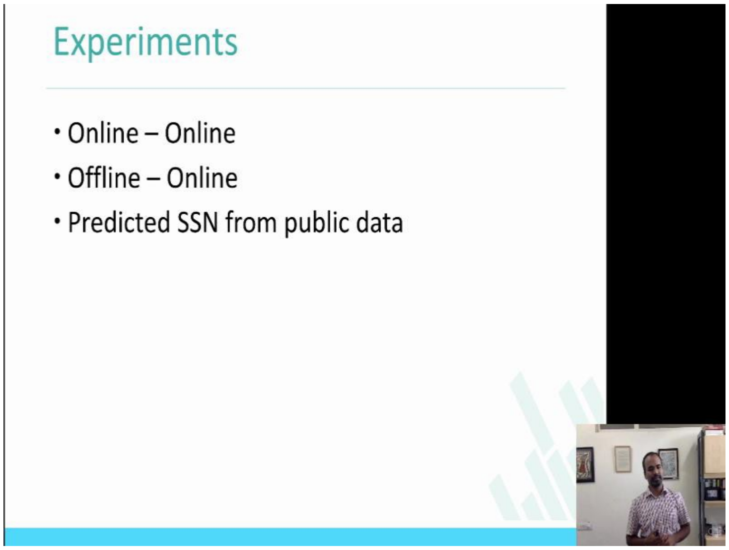


Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- Trust & Credibility
- **Privacy**
- Social Network Analysis, NLTK [Hands-on]
- e-crime
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Policing
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

Thisisthe generaloverview ofthe course,wearelookingat thetopicofprivacy now.

(ReferSlideTime:00:32)



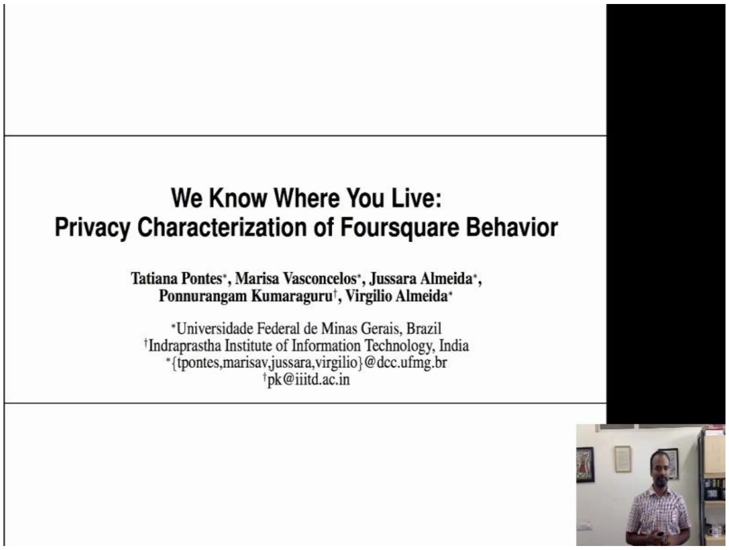
Experiments

- Online – Online
- Offline – Online
- Predicted SSN from public data

Last time intheweekwesawthecontent aboutprivacyand howprivacyimplicationsare going on Online Social Media, in particular we saw some experiments where they are done, an analysis of data from online and online, offline and online and sometimes looking at predicting the Social Security Number also. The online and online study that they did was to take up data from Shaadi dot com type of sites like Match dot com compare it to Facebookpictures, thensee whether theywere able to identifypeople from this matches. Offline and online, they took pictures of students walking around on campus and compared to their Facebook profile and then, said that this is your profileand then got some confirmation about the picturesthat theydownloaded fromFacebook.

Theyuse this publiclyavailable dataonline, offline, allthe datato predict actuallySocial Security Number and they were successful in predicting 1 in 10 percent of the data that theysaw with some confidence ofthe five digits ofSocial SecurityNumber. So,that is a kind of thing that we have seen until now.

(ReferSlideTime:01:46)



We Know Where You Live:
Privacy Characterization of Foursquare Behavior

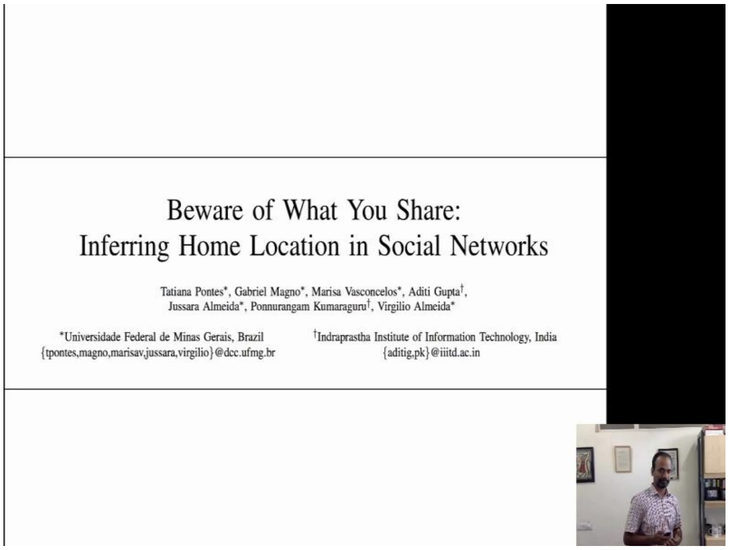
Tatiana Pontes[*], Marisa Vasconcelos[*], Jussara Almeida[*],
Ponnurangam Kumaraguru[†], Virgilio Almeida[*]

[*]Universidade Federal de Minas Gerais, Brazil
[†]Indraprastha Institute of Information Technology, India
[*]{tpontes,marisav,jussara,virgilio}@dcc.ufmg.br
[†]pk@iiitd.ac.in

Let me tell you few things more about privacy on Online Social Media and then wemove on to another topic this week. It is not only that pictures, and you can use these information from social media, but you can actually use something more specific information like a location also to find out where you are and where you have been moving around and things like that.

Here is one study that researchers have done to show that privacy information from Foursquare which is one of thelocation based social networks, one of the very popular location based social network. Data from Foursquare can be actually used to find out where you live.

(ReferSlideTime:02:29)



Beware of What You Share:
Inferring Home Location in Social Networks

Tatiana Pontes*, Gabriel Magno*, Marisa Vasconcelos*, Aditi Gupta†,
Jussara Almeida*, Ponnurangam Kumaraguru†, Virgilio Almeida*

*Universidade Federal de Minas Gerais, Brazil
{tpontes,magno,marisav,jussara,virgilio}@dcc.ufmg.br

†Indraprastha Institute of Information Technology, India
{aditig,pk}@iiitd.ac.in

These two research which wass done back to back is to inferring home location fromthe check insthat you do infoursquare. Iamnot going toget into detailsofstudythat iswhy I put thepointersto thepapers, but I willtalk you ingeneral howthiswasdone, howthis could be done, and how you can actually look at some of the data yourself also.

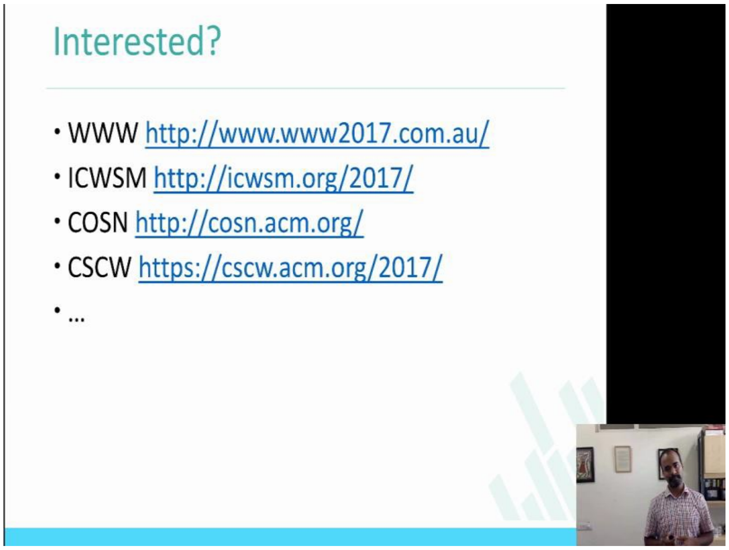
Foursquare is one of the popular online social networks, just for locations it is called location based social network. The different topics and different concepts in foursquare are check ins. Check ins is, you check in to the hotel, you check in to the airports and similarly you check in to a location in foursquare and you can actually alsoleave a tip,let us take if I go to Sarvana Bhavan, Connaught Place in Delhi. I have food there and I can leave a tip saying that food was pretty good. And you can also become a mayor in foursquare which is, if I visit this place, if I visit this location in foursquare the most number of times in the last 60 days, I become the mayor of this location.

The mayor informationcanbeactuallyprettyuseful.Todayorganizationsare monetizing this check ins and mayorship in foursquare.Also someone actually is providing you free parking spots if you are a mayor of that location for the week. This information that is you check in can actually be used to find out your location, your home location also. Peoplehavestudiedotherthings,peoplehavestudiedactuallyfromthepicturesthatyou upload can I actually find out your home location. This work is specifically focused on finding outthe home location fromsocialnetworks like foursquare, and there was a high confidence in finding out the home location with the foursquare check ins that people have done.

Mobility of people is actually not that much. Another conclusion that they also foundwas people do not move a lot from their current location. With this information like check in, mayorship, they were able to actually find out with the high confidence the home location of the person within few kilometers of distance of error.

Therefore, social network data privacy, initially we saw some survey where people actually said about their information of the social networks, then we looked at some studies where pictures uploaded on social media and pictures uploaded onthese publicly available websites which they called as unidentified sources can be actually used to finda person specifically or uniquely identify an individual. So here I am saying that your location can be also inferred from the social networks like foursquare.

(ReferSlideTime:05:26)



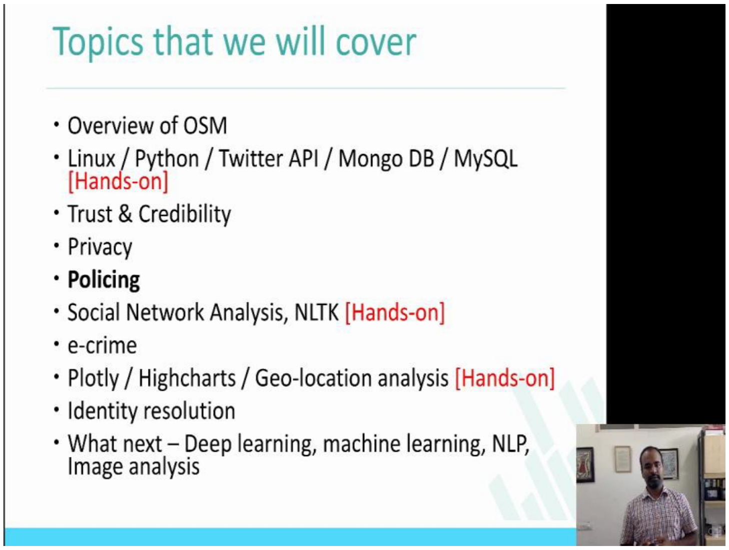
Interested?

- WWW http://www.www2017.com.au/
- ICWSM http://icwsm.org/2017/
- COSN http://cosn.acm.org/
- CSCW https://cscw.acm.org/2017/
- ...

Ifyouareinterestedmoreinthistopicactuallyhighrecommendyoutogetsomeof thesepointers,whichisthesearealltheconferenceswherethetopicsofprivacyand securityinOnlineSocialMedia ingeneralalso getspublished, but also specificallyabout these topics like, image analysis connected to the data from the web, pictures which is what wetalk about and locationand things like that.WWW - which is one ofthe toptier conferences in internet space;then,there isconference onweband socialmedia which is ICWSM. Then there is conference on online social networks and there is also collaborative work which is CSCW.

These are not the comprehensive lists, these are just to tell you that if any of you is interested in looking at these topics more, you should probably be looking into these pointers.And of course, if there is anyquestions feel free to drop it onthe forum.

(ReferSlideTime:06:30)



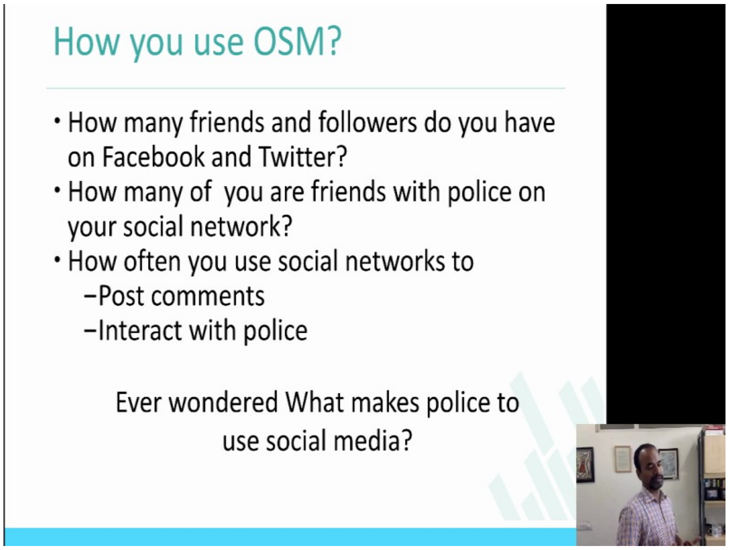
Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- Trust & Credibility
- Privacy
- **Policing**
- Social Network Analysis, NLTK [Hands-on]
- e-crime
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

Now, I thought I will actually move on to another topic given that we are in the week 5 we should actually start looking at other topics also in the course.

So, until now you have done general overview of Online Social Media some Linux, Python, technicaltopics liketrust andcredibilitywhichjust lookedat privacy. So,what I thought I will do now is about spending this week's content mostly on topic called policing.

(ReferSlideTime:07:00)



How you use OSM?

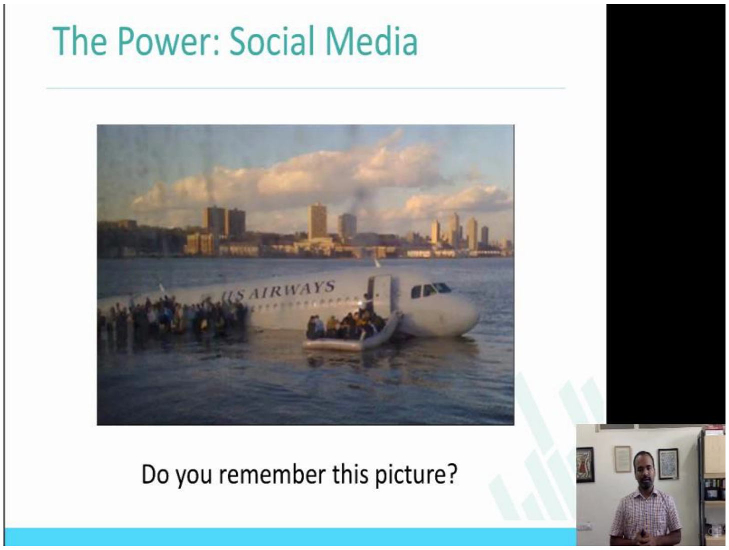
- How many friends and followers do you have on Facebook and Twitter?
- How many of you are friends with police on your social network?
- How often you use social networks to
  - Post comments
  - Interact with police

Ever wondered What makes police to use social media?

I would definitelylike to hear fromstudents inthe class about how you use online social media particularly the kind of questions that I have in the slide. How many friends and followers do you have on Facebook and Twitter? If you can post all this in forum it will be nice to see howthe participants ofthis course are actuallyusing Online SocialMedia.

How many of you are friends with the police on your social network? Police, I mean Police Organizations; how often do you use social networkto post comments or interact with police? And of course, the question that I'm going to be trying to address in this week content is actually what has police been using Online Social Media for, what havethey been doing, what can they do, how we can actually help, how people using social media like you and me can actuallyparticipate with the police in online social networks.

(ReferSlideTime:08:10)



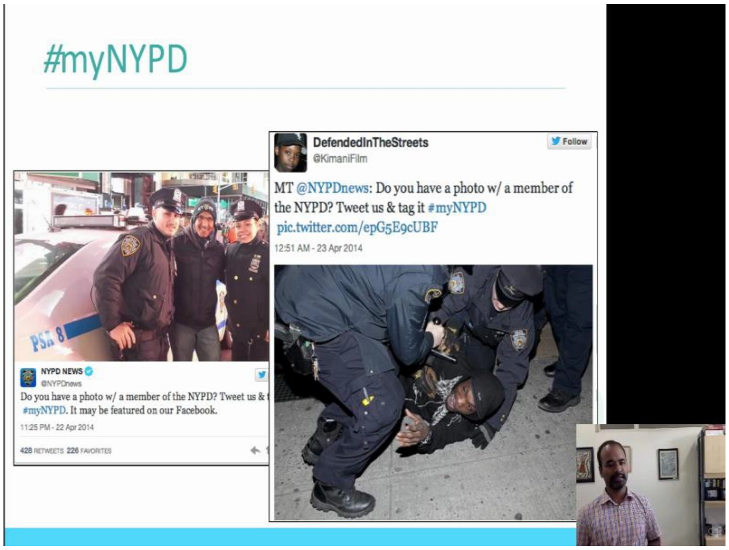
The Power: Social Media

Do you remember this picture?

I do not know how many of you have seen this picture before, but this is the first time ever the socialmedia was actuallyused to attack crisis, to actuallyfind out what is going oninrealworld before people actually, before the first responders, before police actually got to it. This is US airways airplane that landed on Hudson river and you could see people are actually trying to get out or people trying to help. But the first picture and the first tweet that came out, but this picture was actually a person walking on the riverside posting a picture saying that ‘I am going to actually go help them.’

First untilthen this was in 2009 untilthen, the social networkTwitter or anyother social network was actually mostly used for talking to each other, saying what they are doing, hashtag Mondaymorning, things likethat, weretheoncethat were used to talk onsocial networks like Twitter and Facebook. Whereas, first time it was used to solve the problems, solve the crisis is actually this one.

(ReferSlideTime:09:23)



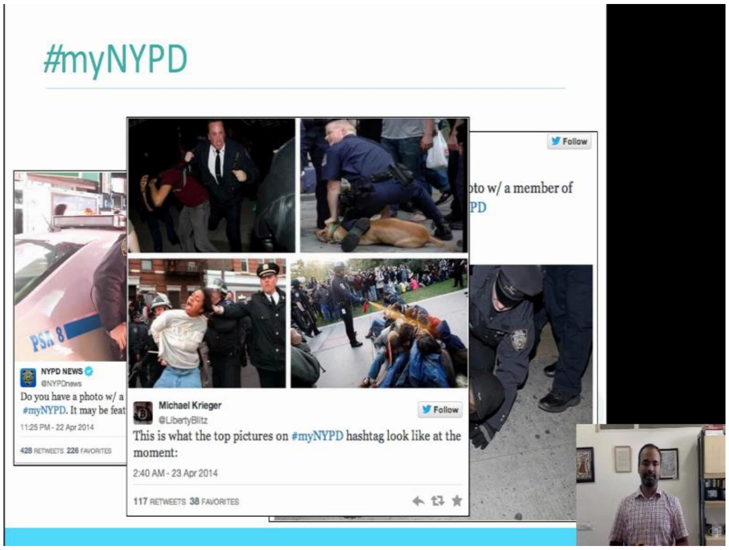
#myNYPD

NYPD NEWS
@NYPDnews
Do you have a photo w/ a member of the NYPD? Tweet us & t
#myNYPD. It may be featured on our Facebook.
11:25 PM - 22 Apr 2014
428 RETWEETS 226 FAVORITES

DefendedInTheStreets
@KimaniFilm
MT @NYPDnews: Do you have a photo w/ a member of the NYPD? Tweet us & tag it #myNYPD pic.twitter.com/epG5E9cUBF
12:51 AM - 23 Apr 2014

And of course, there the police is being part of the social media, socialnetworks for some time now. Here is my example where it became very interesting. So, hash tag myNYPD, is the hashtagthat NYPDpolice posted saying withthis tweet ‘do you have a photo with a member of NYPD, tweet us and tag it, with myNYPD. It may be featuredon our Facebook page. This came out on Twitter and they said that we can actually post the pictures on our Facebook account. Interestingly, with this hash tag they startedgetting pictures like this.

(ReferSlideTime:10:08)



#myNYPD

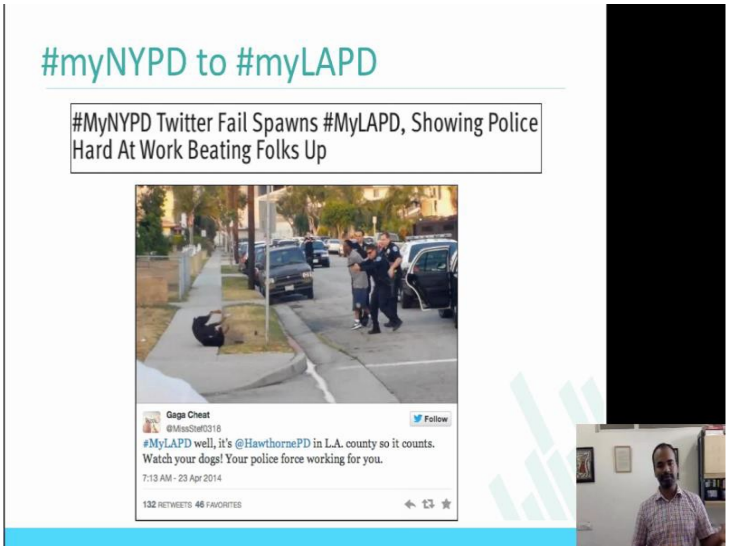
NYPD NEWS
@NYPDnews
Do you have a photo w/ a
#myNYPD. It may be feat
11:25 PM - 22 Apr 2014
428 RETWEETS 226 FAVORITES

Michael Krieger
@LibertyBlitz
This is what the top pictures on #myNYPD hashtag look like at the moment:
2:40 AM - 23 Apr 2014
117 RETWEETS 38 FAVORITES

oto w/ a member of
PD

ItisnotonlylikethattheygotpictureslikethiswhichisfrommyNYPD.

(ReferSlideTime:10:12)



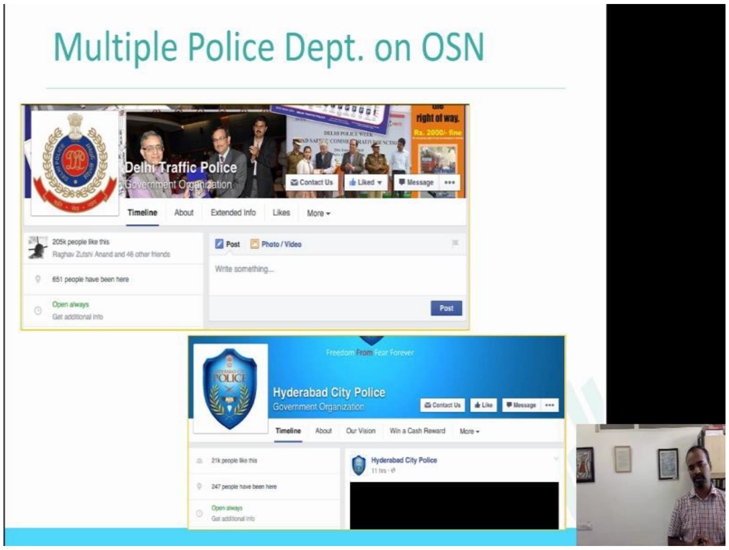
#myNYPD to #myLAPD

#MyNYPD Twitter Fail Spawns #MyLAPD, Showing Police Hard At Work Beating Folks Up

Gaga Cheat
@MissStef0318
#MyLAPD well, it's @HawthornePD in L.A. county so it counts. Watch your dogs! Your police force working for you.
7:13 AM - 23 Apr 2014
132 RETWEETS 46 FAVORITES

But interestingly there were also pictures, this hash tag went from myNYPD tomyLAPD. The point here is that, these kinds of strategies that police use is also happeningonOnlineSocialMediawhichistotakeupthehashtag,getpublicto participate and uploading the picturesand ofcourse sometimes you get different kindsof pictures than what they meant or expected.

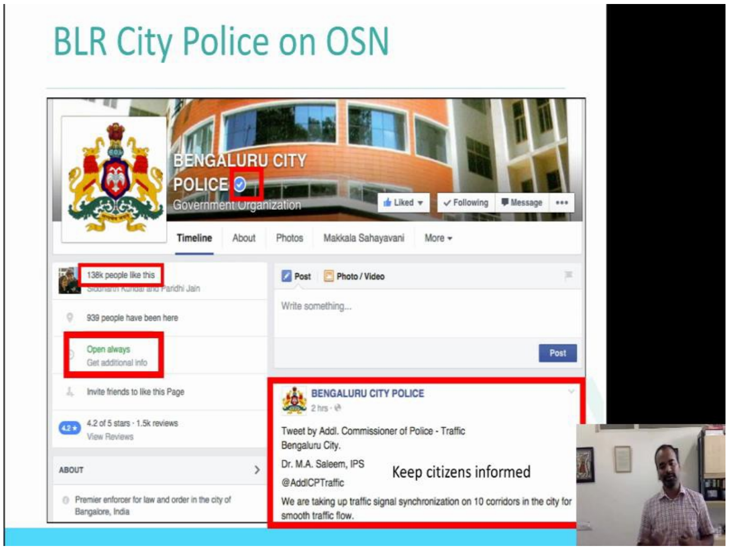
(ReferSlideTime:10:43)



Multiple Police Dept. on OSN

Let uscome backtotheIndiancontext.Inthe lastfew yearsprobablylast oneandhalfor 2 years there is a lot of adoption of Online Social Networks, Online Social Media services in India also. Here are some examples, I am going to go through a lot of accounts, handles which has been using Online Social Media in the last few years. And then, over the process of actually talking about these handles I am also going to inject some technicaltopics interms ofwhat arethe problems, how actuallypeople talking this course and people understanding the social networks technically can also help in identifying these problems. The one on the top leftis actually Delhi traffic police, theone on the bottom is actually Hyderabad city police. I will actually show you moredetails about examples also pretty soon.

(ReferSlideTime:11:37)



BLR City Police on OSN

Premier enforcer for law and order in the city of Bangalore, India

BENGALURU CITY POLICE
2 hrs ·

Tweet by Addl. Commissioner of Police - Traffic Bengaluru City.

Dr. M.A. Saleem, IPS

@AddlCPTraffic

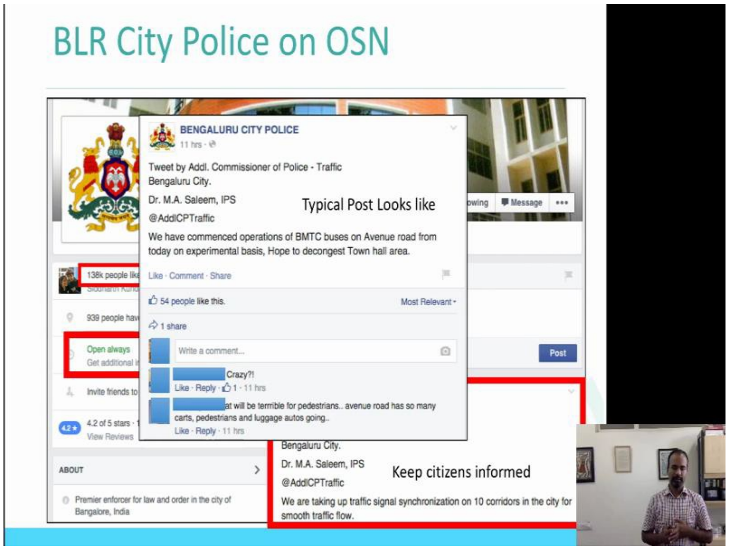
We are taking up traffic signal synchronization on 10 corridors in the city for smooth traffic flow.

Keep citizens informed

Here is the general way by which Police Organizations are actually using the social media services. This is a Facebook page of Bangalore City Police. In India now, Bangalore City Police, Bangalore Traffic Police, Delhi Traffic Page, these are the very popular handles inthe countrynow.And that iswhy you willget,teaching thisparticular course is actually pretty exciting for me, it is because we are actually looking at topics which are very rather relevant, I mean just open your Facebook now and look at Bangalore City Police you will actually look at some of the things I am talking now. Bangalore City Police is the verified page you can see a blue tick next to the top left of Bangalore City Police.

And there is about thousands and thousands of likes this picture that was taken sometime back I am sure the likes have changed now.And in the bottomthe picture showing you an example of a typical post that comes from these kinds of police pages. The post says that 'we are taking up traffic signals synchronization on 10 corridors in the city for smooth traffic flow' and it is coming from handle AddICPTraffic. So the idea is that citizens can inform about what is going on in the decisions that theyare making.

(ReferSlideTime:13:07)



BLR City Police on OSN

BENGALURU CITY POLICE
11 hrs ·

Tweet by Addl. Commissioner of Police - Traffic Bengaluru City.
Dr. M.A. Saleem, IPS
@AddlCPTraffic
We have commenced operations of BMTC buses on Avenue road from today on experimental basis, Hope to decongest Town hall area.

Typical Post Looks like

54 people like this.
1 share

[redacted] Crazy?!
Like · Reply · 1 · 11 hrs

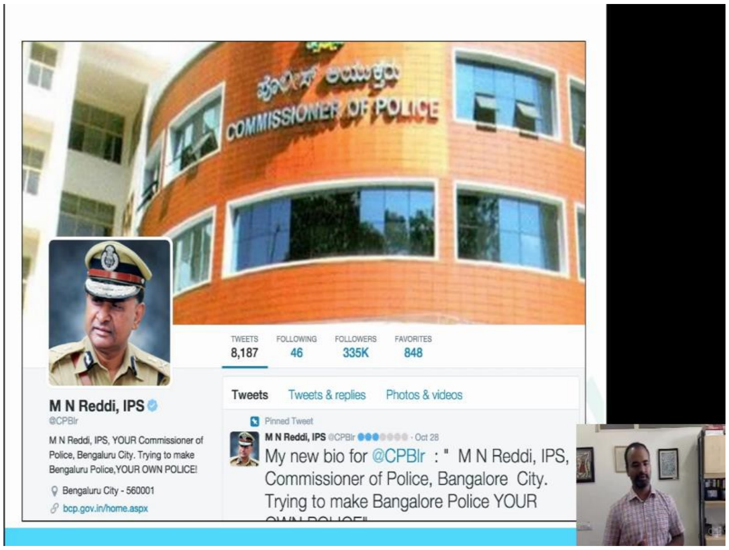
[redacted]at will be terrible for pedestrians.. avenue road has so many carts, pedestrians and luggage autos going..
Like · Reply · 11 hrs

Bengaluru City.
Dr. M.A. Saleem, IPS
@AddlCPTraffic
We are taking up traffic signal synchronization on 10 corridors in the city for smooth traffic flow.

Keep citizens informed

Andyoucan of course see a numberof likes,peopleinteractingwith these posts,with the Police Organizations.

(ReferSlideTime:13:20)



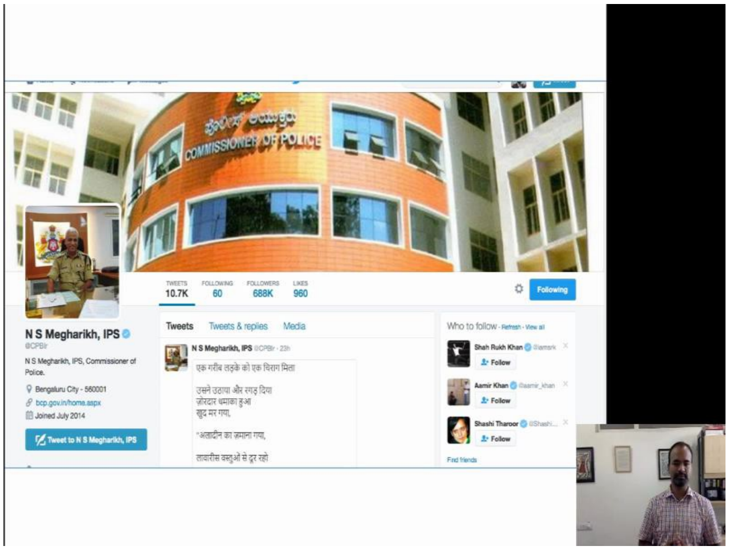
TWEETS	FOLLOWING	FOLLOWERS	FAVORITES
8,187	46	335K	848

M N Reddi, IPS
@CPBlr
M N Reddi, IPS, YOUR Commissioner of Police, Bengaluru City. Trying to make Bengaluru Police,YOUR OWN POLICE!
Bengaluru City - 560001
bcp.gov.in/home.aspx

M N Reddi, IPS @CPBlr · Oct 28
My new bio for @CPBlr : " M N Reddi, IPS, Commissioner of Police, Bangalore City. Trying to make Bangalore Police YOUR OWN POLICE"

SoCPBLR,atCPBLR isahandlethatgotverypopularaboutoneandhalfyearsbefore.

(ReferSlideTime:13:32)



N S Megharikh, IPS
@CPBlr
N S Megharikh, IPS, Commissioner of Police.
Bengaluru City - 560001
bcp.gov.in/home.aspx
Joined July 2014

N S Megharikh, IPS @CPBlr · 23h

एक गरीब लड़के को एक चिराग मिला

उसने उठाया और रगड़ दिया
ज़ोरदार धमाका हुआ
खुद मर गया,

"अलादीन का ज़माना गया,

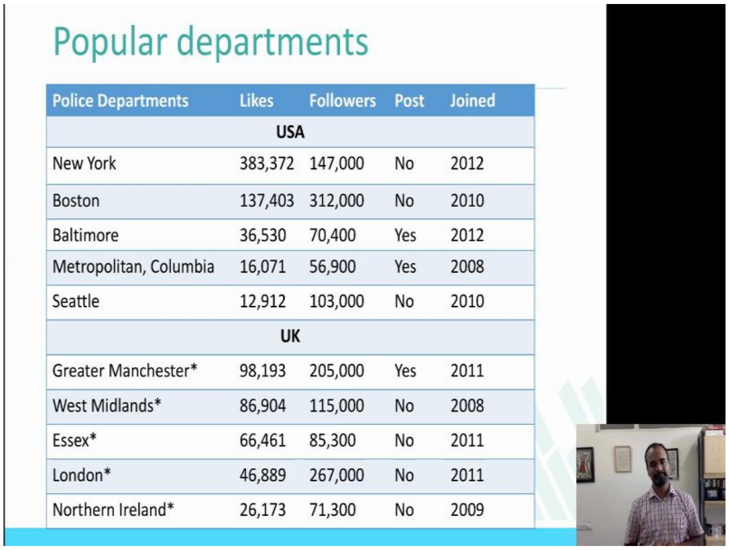
लावारिस वस्तुओं से दूर रहो

This is the current account of CPBLR, it has a part I just took the screenshot a few minutes before actually preparing this deck of slides. It is about 10000 tweets, the account is following 60 people and about 688,000 people are actually following this account. Then this shows that this account is actually very popular.

For example, probably I took the screenshot a year or sometime beforeI took on 28thlast year probably and it says about 335,000. And now it is actually August 2016 it is about 688,000.The activity of this account is actually very high and the number of followers; the interaction that this account has with citizens is also pretty high.Therefore, they gain a lot of followers.

So, what we are going to be looking at this is, how these accounts are doing, in specific we will also look at some data that was collected to analyze Bangalore data itself.

(ReferSlideTime:14:40)



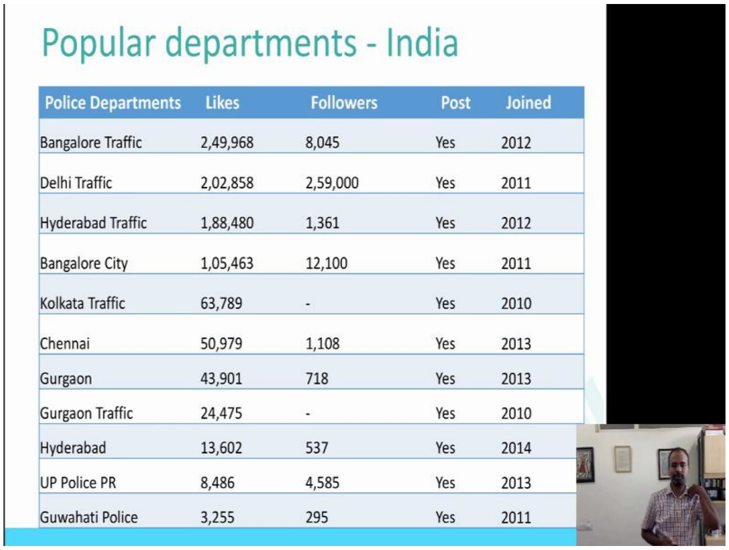
Popular departments

Police Departments	Likes	Followers	Post	Joined
USA				
New York	383,372	147,000	No	2012
Boston	137,403	312,000	No	2010
Baltimore	36,530	70,400	Yes	2012
Metropolitan, Columbia	16,071	56,900	Yes	2008
Seattle	12,912	103,000	No	2010
UK				
Greater Manchester*	98,193	205,000	Yes	2011
West Midlands*	86,904	115,000	No	2008
Essex*	66,461	85,300	No	2011
London*	46,889	267,000	No	2011
Northern Ireland*	26,173	71,300	No	2009

Just to give you a sense of how the Police Organizations around the world are. Police organizations like New York, Boston, Baltimore and many, many other cities have actually adopted Facebook’s and Twitter’s. Here we are just showing you the Facebook which is likes and followers. The common post and joined is that joined is the year that they started. Post is whether they are allowed to actually post the content on theFacebook page, because on Facebook you can actually control whether the people who have liked the post or people who are accessing this page can actually post on to thepage.

In UK also there has been a lot of adoption of Facebook. Of course, if one is interested you could actually look at this topic more closely and do a lots of analysis on analyzing the post, analyzing the content that is going on online social network, both in the policein context. If any of you are interested in doing it,if any of you are interested in looking at this problem more closely I will be happyto actually chat with you.

(ReferSlideTime:15:55)



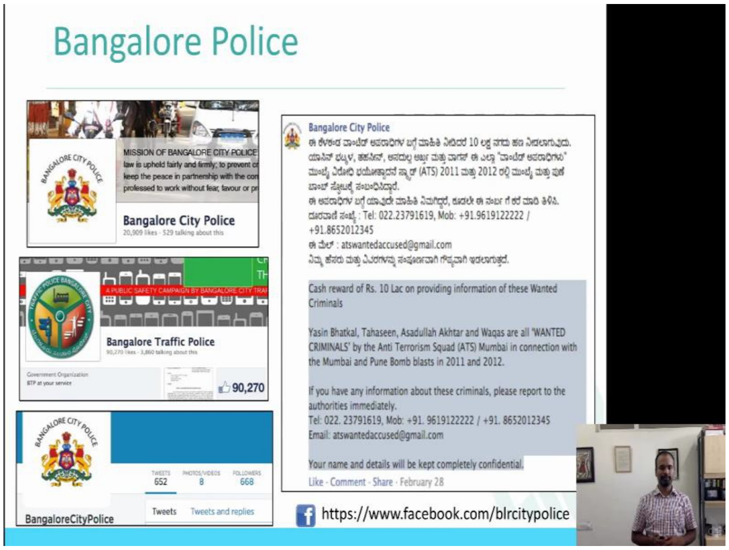
Popular departments - India

Police Departments	Likes	Followers	Post	Joined
Bangalore Traffic	2,49,968	8,045	Yes	2012
Delhi Traffic	2,02,858	2,59,000	Yes	2011
Hyderabad Traffic	1,88,480	1,361	Yes	2012
Bangalore City	1,05,463	12,100	Yes	2011
Kolkata Traffic	63,789	-	Yes	2010
Chennai	50,979	1,108	Yes	2013
Gurgaon	43,901	718	Yes	2013
Gurgaon Traffic	24,475	-	Yes	2010
Hyderabad	13,602	537	Yes	2014
UP Police PR	8,486	4,585	Yes	2013
Guwahati Police	3,255	295	Yes	2011

So, here is the situation what is in India now? This is slightly outdated data but I think this is Bangalore, Delhi Traffic Police which are actually popular, and of course many cities have actuallyadopted getting onto Online SocialMedia and particularlyFacebook and Twitter.

There is a little bit of YouTube but I think YouTube is pretty small. In our case the accounts have probably started after 2010 or 2011 and all the pages that are created in India for using for policing allows citizens to actuallypost onthese pages.

(ReferSlideTime:16:38)



Bangalore Police

MISSION OF BANGALORE CITY POLICE
law is upheld fairly and firmly; to prevent cr
keep the peace in partnership with the com
professed to work without fear, favour or pr
Bangalore City Police
20,909 likes · 529 talking about this
A PUBLIC SAFETY CAMPAIGN BY BANGALORE CITY TRAF
Bangalore Traffic Police
90,270 likes · 3,860 talking about this
Government Organization
BTP at your service
90,270
TWEETS 652
PHOTOS/VIDEOS 8
FOLLOWERS 668
Tweets
Tweets and replies
BangaloreCityPolice

Bangalore City Police
ಈ ಕೆಳಕಂಡ ವಾಂಟೆಡ್ ಅಪರಾಧಿಗಳ ಬಗ್ಗೆ ಮಾಹಿತಿ ನೀಡಿದರೆ 10 ಲಕ್ಷ ನಗದು ಹಣ ನೀಡಲಾಗುವುದು.
ಯಾಸಿನ್ ಭಟ್ಕಳ, ತಹಸೀನ್, ಅಸದುಲ್ಲ ಅಖ್ತರ್ ಮತ್ತು ವಾಕಸ್ ಈ ಎಲ್ಲಾ "ವಾಂಟೆಡ್ ಅಪರಾಧಿಗಳು"
ಮುಂಬೈ ವಿರೋಧಿ ಭಯೋತ್ಪಾದನೆ ಸ್ಕ್ವಾಡ್ (ATS) 2011 ಮತ್ತು 2012 ರಲ್ಲಿ ಮುಂಬೈ ಮತ್ತು ಪುಣೆ
ಬಾಂಬ್ ಸ್ಫೋಟಕ್ಕೆ ಸಂಬಂಧಿಸಿದ್ದಾರೆ.
ಈ ಅಪರಾಧಿಗಳ ಬಗ್ಗೆ ಯಾವುದೇ ಮಾಹಿತಿ ನಿಮಗಿದ್ದರೆ, ಕೂಡಲೇ ಈ ನಂಬರ್ ಗೆ ಕರೆ ಮಾಡಿ ತಿಳಿಸಿ.
ದೂರವಾಣಿ ಸಂಖ್ಯೆ : Tel: 022.23791619, Mob: +91.9619122222 /
+91.8652012345
ಈ ಮೇಲ್ : atswantedaccused@gmail.com
ನಿಮ್ಮ ಹೆಸರು ಮತ್ತು ವಿವರಗಳನ್ನು ಸಂಪೂರ್ಣವಾಗಿ ಗೌಪ್ಯವಾಗಿ ಇಡಲಾಗುತ್ತದೆ.

Cash reward of Rs. 10 Lac on providing information of these Wanted Criminals

Yasin Bhatkal, Tahaseen, Asadullah Akhtar and Waqas are all 'WANTED CRIMINALS' by the Anti Terrorism Squad (ATS) Mumbai in connection with the Mumbai and Pune Bomb blasts in 2011 and 2012.

If you have any information about these criminals, please report to the authorities immediately.
Tel: 022. 23791619, Mob: +91. 9619122222 / +91. 8652012345
Email: atswantedaccused@gmail.com

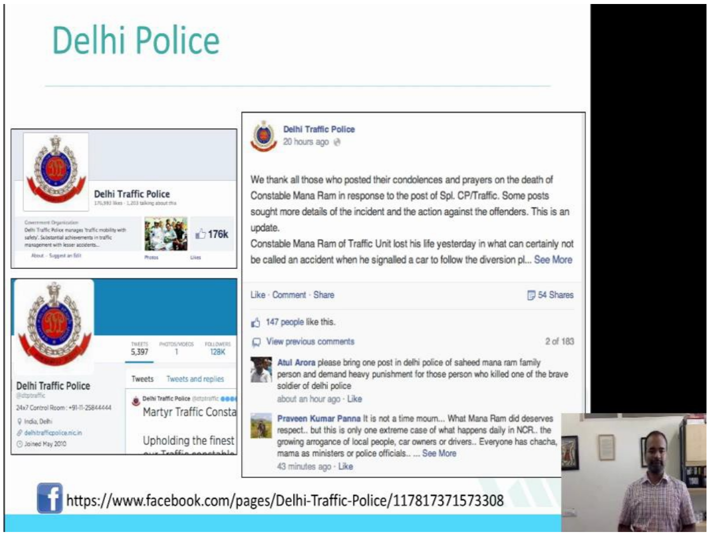
Your name and details will be kept completely confidential.

Like · Comment · Share · February 28

https://www.facebook.com/blrcitypolice

Let me walk you through this some city police handles and then look at what kind of things that they do on these handles. This is Bangalore City Police, Bangalore Traffic Police, Facebook and Twitter page, this is a typical police I told you about traffic earlier the other post that they could doissomethinglike this which is cash reward of Rupees 10 lakh for helping them.

(ReferSlideTime:17:04)



Delhi Police

Delhi Traffic Police
176,593 likes · 1,203 talking about this
Government Organization
Delhi Traffic Police manages traffic mobility with safety. Substantial achievements in traffic management with lesser accidents...
About · Suggest an Edit
Photos
Likes
176k
TWEETS 5,397
PHOTOS/VIDEOS 1
FOLLOWERS 128K
Tweets
Tweets and replies
Delhi Traffic Police
@dtptraffic
24x7 Control Room: +91-11-25844444
India, Delhi
delhitrafficpolice.nic.in
Joined May 2010
Delhi Traffic Police @dtptraffic
Martyr Traffic Consta
Upholding the finest

Delhi Traffic Police
20 hours ago

We thank all those who posted their condolences and prayers on the death of Constable Mana Ram in response to the post of Spl. CP/Traffic. Some posts sought more details of the incident and the action against the offenders. This is an update.
Constable Mana Ram of Traffic Unit lost his life yesterday in what can certainly not be called an accident when he signalled a car to follow the diversion pl... See More

Like · Comment · Share 54 Shares

147 people like this.

View previous comments 2 of 183

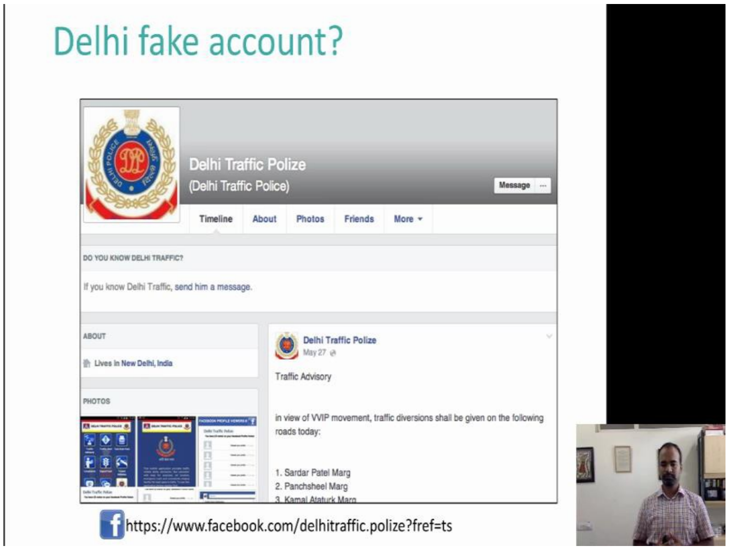
Atul Arora please bring one post in delhi police of saheed mana ram family person and demand heavy punishment for those person who killed one of the brave soldier of delhi police
about an hour ago · Like

Praveen Kumar Panna It is not a time mourn... What Mana Ram did deserves respect.. but this is only one extreme case of what happens daily in NCR.. the growing arrogance of local people, car owners or drivers.. Everyone has chacha, mama as ministers or police officials.. ... See More
43 minutes ago · Like

https://www.facebook.com/pages/Delhi-Traffic-Police/117817371573308

This is Delhitraffic.Again Delhitraffic, Twitter and Facebook pages.This is the kind of post thattheDelhipolicealso doesintermsoftraffic, intermsofactuallyappreciationof their police force, their police officers.

(ReferSlideTime:17:22)



Delhi fake account?

Delhi Traffic Polize
(Delhi Traffic Police)

DO YOU KNOW DELHI TRAFFIC?

If you know Delhi Traffic, send him a message.

ABOUT

Lives in New Delhi, India

PHOTOS

Delhi Traffic Polize
May 27

Traffic Advisory

in view of VVIP movement, traffic diversions shall be given on the following roads today:

1. Sardar Patel Marg
2. Panchsheel Marg
3. Kamal Ataturk Marg

https://www.facebook.com/delhitraffic.polize?fref=ts

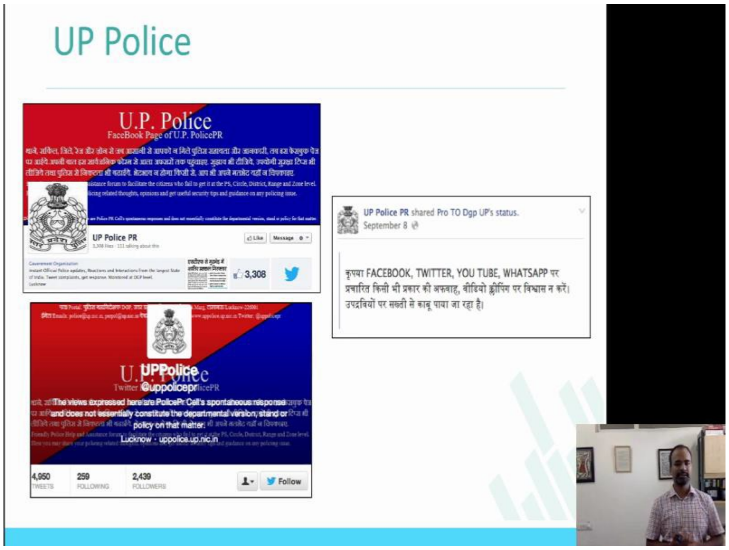
Interestingly, there is not only that these handles have been managed by these Police Organizations, there are also fake handles that are being generated because of using of social media byPolice Organizations also. I think it is typicalproblemthat technologists like students who are taking this class should probably look at.

What are the ways to actually identify this kind of fake content? First of all even to identify whether there are fake handles, 500 accounts which are doing the similar things that the Delhi Traffic Police page is doing or any other Police Organizations. Wherever you are from meaning I am sure the cities like you are sitting in and looking at these lectures you could probably look at whether the Police Organizations in your state, in your cityhas a Facebook page or aTwitter page and see what kind ofactivities that they are doing.

It will be interesting if some of you actually post about your own city or state police organizationactivitiesontheforum.ItwillbeinterestingtolookatwhatPolice

Organizations are actually doing. Maybe, by looking at these topics for sometime you willget a sense but Ithink ifyou are a localite you willprobablyalso understand what is going on. Inthis case, the slide that I have here which is DelhiTraffic Police- po l iz e is a fake account and there are many, many accounts like these for a different Police Organizations.

(ReferSlideTime:18:46)



UP Police

U.P. Police

FaceBook Page of U.P. PolicePR

थाने, सर्किल, जिले, रेंज और ज़ोन से जब आसानी से आपको न मिले पुलिस सहायता और जानकारी, तब इस फेसबुक पेज पर आईये. अपनी बात इस सार्वजनिक फोरम से आला अफसरों तक पहुंचाइए. सुझाव भी दीजिये, उपयोगी सुरक्षा टिप्स भी लीजिये तथा पुलिस से निकटता भी बढ़ाइये.

UP Police PR

UP Police PR shared Pro TO Dgp UP's status.

September 8

कृपया FACEBOOK, TWITTER, YOU TUBE, WHATSAPP पर प्रचारित किसी भी प्रकार की अफवाह, वीडियो क्लीपिंग पर विश्वास न करें। उपद्रवियों पर सख्ती से काबू पाया जा रहा है।

UPPolice

@uppolicepr

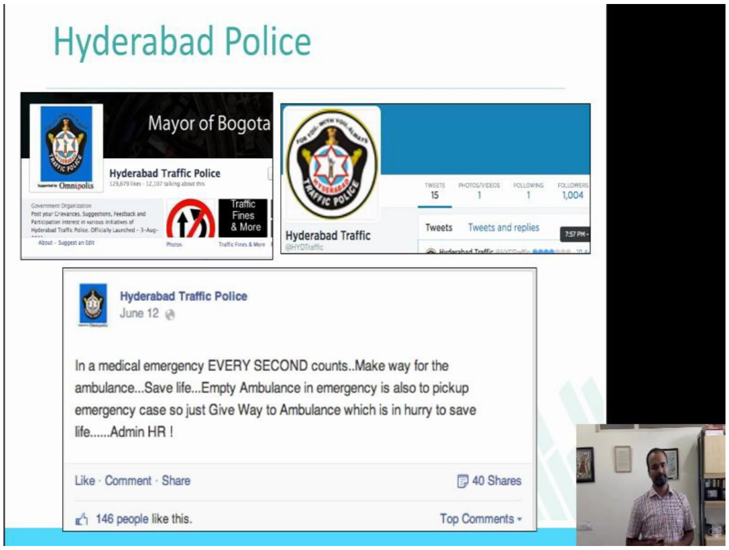
The views expressed here are PolicePr Cell's spontaneous response and does not essentially constitute the departmental version, stand or policy on that matter.

Lucknow · uppolice.up.nic.in

4,950	259	2,439
TWEETS	FOLLOWING	FOLLOWERS

Here is UPPolice. UPPolice also has a Facebook page and aTwitter page and theyalso seem to be talking more about the activities in terms of traffic in terms of helping them and in terms of generally interacting with the citizens.

(ReferSlideTime:19:03)



Hyderabad Police

Mayor of Bogota

Hyderabad Traffic Police

129,679 likes · 12,107 talking about this

Government Organization

Post your Grievances, Suggestions, Feedback and Participation interest in various initiatives of Hyderabad Traffic Police. Officially Launched - 3-Aug-

About · Suggest an Edit

Photos

Traffic Fines & More

Hyderabad Traffic

@HYDTraffic

TWEETS	PHOTOS/VIDEOS	FOLLOWING	FOLLOWERS
15	1	1	1,004

Tweets · Tweets and replies

Hyderabad Traffic Police

June 12

In a medical emergency EVERY SECOND counts..Make way for the ambulance...Save life...Empty Ambulance in emergency is also to pickup emergency case so just Give Way to Ambulance which is in hurry to save life......Admin HR !

Like · Comment · Share

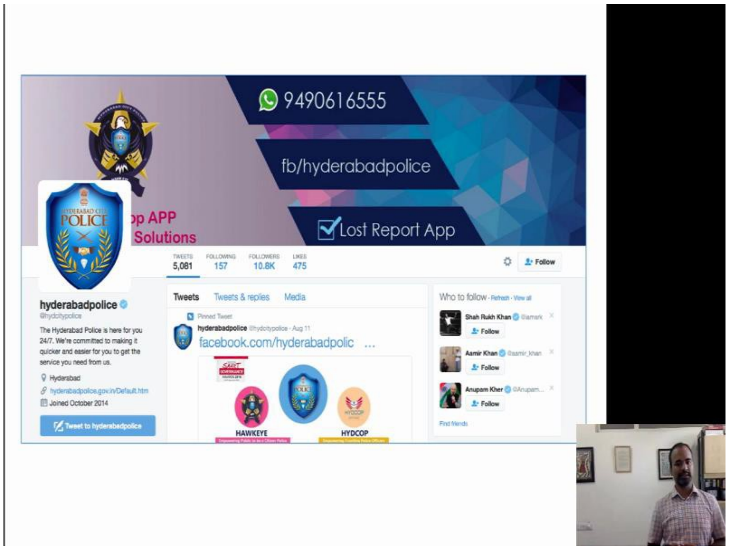
40 Shares

146 people like this.

Top Comments

That isHyderabadPoliceofcourse.

(ReferSlideTime:19:08)



9490616555

fb/hyderabadpolice

☑Lost Report App

op APP

Solutions

TWEETS	FOLLOWING	FOLLOWERS	LIKES
5,081	157	10.8K	475

Follow

hyderabadpolice

@hydcitypolice

The Hyderabad Police is here for you 24/7. We're committed to making it quicker and easier for you to get the service you need from us.

Hyderabad

hyderabadpolice.gov.in/Default.htm

Joined October 2014

Tweet to hyderabadpolice

Tweets · Tweets & replies · Media

Pinned Tweet

hyderabadpolice @hydcitypolice · Aug 11

facebook.com/hyderabadpolic ...

HAWKEYE

HYDCOP

Who to follow · Refresh · View all

Shah Rukh Khan @iamsrk

Follow

Aamir Khan @aamir_khan

Follow

Anupam Kher @Anupam...

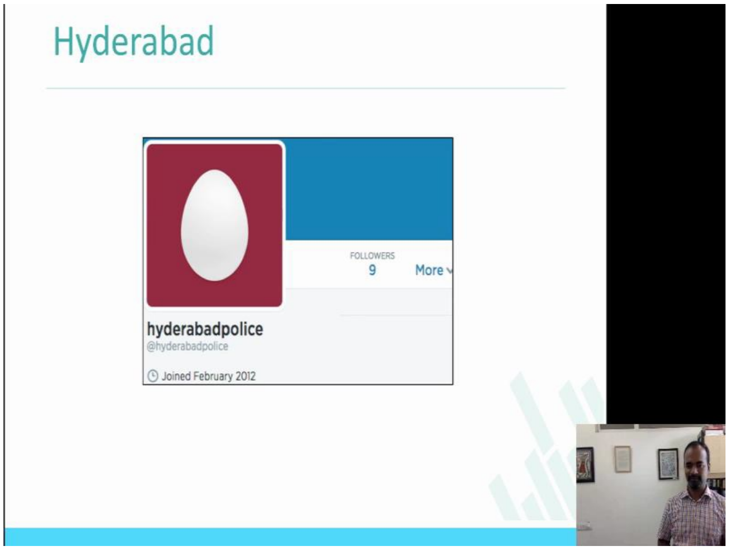
Follow

Find friends

This is the Hyderabad Police recent Twitter handle; I just took the screenshot few minutes before again. This is also a verified account; I will come back to this verified accountinfewminutes.ButIthinkletmejusthighlightwhataverifiedaccountthereis, the blue tick mark next to Hyderabad Police here is the verified account; says that this account is verified that is, Twitter has verified this is reallythe Hyderabad Police.

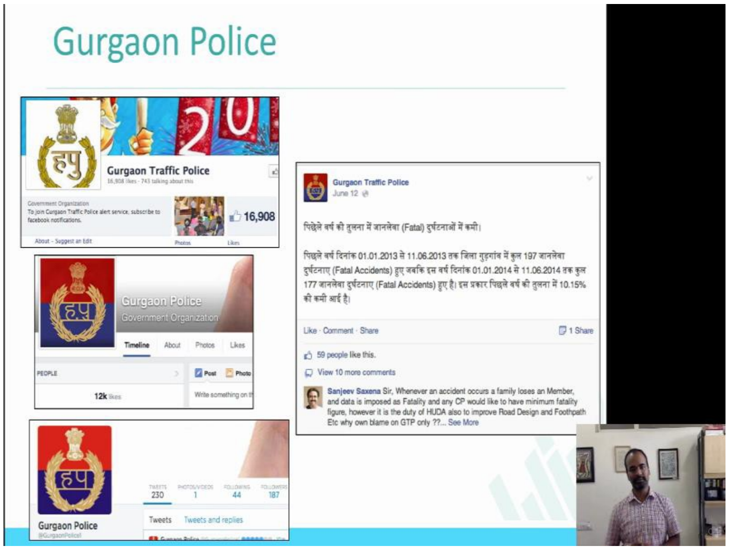
Itisnotthat easyto get averifiedaccount, not allaccountsonTwitterareverified;onlya little fraction of users on Twitter are actually verified, I think it is about a million users now.These verified accountsare veryhelpfulbecause Ithink for Police Organizationsto say that who they are is actually very, very necessary.

(ReferSlideTime:20:09)


Hyderabad

Just to show you this is Hyd City Police which is the real verified account, but as thereare actually accounts like Hyderabad police which does not even have profile picture changed; this is called an egg profile in Twitter terms, and you can see this handle hasbeen from 2012. Sometimes, this handle may not exist now but I think when we wereanalyzing, we keep track of some of these handles also.

(ReferSlideTime:20:30)



Gurgaon Police

Gurgaon Traffic Police

16,908 likes · 743 talking about this

Government Organization

To join Gurgaon Traffic Police alert service, subscribe to facebook notifications.

About · Suggest an Edit

Photos

16,908 Likes

Gurgaon Police

Government Organization

Timeline About Photos Likes

PEOPLE

12k likes

Post Photo

Write something on th

TWEETS	PHOTOS/VIDEOS	FOLLOWING	FOLLOWERS
230	1	44	187

Tweets Tweets and replies

Gurgaon Police

@GurgaonPolice1

Gurgaon Traffic Police

June 12

पिछेले वर्ष की तुलना में जानलेवा (Fatal) दुर्घटनाओं में कमी।

पिछले वर्ष दिनांक 01.01.2013 से 11.06.2013 तक जिला गुड़गांव में कुल 197 जानलेवा दुर्घटनाए (Fatal Accidents) हुए जबकि इस वर्ष दिनांक 01.01.2014 से 11.06.2014 तक कुल 177 जानलेवा दुर्घटनाए (Fatal Accidents) हुए है। इस प्रकार पिछले वर्ष की तुलना में 10.15% की कमी आई है।

Like · Comment · Share 1 Share

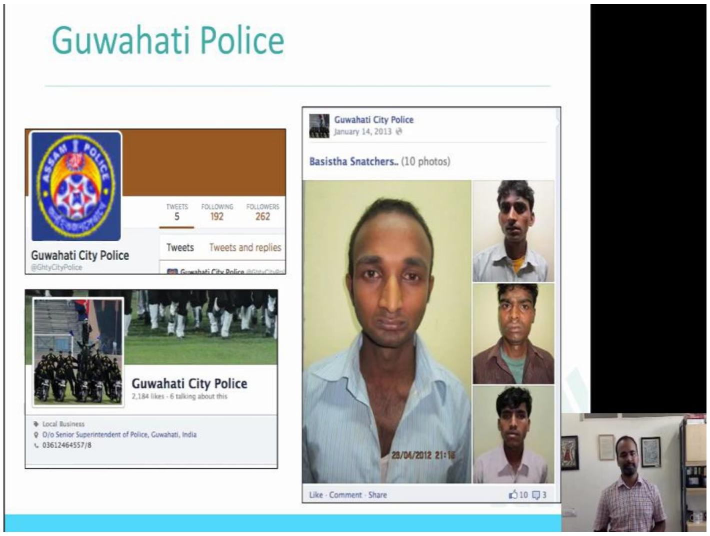
59 people like this.

View 10 more comments

Sanjeev Saxena Sir, Whenever an accident occurs a family loses an Member, and data is imposed as Fatality and any CP would like to have minimum fatality figure, however it is the duty of HUDA also to improve Road Design and Foothpath Etc why own blame on GTP only ??... See More

Sometimes, some states have taken the decision of all the states and all the city police will have just one page, and then everybodywill be posting content and interacting only throughthat. Some organizations have takenthe decision ofhaving multiple pages for at the district level for different activities.

(ReferSlideTime:20:52)



Guwahati Police

TWEETS	FOLLOWING	FOLLOWERS
5	192	262

Tweets Tweets and replies

Guwahati City Police

@GhtyCityPolice

Guwahati City Police

2,184 likes · 6 talking about this

Local Business

O/o Senior Superintendent of Police, Guwahati, India

03612464557/8

Guwahati City Police

January 14, 2013

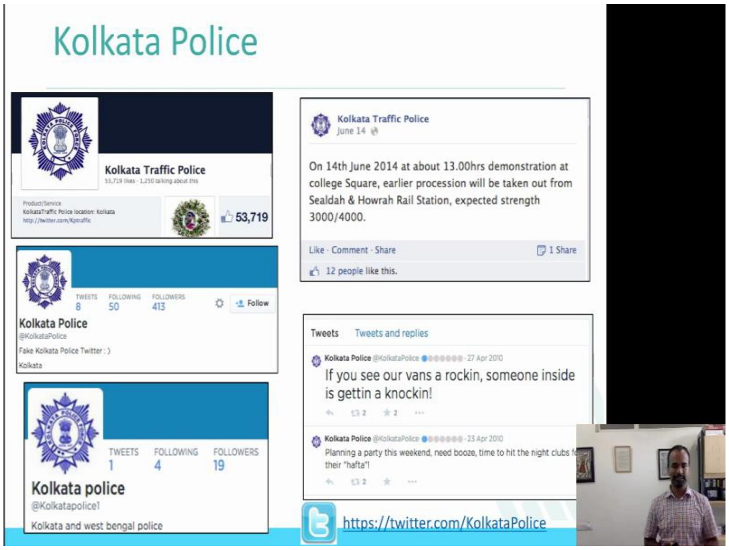
Basistha Snatchers.. (10 photos)

28/04/2012 21:16

Like · Comment · Share 10 3

GuwahatiPolice.

(ReferSlideTime:20:56)



Kolkata Police

Kolkata Traffic Police

53,719 likes · 1,250 talking about this

Product/Service
KolkataTraffic Police location: Kolkata
http://twitter.com/Kptraffic

53,719

TWEETS 8 FOLLOWING 50 FOLLOWERS 413

Follow

Kolkata Police
@KolkataPolice
Fake Kolkata Police Twitter :)
Kolkata

TWEETS 1 FOLLOWING 4 FOLLOWERS 19

Kolkata police
@Kolkatapolice1
Kolkata and west bengal police

Kolkata Traffic Police
June 14

On 14th June 2014 at about 13.00hrs demonstration at college Square, earlier procession will be taken out from Sealdah & Howrah Rail Station, expected strength 3000/4000.

Like · Comment · Share 1 Share

12 people like this.

Tweets Tweets and replies

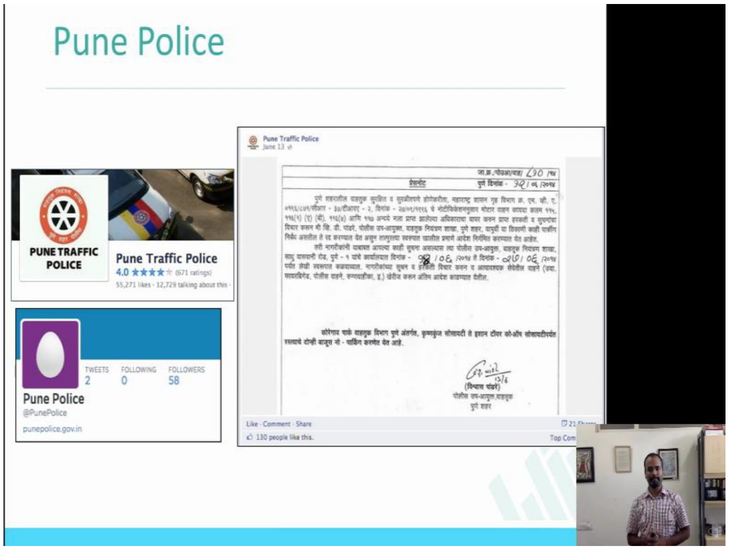
Kolkata Police @KolkataPolice · 27 Apr 2010
If you see our vans a rockin, someone inside is gettin a knockin!

Kolkata Police @KolkataPolice · 23 Apr 2010
Planning a party this weekend, need booze, time to hit the night clubs f
their "hafta"!

https://twitter.com/KolkataPolice

Kolkata Police. I think Kolkata Police also there is a interesting handle which is on the second from the top on the left which says Kolkata Police, fake Kolkata Police twitter,and then a smiley. Therefore, there are these kind of problems where Police Organizationshandles have been mimicked, mastibated to create accountsand see ifyou cansee it actuallyhas at Kolkata Police;the profile picturethat it isusing isprobablythe same as a legitimate Kolkata Police account also.

(ReferSlideTime:21:27)



Pune Police

Pune Traffic Police
June 13

जा.क्र./पोउआ/वाह/ ८९० /१४

प्रेसनोट

पुणे दिनांक - १२ / ०६ /२०१४

पुणे शहरातील वाहतूक सुरळीत व सुरक्षीतपणे होणेकरीता, महाराष्ट्र शासन गृह विभाग क्र. एम. व्ही. ए. ०११६/८७१/सीआर - ३७/टीआरए - २, दिनांक - २७/०९/१९९६ चे नोटीफिकेशननुसार मोटार वाहन कायदा कलम ११५, ११६(१) (ए) (बी), ११६(४) आणि ११७ अन्वये मला प्राप्त झालेल्या अधिकाराचा वापर करुन प्राप्त हरकती व सुचनांचा विचार करुन मी वि. डी. पांढरे, पोलीस उप-आयुक्त, वाहतूक नियंत्रण शाखा, पुणे शहर, यापुर्वी या ठिकाणी काही पार्कीग निर्बंध असतील ते रद्द करण्यात येत असुन तात्पुरत्या स्वरुपात खालील प्रमाणे आदेश निर्गमित करण्यात येत आहेत.

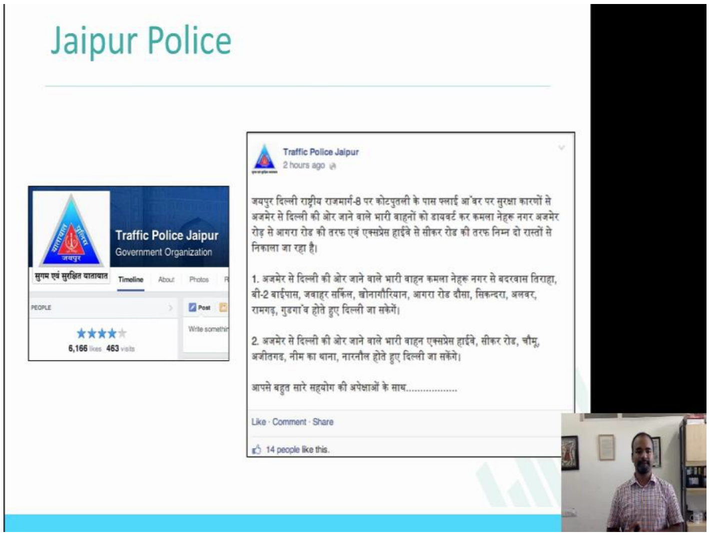
तरी नागरीकांनी याबाबत आपल्या काही सुचना असल्यास त्या पोलीस उप-आयुक्त, वाहतूक नियंत्रण शाखा, साधु वासवानी रोड, पुणे - १ यांचे कार्यालयात दिनांक - १४ / ०६ /२०१४ ते दिनांक - २७ / ०६ /२०१४ पर्यंत लेखी स्वरुपात कळवाव्यात. नागरीकांच्या सुचना व हरकती विचार करुन व आवश्यक सेवेतील वाहने (उदा. फायरब्रिगेड, पोलीस वाहने, रुग्णवाहीका, इ.) वगळून अंतिम आदेश काढण्यात येतील.

कोरेगाव पार्क वाहतूक विभाग पुणे अंतर्गत, कृष्णकुंज सोसायटी ते इशान टॉवर को-ऑप सोसायटीपर्यंत रस्त्याचे दोन्ही बाजूस नो - पार्कीग करणेत येत आहे.

13/6
(विश्वास पांढरे)
पोलीस उप-आयुक्त,वाहतूक
पुणे शहर

PunePolice.AgainPuneTrafficPolice, PunePoliceand it isactuallyveryhardto verify, even for us, even for anybody, even when you go to get your own cities or state police pages as I said now, you yourself will find it hard to get the actual real account which unless it is verified is going to be a veryhard to actually justifyor find out whether they are legitimate.

(ReferSlideTime:21:54)



Jaipur Police

Traffic Police Jaipur
2 hours ago

जयपुर दिल्ली राष्ट्रीय राजमार्ग-8 पर कोटपुतली के पास फ्लाई आॅवर पर सुरक्षा कारणों से अजमेर से दिल्ली की ओर जाने वाले भारी वाहनों को डायवर्ट कर कमला नेहरू नगर अजमेर रोड़ से आगरा रोड की तरफ एवं एक्सप्रेस हाईवे से सीकर रोड की तरफ निम्न दो रास्तों से निकाला जा रहा है।

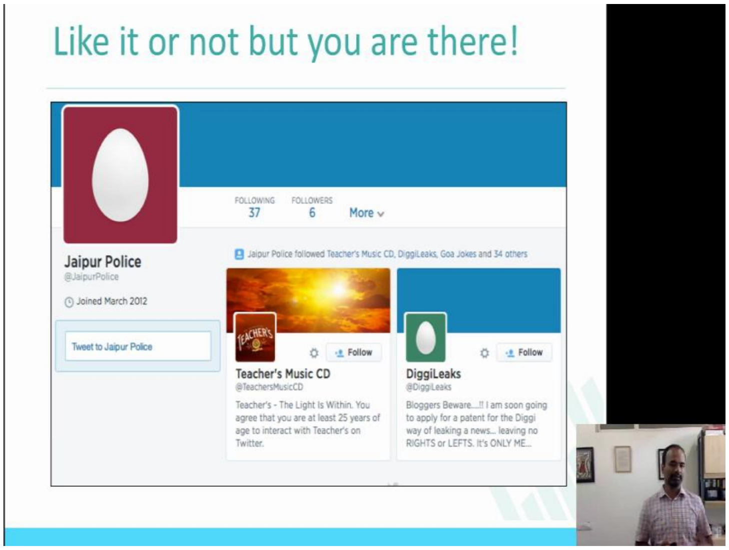
1. अजमेर से दिल्ली की ओर जाने वाले भारी वाहन कमला नेहरू नगर से बदरवास तिराहा, बी-2 बाईपास, जवाहर सर्किल, खोनागौरियान, आगरा रोड दौसा, सिकन्दरा, अलवर, रामगढ़, गुडगांव होते हुए दिल्ली जा सकेगें।

2. अजमेर से दिल्ली की ओर जाने वाले भारी वाहन एक्सप्रेस हाईवे, सीकर रोड, चौमू, अजीतगढ, नीम का थाना, नारनौल होते हुए दिल्ली जा सकेंगे।

आपसे बहुत सारे सहयोग की अपेक्षाओं के साथ..................

ThisisTrafficPoliceJaipur.

(ReferSlideTime:21:59)



Like it or not but you are there!

Jaipur Police
@JaipurPolice
Joined March 2012

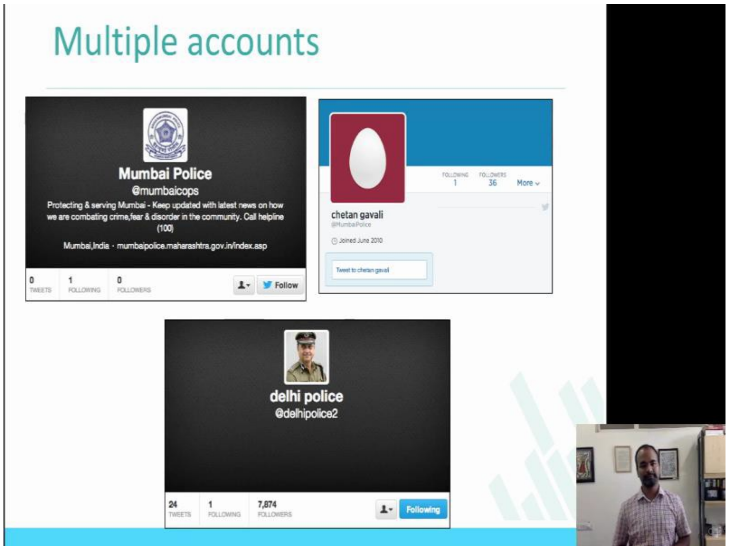
Jaipur Police followed Teacher's Music CD, DiggiLeaks, Goa Jokes and 34 others

Teacher's Music CD
@TeachersMusicCD
Teacher's - The Light Is Within. You agree that you are at least 25 years of age to interact with Teacher's on Twitter.

DiggiLeaks
@DiggiLeaks
Bloggers Beware....!! I am soon going to apply for a patent for the Diggi way of leaking a news... leaving no RIGHTS or LEFTS. It's ONLY ME...

I am just giving you vulnerant tour of some of the handles that we have found. The thought here is that whether you like it or not, whether as citizens want to see their accounts Police Organization accounts in these networks, the accounts are created like I have taken police there are many other polices; it's not clear whether these are actually the Police Organization handles.

(ReferSlideTime:22:24)



Multiple accounts

Mumbai Police

@mumbaicops

Protecting & serving Mumbai - Keep updated with latest news on how we are combating crime,fear & disorder in the community. Call helpline (100)

Mumbai,India · mumbaipolice.maharashtra.gov.in/index.asp

chetan gavali

@MumbaiPolice

Joined June 2010

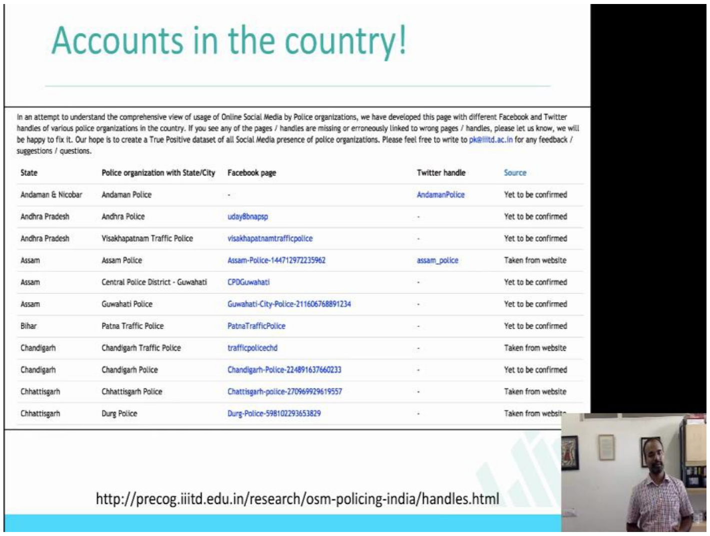
Tweet to chetan gavali

delhi police

@delhipolice2

Here is a quick one in terms of Mumbai Police, at Mumbai cops, it iays protecting and serving Mumbai; keep updated with latest news on how we are combating crime, fearanddisorderinthecommunity. It isnot clear whether it isa legitimateaccount.Andthen there is Chetan Gavali at Mumbai Police, Delhi Police too there is a handle.

(ReferSlideTime:22:47)



Accounts in the country!

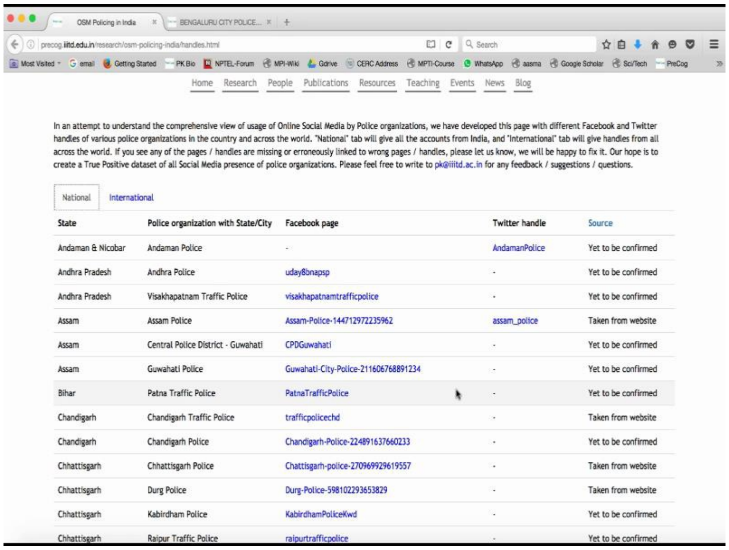
In an attempt to understand the comprehensive view of usage of Online Social Media by Police organizations, we have developed this page with different Facebook and Twitter handles of various police organizations in the country. If you see any of the pages / handles are missing or erroneously linked to wrong pages / handles, please let us know, we will be happy to fix it. Our hope is to create a True Positive dataset of all Social Media presence of police organizations. Please feel free to write to pk@iiitd.ac.in for any feedback / suggestions / questions.

State	Police organization with State/City	Facebook page	Twitter handle	Source
Andaman & Nicobar	Andaman Police	-	AndamanPolice	Yet to be confirmed
Andhra Pradesh	Andhra Police	uday8bnapsp	-	Yet to be confirmed
Andhra Pradesh	Visakhapatnam Traffic Police	visakhapatnamtrafficpolice	-	Yet to be confirmed
Assam	Assam Police	Assam-Police-144712972235962	assam_police	Taken from website
Assam	Central Police District - Guwahati	CPDGuwahati	-	Yet to be confirmed
Assam	Guwahati Police	Guwahati-City-Police-211606768891234	-	Yet to be confirmed
Bihar	Patna Traffic Police	PatnaTrafficPolice	-	Yet to be confirmed
Chandigarh	Chandigarh Traffic Police	trafficpolicechd	-	Taken from website
Chandigarh	Chandigarh Police	Chandigarh-Police-224891637660233	-	Yet to be confirmed
Chhattisgarh	Chhattisgarh Police	Chattisgarh-police-270969929619557	-	Taken from website
Chhattisgarh	Durg Police	Durg-Police-598102293653829	-	Taken from websit

http://precog.iiitd.edu.in/research/osm-policing-india/handles.html

There are multiple handles like this. So, one thing when we were looking at thesehandles we realized is that, hard to find out which is legitimate and which is not legitimate. That is when we created this page in capturing all the State Police Organizations and City Organizations, their Facebook pages, Twitter handles and the source from where we are actually getting this information.

(ReferSlideTime:23:17)



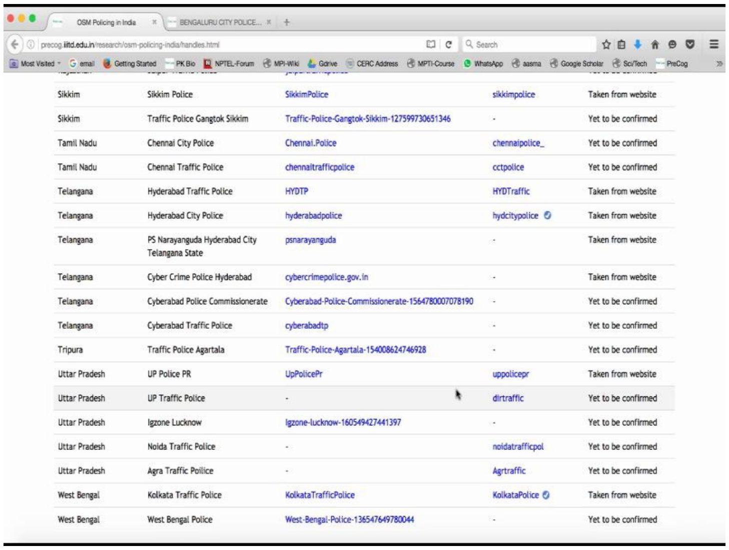
In an attempt to understand the comprehensive view of usage of Online Social Media by Police organizations, we have developed this page with different Facebook and Twitter handles of various police organizations in the country and across the world. "National" tab will give all the accounts from India, and "International" tab will give handles from all across the world. If you see any of the pages / handles are missing or erroneously linked to wrong pages / handles, please let us know, we will be happy to fix it. Our hope is to create a True Positive dataset of all Social Media presence of police organizations. Please feel free to write to pk@iiitd.ac.in for any feedback / suggestions / questions.

National | International

State	Police organization with State/City	Facebook page	Twitter handle	Source
Andaman & Nicobar	Andaman Police	-	AndamanPolice	Yet to be confirmed
Andhra Pradesh	Andhra Police	uday8bnapsp	-	Yet to be confirmed
Andhra Pradesh	Visakhapatnam Traffic Police	visakhapatnamtrafficpolice	-	Yet to be confirmed
Assam	Assam Police	Assam-Police-144712972235962	assam_police	Taken from website
Assam	Central Police District - Guwahati	CPDGuwahati	-	Yet to be confirmed
Assam	Guwahati Police	Guwahati-City-Police-211606768891234	-	Yet to be confirmed
Bihar	Patna Traffic Police	PatnaTrafficPolice	-	Yet to be confirmed
Chandigarh	Chandigarh Traffic Police	trafficpolicechd	-	Taken from website
Chandigarh	Chandigarh Police	Chandigarh-Police-224891637660233	-	Yet to be confirmed
Chhattisgarh	Chhattisgarh Police	Chattisgarh-police-270969929619557	-	Taken from website
Chhattisgarh	Durg Police	Durg-Police-598102293653829	-	Taken from website
Chhattisgarh	Kabirdham Police	KabirdhamPoliceKwd	-	Yet to be confirmed
Chhattisgarh	Raipur Traffic Police	raipurtrafficpolice	-	Yet to be confirmed

So, now let me walk you through the website that I have just now mentioned which is looking at different handles. If you see here, we have been capturing different State Police Organizations, their Facebook pages and their Twitter handle. Source column is basically to show that whether we were able to confirm whether this is the legitimate handle.

(ReferSlideTime:23:48)



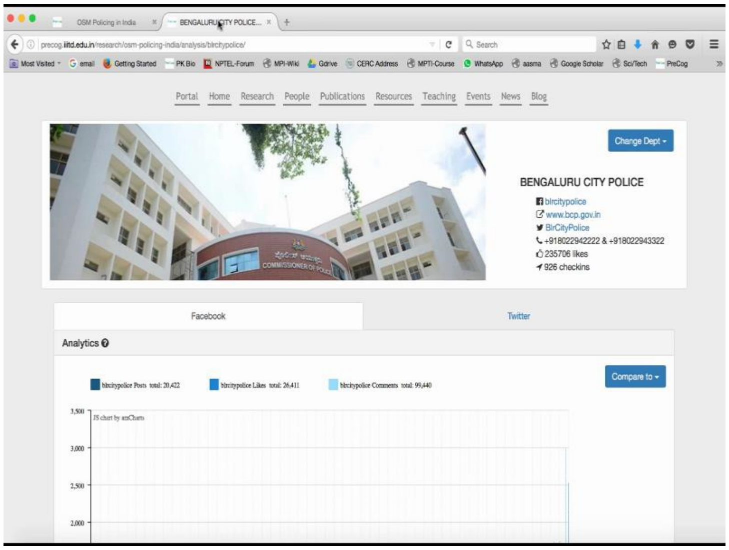
Sikkim	Sikkim Police	SikkimPolice	sikkimpolice	Taken from website
Sikkim	Traffic Police Gangtok Sikkim	Traffic-Police-Gangtok-Sikkim-127599730651346	-	Yet to be confirmed
Tamil Nadu	Chennai City Police	Chennai.Police	chennaipolice_	Yet to be confirmed
Tamil Nadu	Chennai Traffic Police	chennaitrafficpolice	cctpolice	Yet to be confirmed
Telangana	Hyderabad Traffic Police	HYDTP	HYDTraffic	Taken from website
Telangana	Hyderabad City Police	hyderabadpolice	hydcitypolice ✓	Taken from website
Telangana	PS Narayanguda Hyderabad City Telangana State	psnarayanguda	-	Taken from website
Telangana	Cyber Crime Police Hyderabad	cybercrimepolice.gov.in	-	Taken from website
Telangana	Cyberabad Police Commissionerate	Cyberabad-Police-Commissionerate-1564780007078190	-	Yet to be confirmed
Telangana	Cyberabad Traffic Police	cyberabadtp	-	Yet to be confirmed
Tripura	Traffic Police Agartala	Traffic-Police-Agartala-154008624746928	-	Yet to be confirmed
Uttar Pradesh	UP Police PR	UpPolicePr	uppolicepr	Taken from website
Uttar Pradesh	UP Traffic Police	-	dirtraffic	Yet to be confirmed
Uttar Pradesh	Igzone Lucknow	Igzone-lucknow-160549427441397	-	Yet to be confirmed
Uttar Pradesh	Noida Traffic Police	-	noidatrafficpol	Yet to be confirmed
Uttar Pradesh	Agra Traffic Poilice	-	Agrtraffic	Yet to be confirmed
West Bengal	Kolkata Traffic Police	KolkataTrafficPolice	KolkataPolice ✓	Taken from website
West Bengal	West Bengal Police	West-Bengal-Police-136547649780044	-	Yet to be confirmed

Soyoucanclearlyseethereabout 80pages in80accounts, 80rowsinthispage.Youcan clearlyseethatthereareonlyfewofthemwhichareverified, because verified istheonly way to actually confirm that this page is actually legitimately the Police Organization page. And the other way that Police Organizations are doing is to connect the actual website which is Delhi Police dot gov dot in and probably link it to the actual Facebook or Twitter handle from that page. This just helps us to confirm that the legitimate Facebook page or Twitter handle.

You cansee that there areonlyfew ofthemare verified, Kolkata Police, Hyderabad City Police, Bangalore City Police, Bangalore City Traffic, both of them in Facebook and Twitter, and then Delhi Police and CPDelhi. We trykeep this page updated ifand when we get to know that if there is any page which we are missing.

Againif there's any page that you think we are missing please let us now in the forum,we will be actually happy. The idea here is for you to get a sense of how social media is being used in these context. So, keeping the topic of the course in mind in terms of the data collection analysis, let me show you how one can actually collect the data fromthese social networks and actually do some analysis which could be very useful for making some decisions.

(ReferSlideTime:25:12)



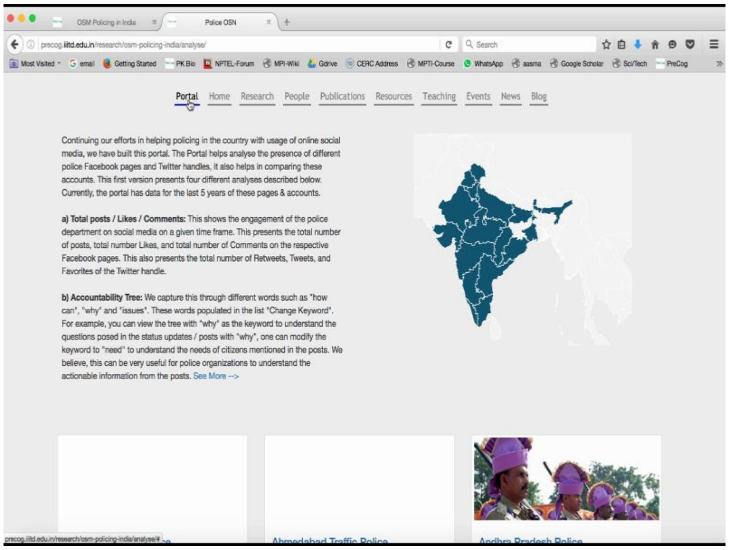
BENGALURU CITY POLICE

- blrcitypolice
- www.bcp.gov.in
- BlrCityPolice
- +918022942222 & +918022943322
- 235706 likes
- 926 checkins

Analytics

blrcitypolice Posts total: 20,422 | blrcitypolice Likes total: 26,411 | blrcitypolice Comments total: 99,440

(ReferSlideTime:25:20)



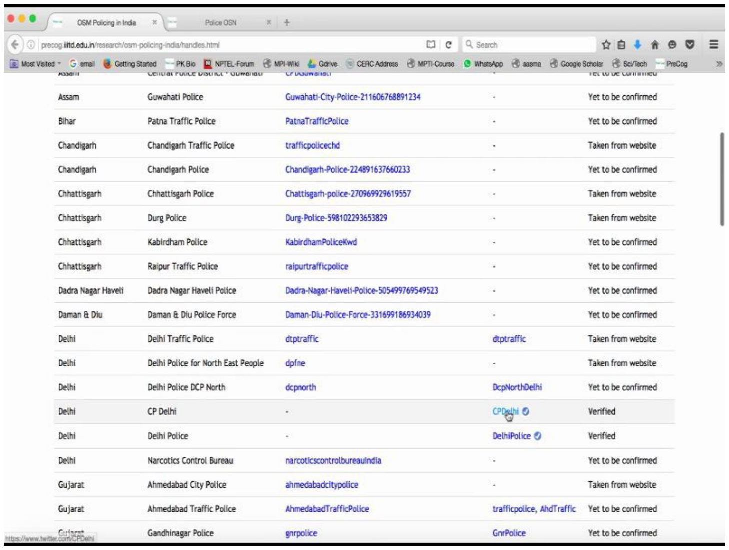
Continuing our efforts in helping policing in the country with usage of online social media, we have built this portal. The Portal helps analyse the presence of different police Facebook pages and Twitter handles, it also helps in comparing these accounts. This first version presents four different analyses described below. Currently, the portal has data for the last 5 years of these pages & accounts.

a) Total posts / Likes / Comments: This shows the engagement of the police department on social media on a given time frame. This presents the total number of posts, total number Likes, and total number of Comments on the respective Facebook pages. This also presents the total number of Retweets, Tweets, and Favorites of the Twitter handle.

b) Accountability Tree: We capture this through different words such as "how can", "why" and "issues". These words populated in the list "Change Keyword". For example, you can view the tree with "why" as the keyword to understand the questions posed in the status updates / posts with "why", one can modify the keyword to "need" to understand the needs of citizens mentioned in the posts. We believe, this can be very useful for police organizations to understand the actionable information from the posts. See More -->

Here isa pagethat wehave set upfor sometime which isbasicallya list ofpagesthat are taken fromthe earlier page handles and then it is not only the link to those pages. So let me show you.

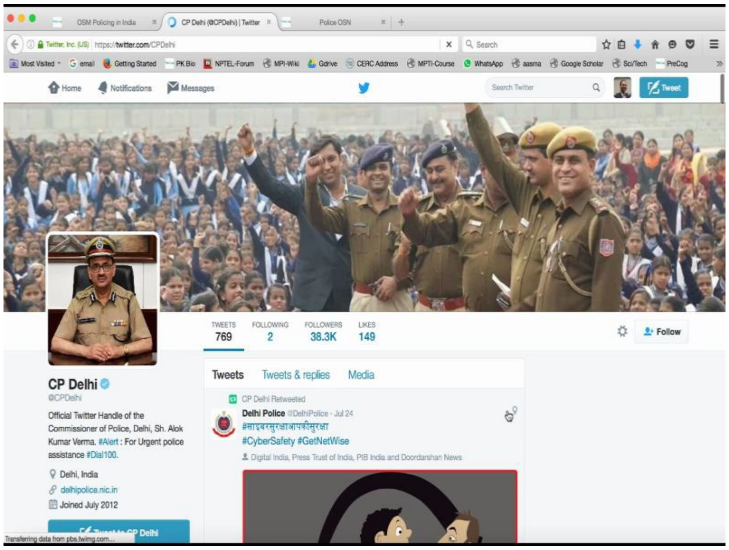
(ReferSlideTime:25:27)



Assam	Central Police District - Guwahati	CPDGuwahati	-	Yet to be confirmed
Assam	Guwahati Police	Guwahati-City-Police-211606768891234	-	Yet to be confirmed
Bihar	Patna Traffic Police	PatnaTrafficPolice	-	Yet to be confirmed
Chandigarh	Chandigarh Traffic Police	trafficpolicechd	-	Taken from website
Chandigarh	Chandigarh Police	Chandigarh-Police-224891637660233	-	Yet to be confirmed
Chhattisgarh	Chhattisgarh Police	Chattisgarh-police-270969929619557	-	Taken from website
Chhattisgarh	Durg Police	Durg-Police-598102293653829	-	Taken from website
Chhattisgarh	Kabirdham Police	KabirdhamPoliceKwd	-	Yet to be confirmed
Chhattisgarh	Raipur Traffic Police	raipurtrafficpolice	-	Yet to be confirmed
Dadra Nagar Haveli	Dadra Nagar Haveli Police	Dadra-Nagar-Haveli-Police-505499769549523	-	Yet to be confirmed
Daman & Diu	Daman & Diu Police Force	Daman-Diu-Police-Force-331699186934039	-	Yet to be confirmed
Delhi	Delhi Traffic Police	dtptraffic	dtptraffic	Taken from website
Delhi	Delhi Police for North East People	dpfne	-	Taken from website
Delhi	Delhi Police DCP North	dcpnorth	DcpNorthDelhi	Yet to be confirmed
Delhi	CP Delhi	-	CPDelhi ✓	Verified
Delhi	Delhi Police	-	DelhiPolice ✓	Verified
Delhi	Narcotics Control Bureau	narcoticscontrolbureauindia	-	Yet to be confirmed
Gujarat	Ahmedabad City Police	ahmedabadcitypolice	-	Taken from website
Gujarat	Ahmedabad Traffic Police	AhmedabadTrafficPolice	trafficpolice, AhdTraffic	Yet to be confirmed
Gujarat	Gandhinagar Police	gnrpolice	GnrPolice	Yet to be confirmed

https://www.twitter.com/CPDelhi

(ReferSlideTime:25:29)



TWEETS 769 FOLLOWING 2 FOLLOWERS 38.3K LIKES 149

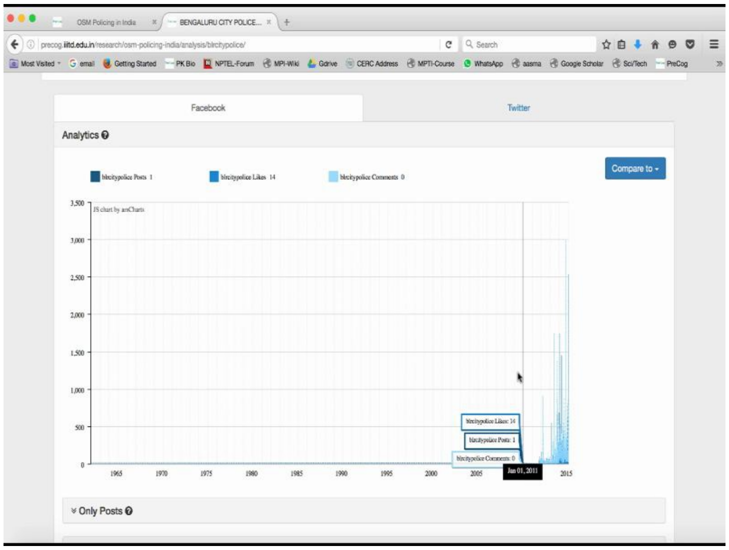
CP Delhi ✓
@CPDelhi
Official Twitter Handle of the Commissioner of Police, Delhi, Sh. Alok Kumar Verma. #Alert : For Urgent police assistance #Dial100.
Delhi, India
delhipolice.nic.in
Joined July 2012

CP Delhi Retweeted
Delhi Police @DelhiPolice · Jul 24
#साइबरसुरक्षाआपकीसुरक्षा
#CyberSafety #GetNetWise
Digital India, Press Trust of India, PIB India and Doordarshan News

Here what we have done is we just have linked to these pages, it just opens up the CP Delhipage for nowfromthis link.Whereas, herewhat we did waswethought we should collect the data and do analysis on the data itself. There is going to capture some things that you have done in the tutorials which is taking data from these networks.

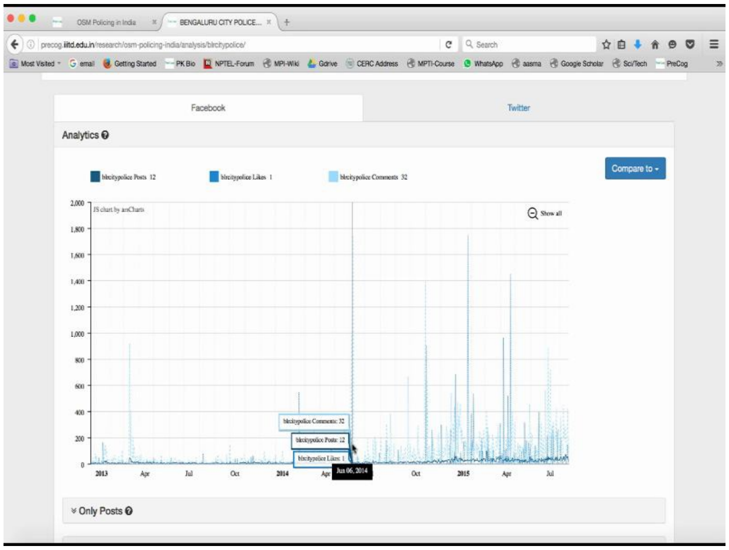
So now let me go to the Karnataka and let me go to Bangalore CityPolice to show you. Essentially the India map on top is capturing all the handles that we have in the database and fromthere if you click ona state it is going to take youto the state and show you all the pages that are there in that state, from there you can go down, go to a specific account. What I am doing here is, specifically looking at a Twitter handle of Bangalore CityPolice, so here isthe citypoliceTwitterhandle and wehave also capturedthe actual website which is dcp dot gov dot in which is useful for us.

(ReferSlideTime:26:28)



Nowwhat we arelooking at isthe Facebookand youcangotoTwitteralso.InFacebook if you remember inthe tutorialwe talked about data collectionand so now collecting the data from these accounts, putting them into the database from the json and doing the graphright. So wearegoing to slicker some other graphs that wecanactuallyprettyuse. Here if you see, you can actually zoom in to the data to see that how these handles have been posting.

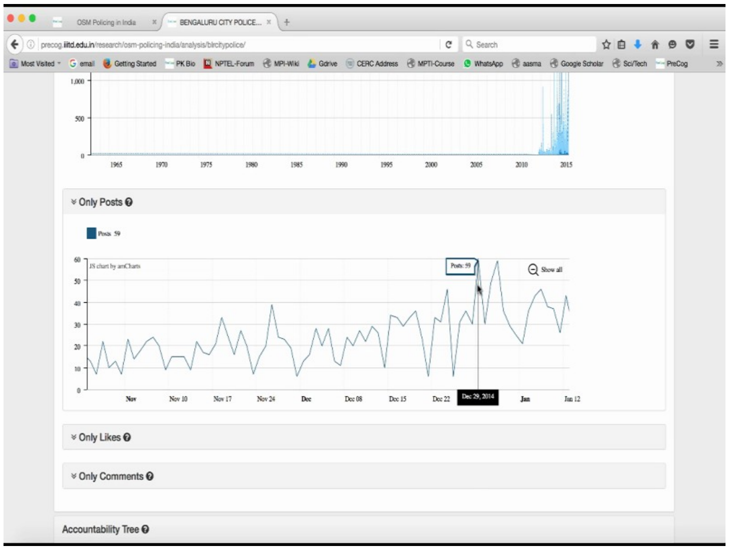
(ReferSlideTime:26:58)



For example here, it says let us look at it here, Bangalore CityPolice, the likes that they got were1 and thenthepost that theydid was12 and the commentsthat theygot was32. This kind of gives the sense of how active these handles are. You could do interesting analysis with these kinds of data from these social networks.

You have enough skills now in the course to collect such data and look at the data also. This week even the course questions that we are going to be looking at for the course is going to be based on something, some data that you can collect and some analysis that you can do. So that gives you a good sense.

(ReferSlideTime:27:44)



Only Posts

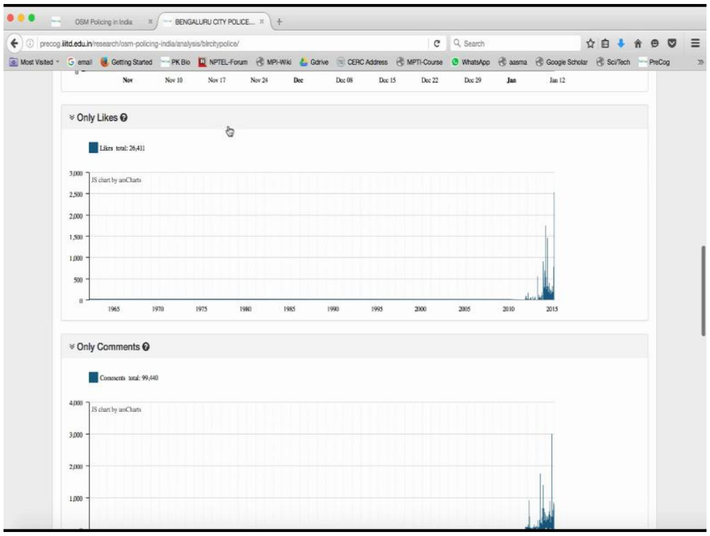
Only Likes

Only Comments

Accountability Tree

Then let us look at only the post analysis, which is the post that they have done but if theydo, what time did theydo, what post you can keep zooming into the datato look at, what kind ofpost, when did they do, what number ofpoststheydo. For example, here itsays in December 29-2014 they did 59 posts.

(ReferSlideTime:28:10)



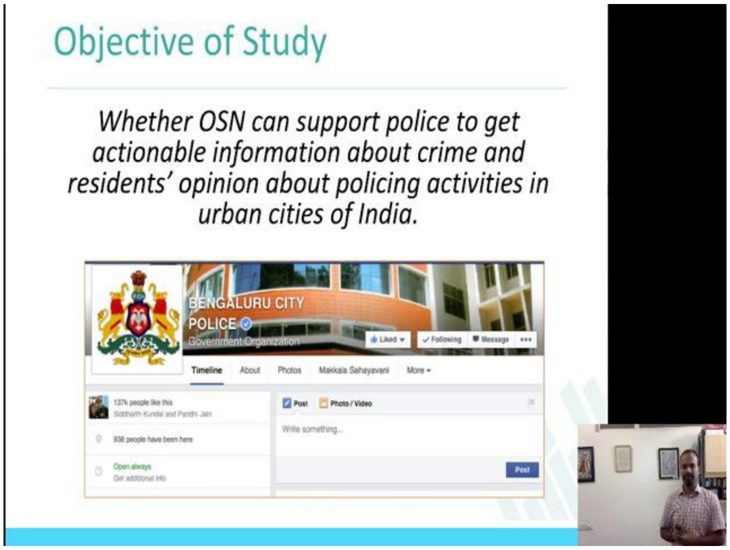
Only Likes

Only Comments

Similarly, you could look at likes and comments also. This kind of gives you a good sense of how these Police Organizations are using the data, how this data can be collected, what analysis can be done, and I am sure you can do more analysis with this data also. I am going to be also specifically talking about one specific set of questions that we were trying to answer with the data that we collected from this social media services for the Police Organizations.

Policing and Online Social Media Part-II

(ReferSlideTime:00:30)



Objective of Study

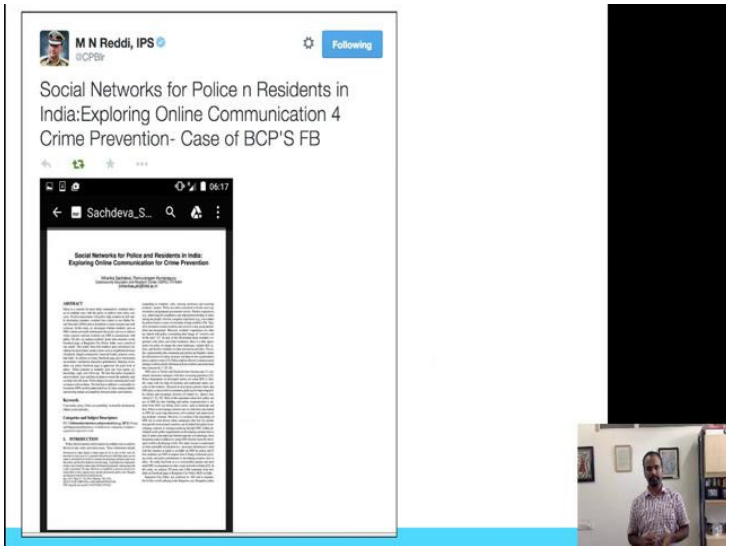
Whether OSN can support police to get actionable information about crime and residents' opinion about policing activities in urban cities of India.

Welcome back. So, let me now talk a little bit about the specifics of how the data from police organizations can be collected, and what kind of analysis can be done to find out some interesting things.

Here is one research question; here is one question that you can think about - objectives of the study. And then I am going to be taking about whether online social media can support police to get actionable information about crime and residents' opinion about policing activities in urban cities yeah, so that is the goal. So, let us try and, see if youcanactuallyteachthisobjectiveto studysomedata fromFacebook andTwitterand make some useful inferences.

So, let me just breakthisobjective into pieces, which is, canweuse Facebookto support policetogetactionableinformation?Whatisanactionableinformation,actionable information is something like do this, can you actually get this done, I mean I am having a problem inthe streetthat is traffic issues inthe roadthere is a pot hole which is broken on this street, a car broken down. So, these are actionable information that police organizations cantake fromthe post and that is actuallyuseful for decision making.And residents'opinion, ofcourse, what people think about police, what arethey talking about police is also useful information for police organizations.

(ReferSlideTime:01:46)



M N Reddi, IPS
@CPBlr
Following

Social Networks for Police n Residents in India:Exploring Online Communication 4 Crime Prevention- Case of BCP'S FB

06:17
Sachdeva_S...

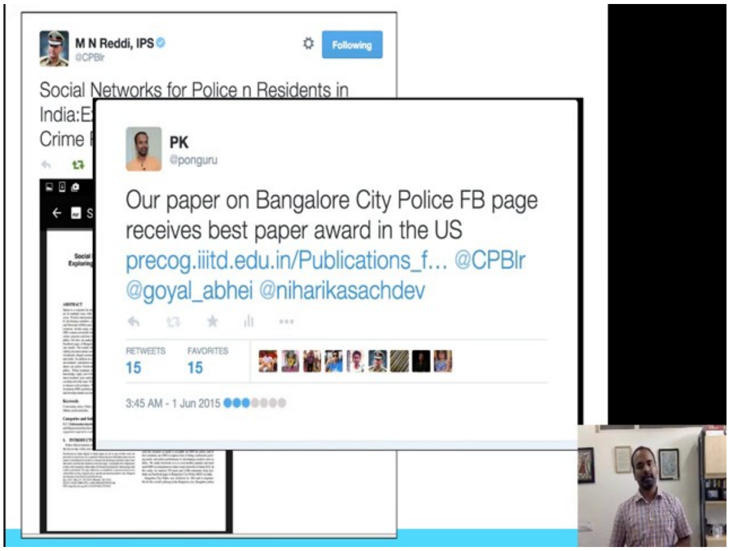
Social Networks for Police and Residents in India: Exploring Online Communication for Crime Prevention

So, before I get into further I think there was a question in the forum asking aboutwhatisre-identification. IthoughtIwillactuallymentionit here ratherthanactuallypacking it in the forum itself. So, re-identification is nothing but, take some information, you want to,actually, yougot ainformationabout PK,thereissomepubliclyavailable information which has no reference to PK. For example, if you remember the LatanyaSweeny(Refer Time: 02:14) slide where we talked about voters record and medical record. Justin medical records they are actually identifiable. Just in voter records also they are identifiable. If we put together the identification actually becomes much stronger, youare able to uniquely identify more people with more data put together.

For example, again, let me go to myown example, you cantake some publiclyavailable information about me on some websites. Say oh faculty at IIIT and things like that. And you go back to Facebook, and then you use the Facebook pictures that are publicly available about me, take those pictures connectit with these posts you can say it oh this is actually PK, this is how I willalso I mean insides your faculty and IIIT, NPTEL.

So, re-identification of information of a particular individual, of a particular thing is actually the concept that we discussed last week. I hope that makes it clear which is unidentified datasetswhich inthe class that wetalked about max.comand identified data sets we stored which where there is if your time you know one can find out the this is here. So, taking some unidentified data and using some identified data putting them together and identifying the users actually is an (Refer Time: 03:26).

(ReferSlideTime:03:48)



M N Reddi, IPS
@CPBlr
Following

Social Networks for Police n Residents in India:E
Crime

PK
@ponguru

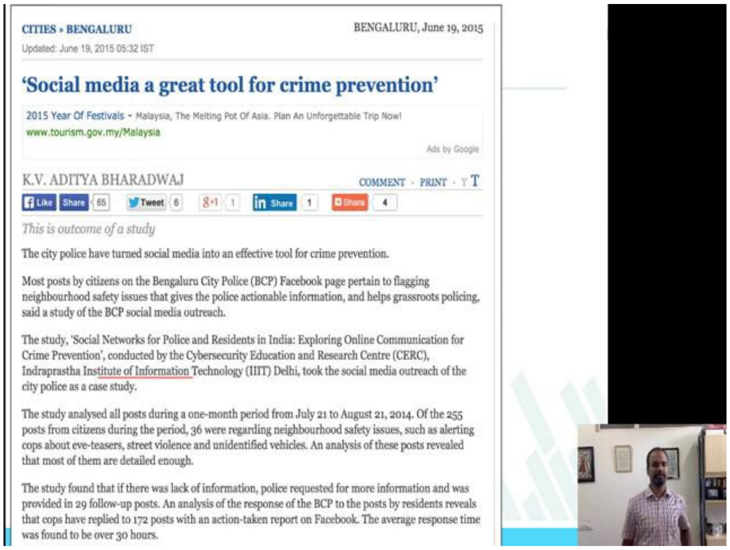
Our paper on Bangalore City Police FB page receives best paper award in the US precog.iiitd.edu.in/Publications_f... @CPBlr @goyal_abhei @niharikasachdev

RETWEETS 15 FAVORITES 15

3:45 AM - 1 Jun 2015

So, wedidthisworkonsocialnetworksforpolice andresidents inIndiaexploringonline communications. So, this is the paper that with I am going to be talking about, but this actually more than a paper that the data that I will be talking about right now.

(ReferSlideTime:03:47)



CITIES » BENGALURU

BENGALURU, June 19, 2015

Updated: June 19, 2015 05:32 IST

'Social media a great tool for crime prevention'

2015 Year Of Festivals - Malaysia, The Melting Pot Of Asia. Plan An Unforgettable Trip Now!
www.tourism.gov.my/Malaysia

Ads by Google

K.V. ADITYA BHARADWAJ

This is outcome of a study

The city police have turned social media into an effective tool for crime prevention.

Most posts by citizens on the Bengaluru City Police (BCP) Facebook page pertain to flagging neighbourhood safety issues that gives the police actionable information, and helps grassroots policing, said a study of the BCP social media outreach.

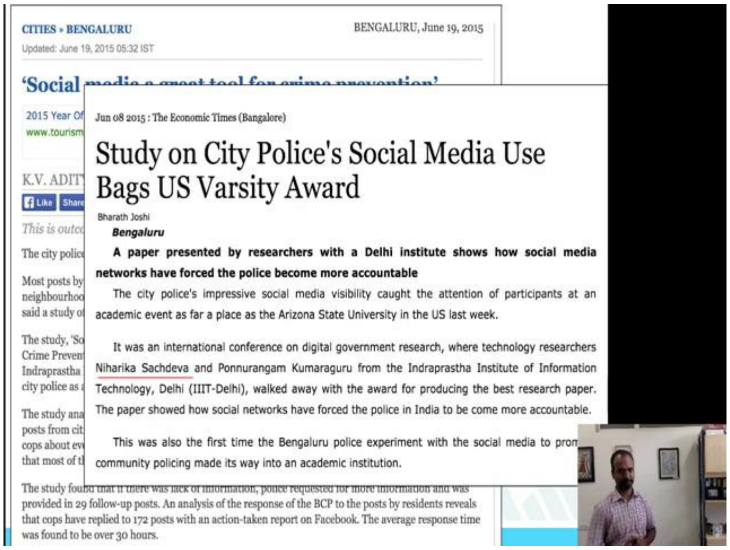
The study, 'Social Networks for Police and Residents in India: Exploring Online Communication for Crime Prevention', conducted by the Cybersecurity Education and Research Centre (CERC), Indraprastha Institute of Information Technology (IIIT) Delhi, took the social media outreach of the city police as a case study.

The study analysed all posts during a one-month period from July 21 to August 21, 2014. Of the 255 posts from citizens during the period, 36 were regarding neighbourhood safety issues, such as alerting cops about eve-teasers, street violence and unidentified vehicles. An analysis of these posts revealed that most of them are detailed enough.

The study found that if there was lack of information, police requested for more information and was provided in 29 follow-up posts. An analysis of the response of the BCP to the posts by residents reveals that cops have replied to 172 posts with an action-taken report on Facebook. The average response time was found to be over 30 hours.

So, in general, actually in the last two years or so social media has been used for crime prevention. It can be used effectively for finding out what people are saying, you can actually collect the information lot of things about what is going on in the society, because it is going to be a very hard to have police organization, police personnel ateverygivenpoint intime, at anygivenpoint inthe societyalso. So you canactuallyget a lot of information frompublic throughthe social networks, which can be used to prevent crime. So, essentially you can build societies which are safer if were to actually analyze use social media services.

(ReferSlideTime:04:26)



CITIES » BENGALURU

BENGALURU, June 19, 2015

Updated: June 19, 2015 05:32 IST

‘Social

2015 Year Of
www.tourism

K.V. ADIT

Like Share

This is outco

The city police

Most posts by
neighbourhoo
said a study o

The study, ‘So
Crime Preven
Indraprastha
city police as

The study ana
posts from cit
cops about ev
that most of t

Jun 08 2015 : The Economic Times (Bangalore)

Study on City Police's Social Media Use Bags US Varsity Award

Bharath Joshi

Bengaluru

A paper presented by researchers with a Delhi institute shows how social media networks have forced the police become more accountable

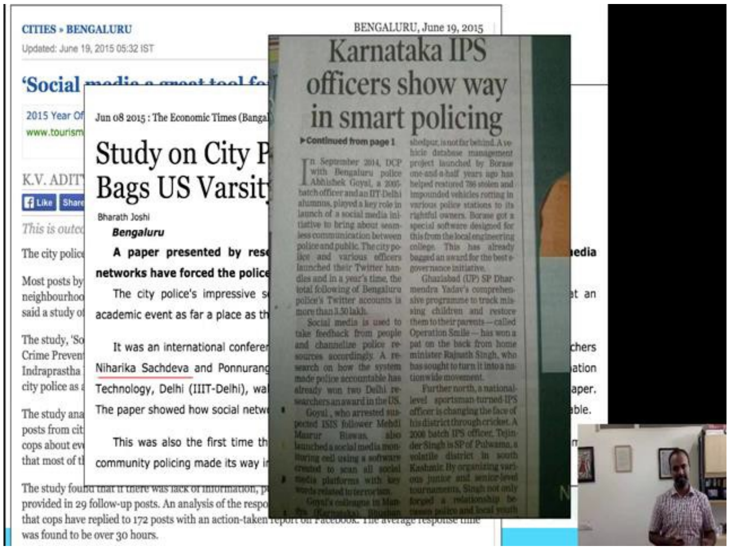
The city police's impressive social media visibility caught the attention of participants at an academic event as far a place as the Arizona State University in the US last week.

It was an international conference on digital government research, where technology researchers Niharika Sachdeva and Ponnurangam Kumaraguru from the Indraprastha Institute of Information Technology, Delhi (IIIT-Delhi), walked away with the award for producing the best research paper. The paper showed how social networks have forced the police in India to be come more accountable.

This was also the first time the Bengaluru police experiment with the social media to prom community policing made its way into an academic institution.

The study fou provided in 29 follow-up posts. An analysis of the response of the BCP to the posts by residents reveals that cops have replied to 172 posts with an action-taken report on Facebook. The average response time was found to be over 30 hours.

(ReferSlideTime:04:30)



CITIES » BENGALURU

BENGALURU, June 19, 2015

Updated: June 19, 2015 05:32 IST

Jun 08 2015 : The Economic Times (Bangal

Study on City P
Bags US Varsit

Bharath Joshi

Bengaluru

Karnataka IPS officers show way in smart policing

▶Continued from page 1

In September 2014, DCP with Bengaluru police Abhishek Goyal, a 2005-batch officer and an IIT-Delhi alumnus, played a key role in launch of a social media initiative to bring about seamless communication between police and public. The city police and various officers launched their Twitter handles and in a year's time, the total following of Bengaluru police's Twitter accounts is more than 3.50 lakh.

Social media is used to take feedback from people and channelize police resources accordingly. A research on how the system made police accountable has already won two Delhi researchers an award in the US.

Goyal, who arrested suspected ISIS follower Mehdi Masrur Biswas, also launched a social media monitoring cell using a software created to scan all social media platforms with key words related to terrorism.

Goyal's colleague in Mandya (Karnataka), Bhushan

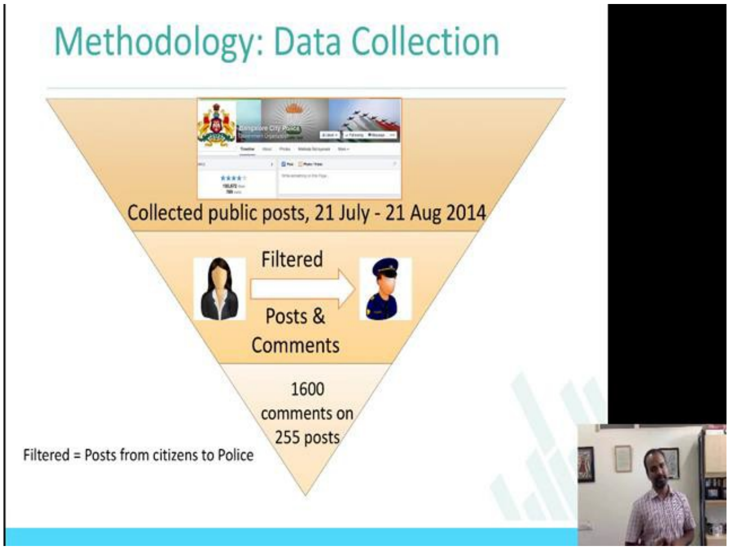
shedpur, is not far behind. A vehicle database management project launched by Borase one-and-a-half years ago has helped restored 786 stolen and impounded vehicles rotting in various police stations to its rightful owners. Borase got a special software designed for this from the local engineering college. This has already bagged an award for the best e-governance initiative.

Ghaziabad (UP) SP Dharmendra Yadav's comprehensive programme to track missing children and restore them to their parents — called Operation Smile — has won a pat on the back from home minister Rajnath Singh, who has sought to turn it into a nationwide movement.

Further north, a national-level sportsman-turned-IPS officer is changing the face of his district through cricket. A 2008 batch IPS officer, Tejinder Singh is SP of Pulwama, a volatile district in south Kashmir. By organizing various junior and senior-level tournaments, Singh not only forged a relationship between police and local youth

So,intermsofactuallythethemeitselfthedatathatwelookedatisactuallyfrom Bangalore, Karnataka.

(ReferSlideTime:04:38)



Methodology: Data Collection

Collected public posts, 21 July - 21 Aug 2014

Filtered

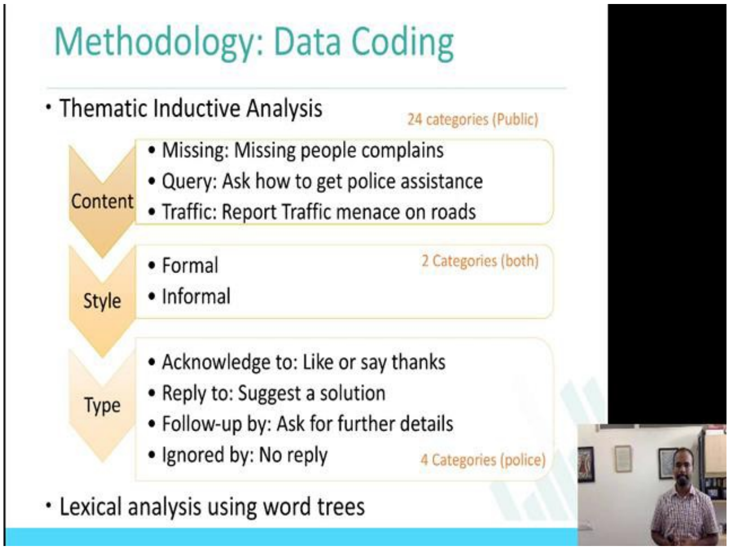
Posts & Comments

1600 comments on 255 posts

Filtered = Posts from citizens to Police

So, in terms of methodology, what kind of data did we collect? So, keeping the goal for studying whether we can actually collect actionable information from social media we started looking at this data, we collected the data from the Facebook page of Bangalore City Police in 2014. Looking at what are the posts that was done. And we filtered the posts and comments, because we wanted to studywhat public said to the police in terms of what post that they did, what comments that they say on the Facebook page. Andabout 1600 comments and 255 posts were actually collected.

(ReferSlideTime:05:19)



Methodology: Data Coding

- Thematic Inductive Analysis

24 categories (Public)

Content
- Missing: Missing people complains
- Query: Ask how to get police assistance
- Traffic: Report Traffic menace on roads

Style
- Formal
- Informal

2 Categories (both)

Type
- Acknowledge to: Like or say thanks
- Reply to: Suggest a solution
- Follow-up by: Ask for further details
- Ignored by: No reply

4 Categories (police)

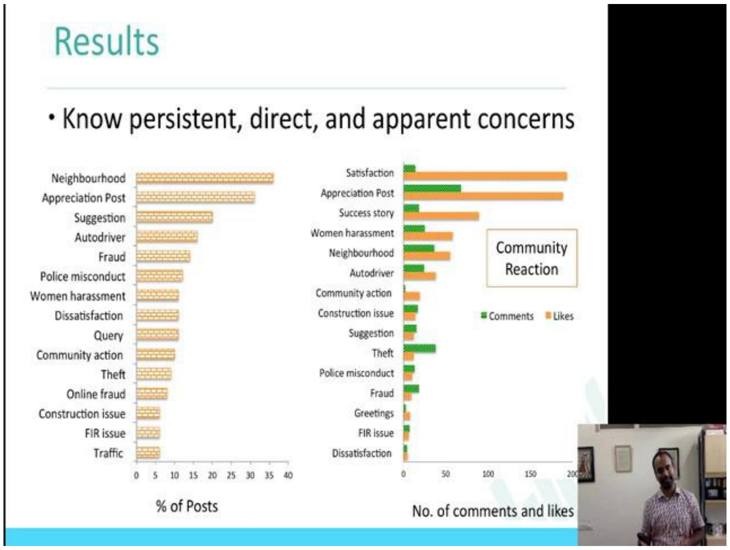
- Lexical analysis using word trees

So, interms of methodology, there are actually multiple ways the people actuallylook at this data type. We are looking at the post and we are looking at the comments we can analyze indifferent ways. So,oneapproachthat wetookwasfindingout what peopleare talkingaboutwhichis misinformation,query,trafficdetailsthat isaboutthecontent.And then we looked at for the style of I think which is formal or informal.

And interms oftypes ofpostthat wereshowing up which is acknowledge to, like, or say thanks, reply to, such suggest a solution and the follow up by asking further details, ignored byno reply, because these are the ones that are coming fromthe police side. So, citizens post and what do police do about it.

So, if you look at in the right hand side, it is says twenty four categories for the public post and two categories for the style, and four categories of the police responses. So, again given that we are talking about content and injecting some analysis that you could do with the data yourself also in terms of lexical analysis in terms of actually (Refer Time: 06:27) the content itself that is from the post.

(ReferSlideTime:06:33)



Results

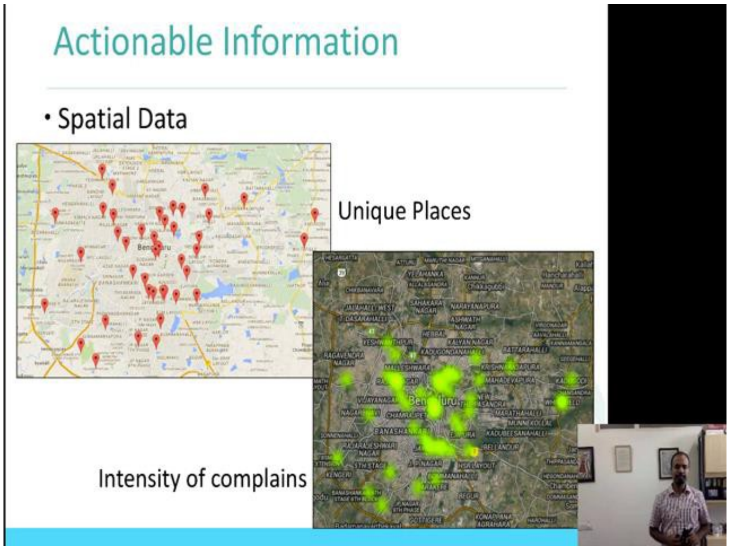
- Know persistent, direct, and apparent concerns

So, if you look at the results, some of the results are very interesting in terms of what kind of post were done by citizens for these on this page. Majority of the things were actually from the neighborhood concerns right. Then it is appreciation which is talking about thanks to police and appreciating the things that police does. And it kind of goesdown. suggestions, auto driver related, fraud, till traffic issues.

And, if you look at the comments and the likes, the comments for actions like appreciation are actually higher than the comments for satisfaction; whereas, if you look at the likes, the likes for satisfaction, appreciation and success stories are actually veryhighe. It is probably very intuitive that (Refer Time: 07:20) how the police post gets reactions fromthe society, the likes are actually pretty high for satisfaction, appreciation post and for success stories compared to some of the other ones.

So,this isgives you asenseoftheanalysisthat you cando withanykind ofdatathat you collect.You remember we talked about Bostonblasts and Hurricane Sandy (Refer Time: 07:44) and those kind of events in the context of credibility and trust (Refer Time:07:50). So here, we aredoing these similar kinds of analysis, similar kinds of questions that we are asking, but we are actually using different sets of data and different kinds of graphsthatweareproducing.So,thiswillhelpyoutogetasenseofwhatarekindof posts are actually showing up on these pages, and what kind of reactions are being seen on for these posts also.

(ReferSlideTime:08:13)

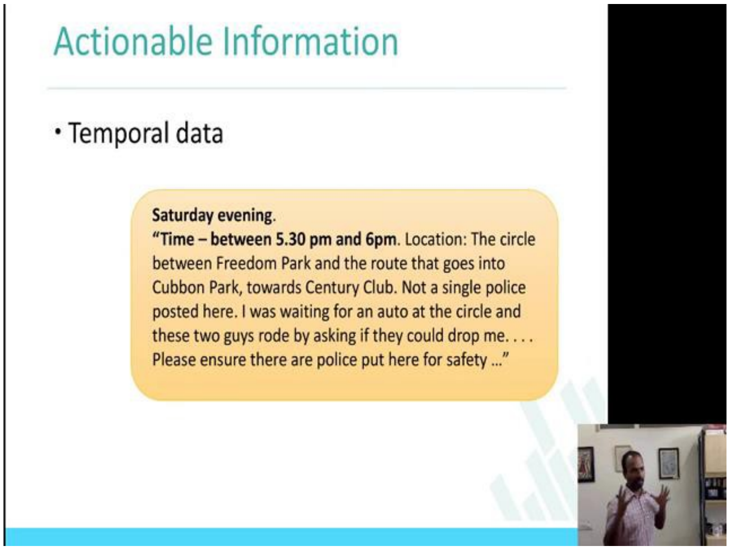


Actionable Information

- Spatial Data

Similar to the analysis that we did in Boston and Hurricane Sandy, we can do the geospatial analysis also with this data here (Refer Time: 08:22). The one on the top is showing you the poststhat arecoming fromthe different partsofBangalore forthe posts that we saw in the page.And ofcourse, one could do some heat map, one could find out where are the places fromwhere majorityofthe posts are coming and you could use that for making some decisions.

(ReferSlideTime:08:44)



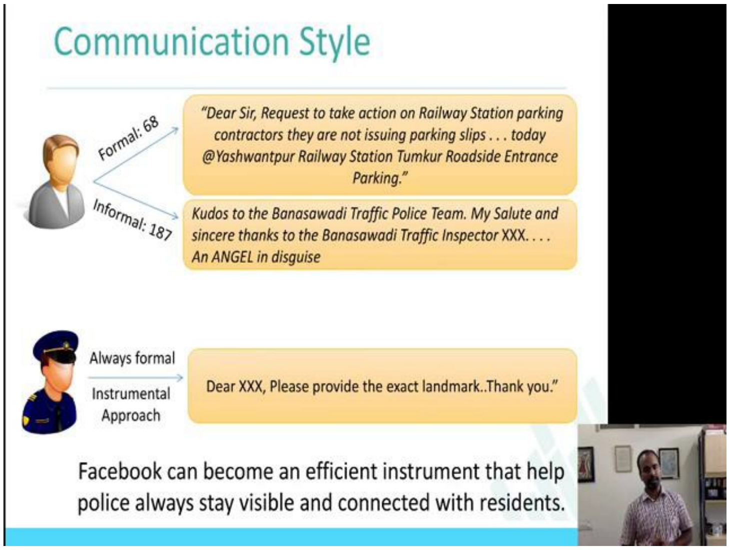
Actionable Information

- Temporal data

Saturday evening.
"Time – between 5.30 pm and 6pm. Location: The circle between Freedom Park and the route that goes into Cubbon Park, towards Century Club. Not a single police posted here. I was waiting for an auto at the circle and these two guys rode by asking if they could drop me. . . . Please ensure there are police put here for safety ..."

So, given the goal was actionable information, we were actually focused on finding out fromthe content what kind ofinformationcan be drawn. So, here is one post whichtalks about temporaldata at least which can be drawn, (Refer Time: 09:00)time between 5.30 and 6 pm. Location, blah, blah, blah, not a single police postedhere,I waswaiting for an auto at the circle blah, blah, blah right. So, this gives them, this gives the police organizations a good sense of what time is it, what location is it, what should be done, what is the problem, it is easy to actually collect this information. If it was not given in this form, if this information was not there, the police has to actually ask saying what location is it, what time is it, and things like that.

(ReferSlideTime:09:31)



Communication Style

Formal: 68

"Dear Sir, Request to take action on Railway Station parking contractors they are not issuing parking slips . . . today @Yashwantpur Railway Station Tumkur Roadside Entrance Parking."

Informal: 187

Kudos to the Banasawadi Traffic Police Team. My Salute and sincere thanks to the Banasawadi Traffic Inspector XXX. . . . An ANGEL in disguise

Always formal

Instrumental Approach

Dear XXX, Please provide the exact landmark..Thank you."

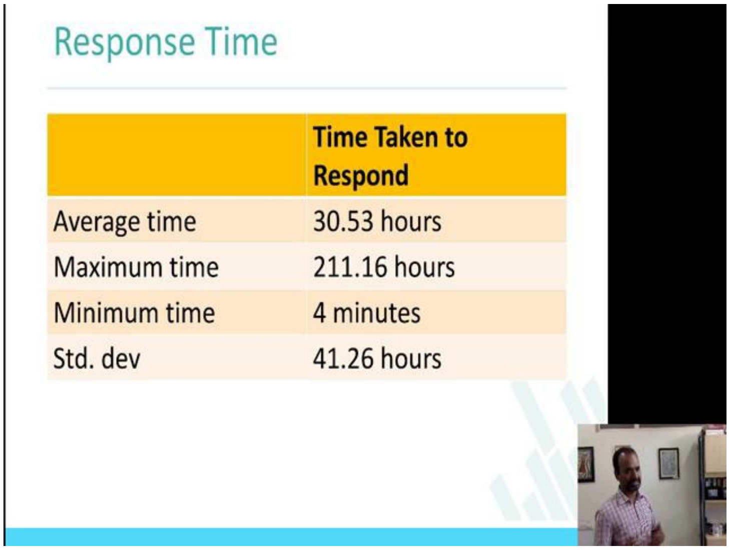
Facebook can become an efficient instrument that help police always stay visible and connected with residents.

So, if you look at the communication style, the style is also interesting (Refer Time: 09:37) that lot ofdiscussions that happen on from the police side is actually very formal; formal versus in informal. Dear Sir, Request to take action on Railway Station this is from the citizens, parking contractors they are not issuing parking slips right. Kudos to the BanasawadiTraffic Police Team. My salute and this is post for appreciation.

So, from the police if it comes, it is (Refer Time: 10:07) almost going to be always formal.And stayvisible ofcourse,this isthe point Ihave said earlier, which isFacebook and any social network for that matter can become the way by which police can actually connect with society most strongly.

And I am sure as you are going through the course you will also start looking at, I hope you will also start looking at the police pages of a local city from your location are actually start saying, what kind of post they are doing, what kind of things that they are looking for, what kind of interactions are theyhaving. So, the whole bodyofknowledge, body of research, body of work is to actually look at increase the community policing right. So, you can actually increase the interactions with the society to get more information from them.

(ReferSlideTime:10:53)

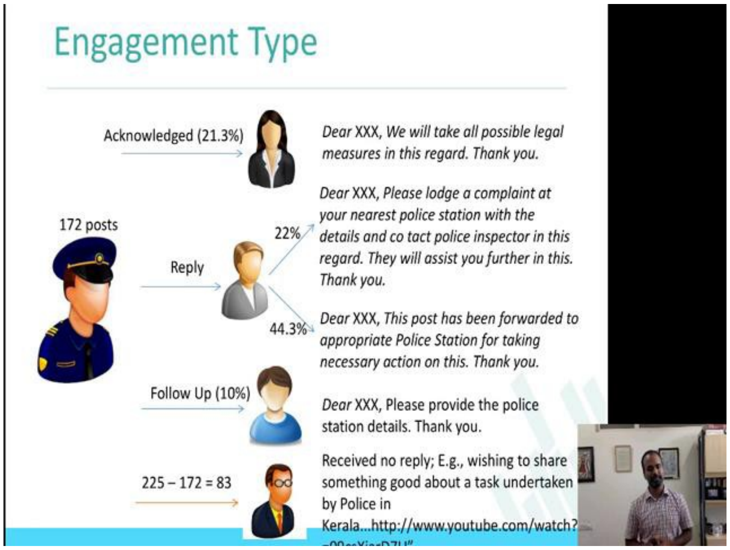


Response Time

	Time Taken to Respond
Average time	30.53 hours
Maximum time	211.16 hours
Minimum time	4 minutes
Std. dev	41.26 hours

Of course, these are some details, I will go through them slightly quickly. Average time response 30 hours, maximum time was 211 hours, minimum time was about 4 minutes. Showing that there is large variance in terms of actually responses that they get 4minutesto 211hours, sothatisa lotofdifference intermsofthetimesresponsethat they get.

(ReferSlideTime:11:18)



Engagement Type

Dear XXX, We will take all possible legal measures in this regard. Thank you.

Dear XXX, Please lodge a complaint at your nearest police station with the details and co tact police inspector in this regard. They will assist you further in this. Thank you.

Dear XXX, This post has been forwarded to appropriate Police Station for taking necessary action on this. Thank you.

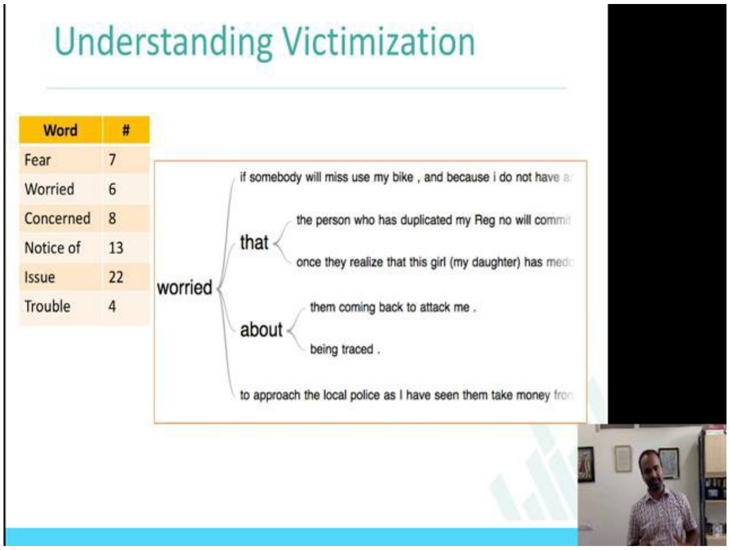
Dear XXX, Please provide the police station details. Thank you.

Received no reply; E.g., wishing to share something good about a task undertaken by Police in Kerala...http://www.youtube.com/watch?

So, herearethedifferent typesofpost that come frompoliceand thekind ofengagement that they have. So, acknowledged 21 percent of the post police actually acknowledged. Dear XXX, we will take all possible legal measures in this regard, thank you. And as a reply Dear XXX, Please lodge a complaint – 22 percent. And Dear XXX, This post has been forwarded to appropriate police station. And about follow up, Dear XXX, Please provide the police station details. Thank you.

So, this kind tells you what kind of interactions of police organizations having. Andabout anything postednot have getting response. So, the goalis to find out theoneofthe one ofthe interesting questions that you could also think about is how to actually have a post whichwill have the response frompolice, that would be also an interesting question to look at.

(ReferSlideTime:12:20)



Understanding Victimization

Word	#
Fear	7
Worried	6
Concerned	8
Notice of	13
Issue	22
Trouble	4

worried

if somebody will miss use my bike , and because i do not have a

that

the person who has duplicated my Reg no will comm

once they realize that this girl (my daughter) has med

about

them coming back to attack me .

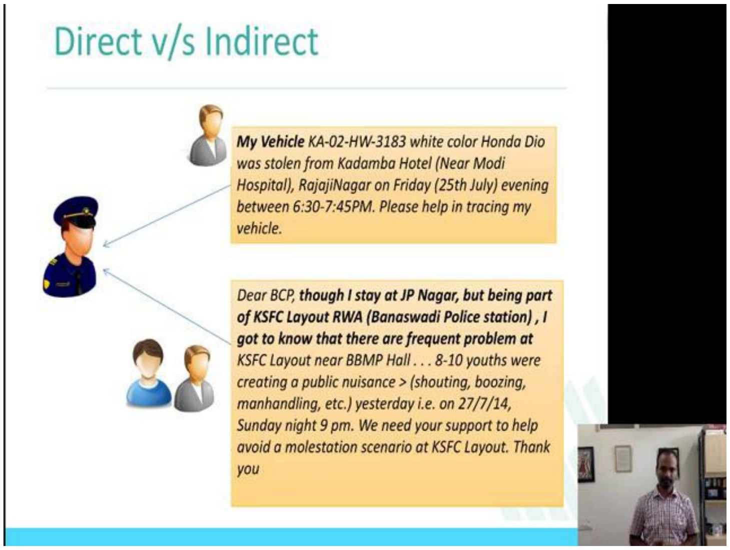
being traced .

to approach the local police as I have seen them take money fro

So, if you look at the concept of finding out what citizens are worried about, what citizens are talking about, I will go through some tweets what we so to say in terms of actually looking at what poststhe citizens are doing, how we can actually take out some useful information from these posts. So, in this case, we aretalking about worried as the starting pointwhich isfrom the postyou can actually look at worried,if somebody will misuse my bike, worried at the person who is duplicated my registration number will commit, blah, blah, blah, worried about they coming back to attack me.

Forinstance, thisconceptofidentifyingthecontent,textualcontent andseeingwhat kind of posts that citizens are doing can be extremely useful. If only if one can generate these trees in real time it can be very useful for police to make some decisions.And if this can be done in real time to showing up, oh, currently there is a post on Facebook which hasactionable information and the actionable information is the time, the details and this citizen’s post is actually having about worried about few things can be veryuseful.

(ReferSlideTime:13:37)



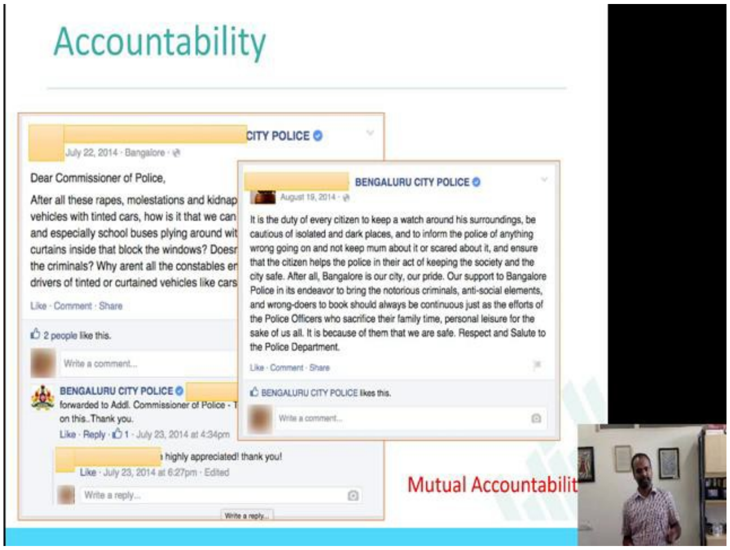
Direct v/s Indirect

***My Vehicle** KA-02-HW-3183 white color Honda Dio was stolen from Kadamba Hotel (Near Modi Hospital), RajajiNagar on Friday (25th July) evening between 6:30-7:45PM. Please help in tracing my vehicle.*

*Dear BCP, **though I stay at JP Nagar, but being part of KSFC Layout RWA (Banaswadi Police station) , I got to know that there are frequent problem at** KSFC Layout near BBMP Hall . . . 8-10 youths were creating a public nuisance > (shouting, boozing, manhandling, etc.) yesterday i.e. on 27/7/14, Sunday night 9 pm. We need your support to help avoid a molestation scenario at KSFC Layout. Thank you*

Of course, the direct versus indirect information drawing from this post - Direct information, it going on with me, I am actually going through the problem or I am actually part of the situation that I am talking about. Sometimes it could be indirectwhich is 'Dear BCP, though I stayat JPNagar, but being part ofKSFC layout near blah, blah, blah. Iamnot fromthere, but Isee aproblemthere, so I’mletting you know (Refer Time:14:05)It could bethat myfriendsaysthis,myfriend livesthereontheregardingly I post, post on Facebook, I do a post on Facebook about the friend that who lives in a different location not about myself.

(ReferSlideTime:14:26)



Accountability

[redacted] CITY POLICE
July 22, 2014 · Bangalore ·

Dear Commissioner of Police,

After all these rapes, molestations and kidnap
vehicles with tinted cars, how is it that we can
and especially school buses plying around wit
curtains inside that block the windows? Doesr
the criminals? Why arent all the constables er
drivers of tinted or curtained vehicles like cars

Like · Comment · Share

2 people like this.

Write a comment...

BENGALURU CITY POLICE [redacted]
forwarded to Addl. Commissioner of Police - T
on this..Thank you.
Like · Reply · 1 · July 23, 2014 at 4:34pm

[redacted] highly appreciated! thank you!
Like · July 23, 2014 at 6:27pm · Edited

Write a reply...

Write a reply...

[redacted] BENGALURU CITY POLICE
August 19, 2014 ·

It is the duty of every citizen to keep a watch around his surroundings, be cautious of isolated and dark places, and to inform the police of anything wrong going on and not keep mum about it or scared about it, and ensure that the citizen helps the police in their act of keeping the society and the city safe. After all, Bangalore is our city, our pride. Our support to Bangalore Police in its endeavor to bring the notorious criminals, anti-social elements, and wrong-doers to book should always be continuous just as the efforts of the Police Officers who sacrifice their family time, personal leisure for the sake of us all. It is because of them that we are safe. Respect and Salute to the Police Department.

Like · Comment · Share

BENGALURU CITY POLICE likes this.

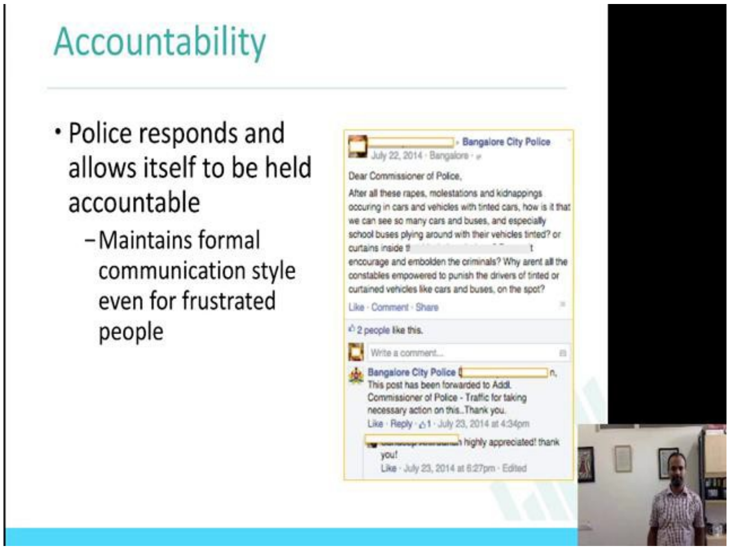
Write a comment...

Mutual Accountabilit

So, there are recent posts also which is directly about myself and indirectly about somebody else. And of course, being it is probably very intuitive or to realize that you could take the content from the social networks, and actually took for accountability of both sides for example, accountabilityofpolice and accountabilityofcitizens also. How we candothat we could lookatthe post and see howfast theyare responding, what kind of responses they are it is coming and how citizens are also responding to these queries that the police is making.

So, accountability can be good question to ask from the post that is collected from the social network. And of course, the little that we have seen little that is being looked at there is also mutual accountability that is going on citizens think that police should be doing and (Refer Time: 15:21) police think that citizens should be doing it, there is a accountability, in terms of, because this platforms publicly available.

(ReferSlideTime:15:24)



Accountability

- Police responds and allows itself to be held accountable
  - Maintains formal communication style even for frustrated people

[redacted] › Bangalore City Police
July 22, 2014 · Bangalore ·

Dear Commissioner of Police,

After all these rapes, molestations and kidnappings occuring in cars and vehicles with tinted cars, how is it that we can see so many cars and buses, and especially school buses plying around with their vehicles tinted? or curtains inside t[redacted]t encourage and embolden the criminals? Why arent all the constables empowered to punish the drivers of tinted or curtained vehicles like cars and buses, on the spot?

Like · Comment · Share

2 people like this.

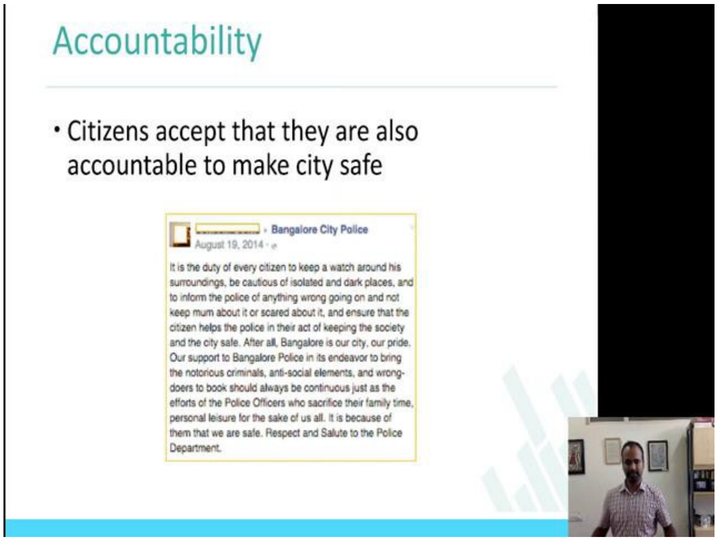
Write a comment...

Bangalore City Police [redacted]n, This post has been forwarded to Addl. Commissioner of Police - Traffic for taking necessary action on this..Thank you.
Like · Reply · 1 · July 23, 2014 at 4:34pm

[redacted] highly appreciated! thank you!
Like · July 23, 2014 at 6:27pm · Edited

And ofcourse, police organizations respond to this post, and request for informationand follow up onthings also making themselves accountable for the activitythat is going on.

(ReferSlideTime:15:43)



Accountability

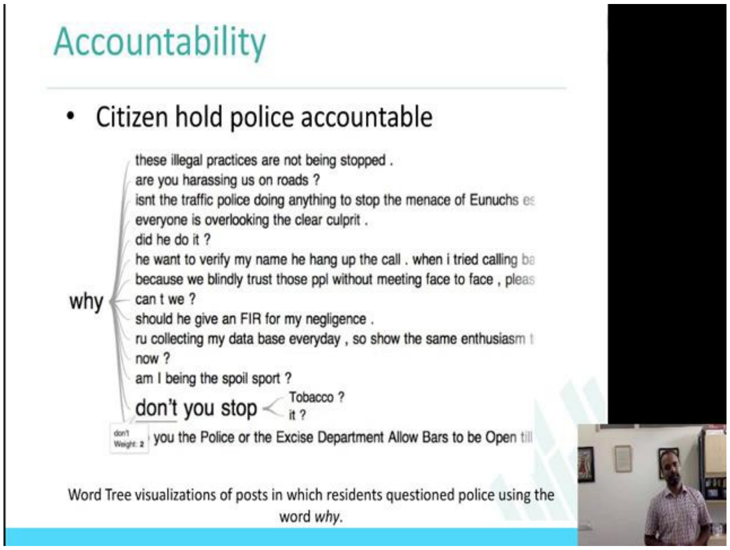
- Citizens accept that they are also accountable to make city safe

[redacted] › Bangalore City Police
August 19, 2014 ·

It is the duty of every citizen to keep a watch around his surroundings, be cautious of isolated and dark places, and to inform the police of anything wrong going on and not keep mum about it or scared about it, and ensure that the citizen helps the police in their act of keeping the society and the city safe. After all, Bangalore is our city, our pride. Our support to Bangalore Police in its endeavor to bring the notorious criminals, anti-social elements, and wrong-doers to book should always be continuous just as the efforts of the Police Officers who sacrifice their family time, personal leisure for the sake of us all. It is because of them that we are safe. Respect and Salute to the Police Department.

A citizen accepts that they are also accountable to make the city safe. Citizens also believethattheyshouldbeparticipating intheseactivities intermsofposting, interacting with police, giving information and making sure that the city is safe.

(ReferSlideTime:16:01)



Accountability

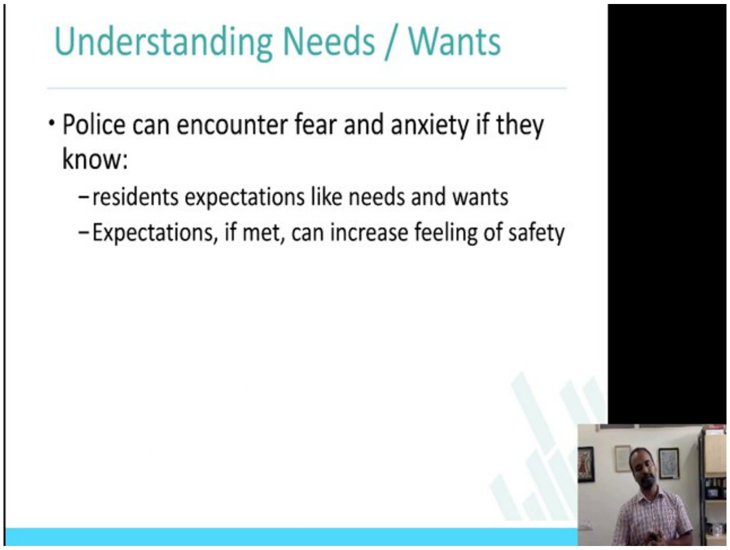
- Citizen hold police accountable

why
these illegal practices are not being stopped .
are you harassing us on roads ?
isnt the traffic police doing anything to stop the menace of Eunuchs es
everyone is overlooking the clear culprit .
did he do it ?
he want to verify my name he hang up the call . when i tried calling ba
because we blindly trust those ppl without meeting face to face , pleas
can t we ?
should he give an FIR for my negligence .
ru collecting my data base everyday , so show the same enthusiasm t
now ?
am I being the spoil sport ?
don't you stop
Tobacco ?
it ?
you the Police or the Excise Department Allow Bars to be Open till

Word Tree visualizations of posts in which residents questioned police using the word *why*.

So, if you look at the tree again earlier the example was worried now if you look at the other concept ofwhy. So, whythese illegal practices are not being stopped?Why do not you stop tobacco?Why this, whythat, right. So,this could also be a good wayto look at the content and cull out the actionable information from these posts right. So, these are things that you can do this is the types of analysis that you can do in terms of what citizens are talking about, what police organizations are actually posting.

(ReferSlideTime:16:29)

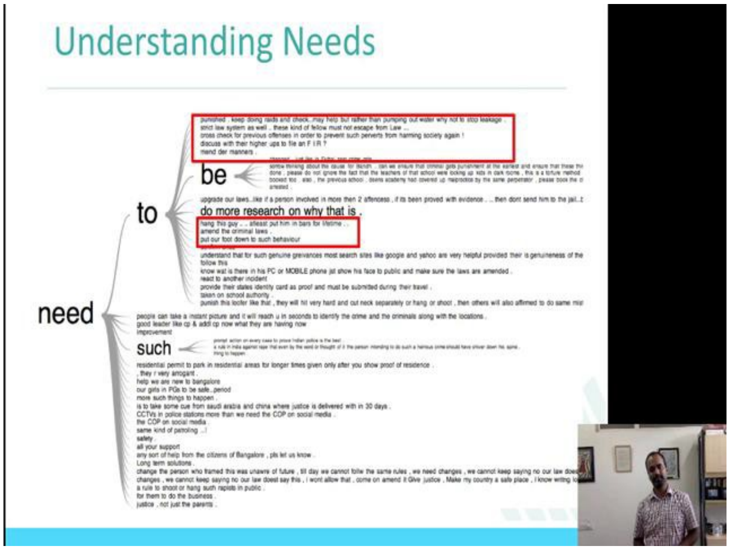


Understanding Needs / Wants

- Police can encounter fear and anxiety if they know:
  - residents expectations like needs and wants
  - Expectations, if met, can increase feeling of safety

Of course, police can also understand needs and wants right police can encounter fearand anxiety if they know resident expectations like needs and wants I want something living in this place I want some specific safety, I know that this is happening, I'm complaining, please take care of it.If only all this can be done using the content, using the information that is coming on social media, it could be very helpful. Of course, it is notthat onlythis is the onlysource for making allthese judgements (ReferTime: 17:10).

(ReferSlideTime:17:11)



Understanding Needs

need

to

punished . keep doing raids and check...may help but rather than pumping out water why not to stop leakage .
strict law system as well .. these kind of fellow must not escape from Law ...
cross check for previous offenses in order to prevent such perverts from harming society again !
discuss with their higher ups to file an F I R ?
mend der manners .

be

changed .. just like in Dubai zero crime rate
sorrow thinking about the cause for Bandh .. can we ensure that criminal gets punishment at the earliest and ensure that these thi
done , please do not ignore the fact that the teachers of that school were locking up kids in dark rooms , this is a torture method
booked too . also , the previous school , deens academy had covered up malpractice by the same perpetrator , please book the c
arrested .

upgrade our laws...like if a person involved in more then 2 affencess , if its been proved with evidence . .. then dont send him to the jail...
do more research on why that is .
hang this guy . . . atleast put him in bars for lifetime . .
amend the criminal laws .
put our foot down to such behaviour
understand that for such genuine grievances most search sites like google and yahoo are very helpful provided their is genuineness of the
follow this
know wat is there in his PC or MOBILE phone jst show his face to public and make sure the laws are amended .
react to another incident
provide their states identity card as proof and must be submitted during their travel .
taken on school authority .
punish this loofer like that , they will hit very hard and cut neck separately or hang or shoot , then others will also affirmed to do same mis

people can take a instant picture and it will reach u in seconds to identify the crime and the criminals along with the locations .
good leader like cp & addl cp now what they are having now
improvement

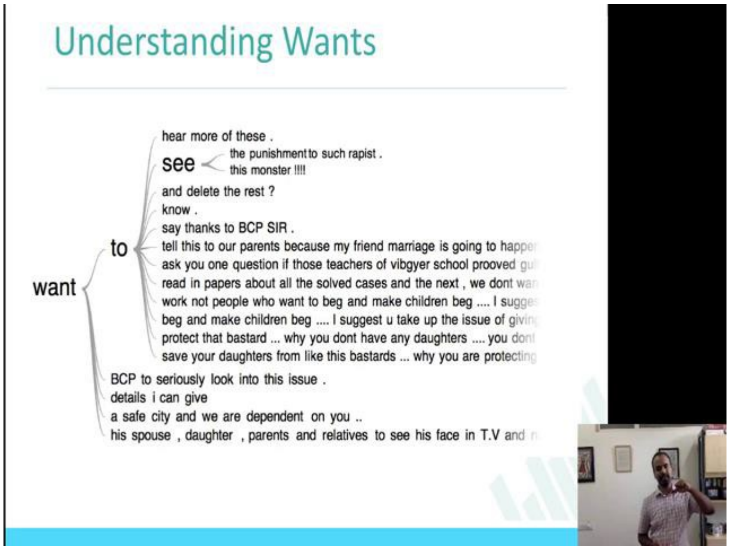
such

prompt action on every case to prove Indian police is the best .
a rule in india against rape that even by the word or thought of it the person intending to do such a heinous crime should have shiver down his spine .
thing to happen

residential permit to park in residential areas for longer times given only after you show proof of residence .
, they r very arrogant .
help we are new to bangalore
our girls in PGs to be safe...period
more such things to happen .
is to take some cue from saudi arabia and china where justice is delivered with in 30 days .
CCTVs in police stations more than we need the COP on social media .
the COP on social media .
same kind of patroling ...!
safety .
all your support
any sort of help from the citizens of Bangalore , pls let us know .
Long term solutions .
change the person who framed this was unaware of future , till day we cannot follw the same rules , we need changes , we cannot keep saying no our law does
changes , we cannot keep saying no our law doest say this , i wont allow that , come on amend it Give justice , Make my country a safe place , I know writing lo
a rule to shoot or hang such rapists in public .
for them to do the business .
justice , not just the parents .

Looking at few more examples in terms of need to be punished blah, blah, blah, need to be so and so, need to hang this guy, need to do more research on why that is going on, need such information for doing this. So, this kind of tree(Refer Time: 17:28)information can be actually very helpful, I think I have emphasised enough about thistree, I’llgo through.

(ReferSlideTime:17:31)



Understanding Wants

want

to

hear more of these .

see

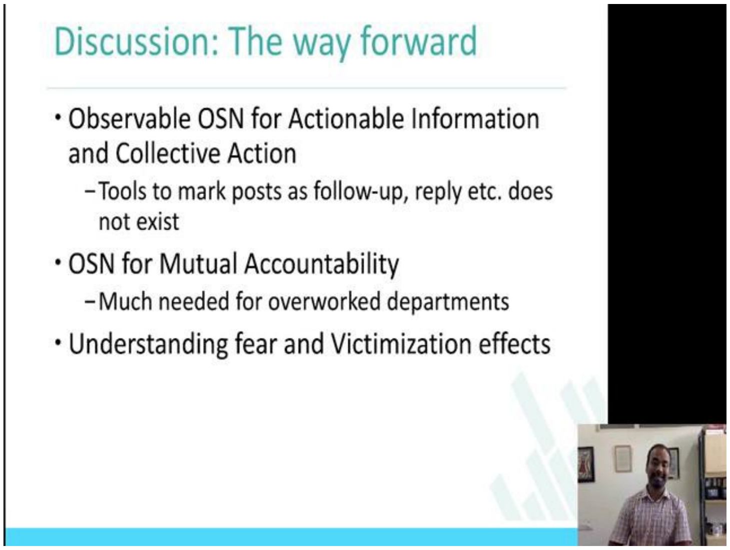
the punishment to such rapist .
this monster !!!!

and delete the rest ?
know .
say thanks to BCP SIR .
tell this to our parents because my friend marriage is going to happe
ask you one question if those teachers of vibgyer school prooved gu
read in papers about all the solved cases and the next , we dont wa
work not people who want to beg and make children beg .... I sugge
beg and make children beg .... I suggest u take up the issue of givin
protect that bastard ... why you dont have any daughters .... you don
save your daughters from like this bastards ... why you are protecting

BCP to seriously look into this issue .
details i can give
a safe city and we are dependent on you ..
his spouse , daughter , parents and relatives to see his face in T.V and r

So, this is about needs which is what this needed by the citizens, and what did theywant also. Want to hear more of these, want to see the punishment of xyz, want to and deletetherest (ReferTime:17:46) want tosaythankstoBCPSIRright.Thiskindofanalysis in terms of wants and needs which is also connected to the actionable information that we talked about is very helpful.

(ReferSlideTime:17:59)



Discussion: The way forward

- Observable OSN for Actionable Information and Collective Action
  - Tools to mark posts as follow-up, reply etc. does not exist
- OSN for Mutual Accountability
  - Much needed for overworked departments
- Understanding fear and Victimization effects

So, now just keeping these things in mind the data that we have collected from the Facebook Bangalore Citypolice what arethe things that we canthink about. Just aquick summary of what we looked at also, right, (Refer Time: 18:10) in terms of the data one could actually look at collecting allthese information, and helping understand actionable information.

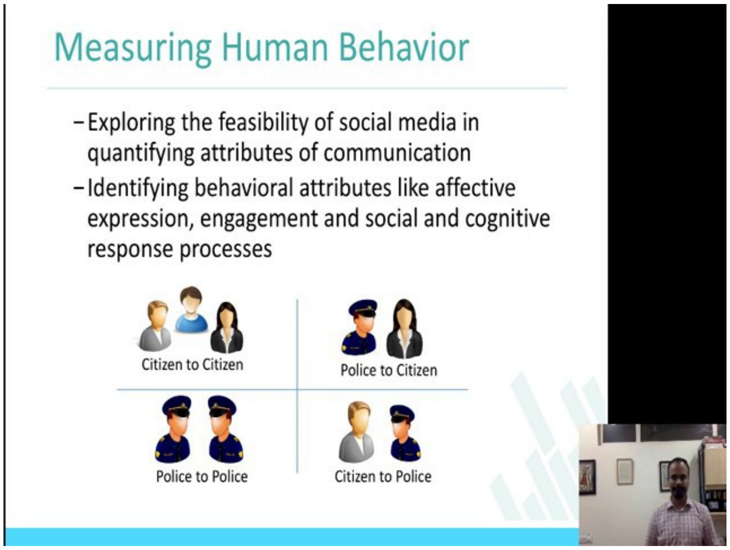
Actionable information, in the sense that I showed you it was just a tree, but how you actually take this and give it to police organizations to look at. It could be that the same tree could be shown, but I think highlighting some post saying that here is the post that you should lookat more carefully.And probablywhenproposing what kind ofpost to be givenand for a specific post here is a template for the replythat you should produce and thingslikethat.Increasingtheproductivityofthepoliceorganizationslookingatthis post canbe very, veryuseful. Ofcourse, wesawthat bothcitizens andpoliceareactually accountable because they are actually interacting on this public forum.

And of course, that is also understanding of fear; understanding of wants, understanding of needs from the citizens for police also. With that I will stop with this part of thelecture which is so to in this week, we looked at how initially we just started off with privacy, closing up the topic on privacy, then we look at different police organizations Facebook handles why they should do, what kind of post show up on these Facebook pagesandTwitterhandles,what kindofhandlesexist.Andthenwe lookedat specifically analyzing the post for identifying actionable information. With this, I will stop this lecture.

Policing and Online Social Media Part-III

Welcome back. Letuscontinueonthetopicofhow manySocialMediaand thePolicing. Now,I am going to look at different set of questions within the context of using content from social media from the Police Organizations itself.

(ReferSlideTime:00:25)



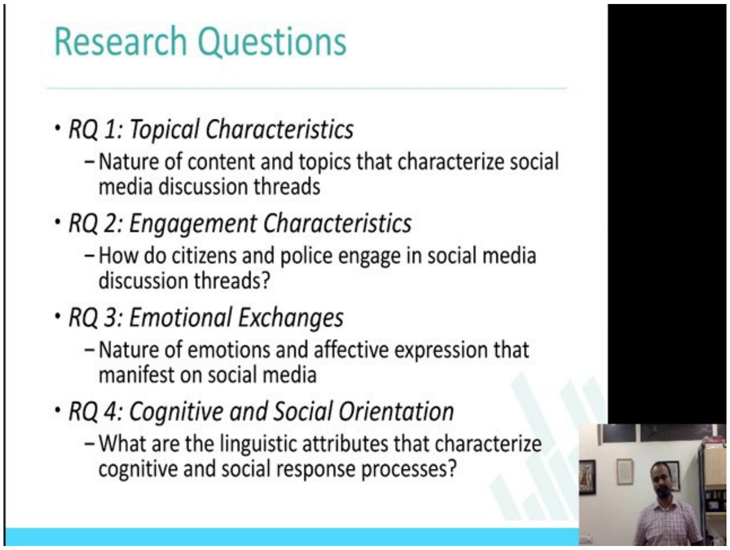
Measuring Human Behavior

- Exploring the feasibility of social media in quantifying attributes of communication
- Identifying behavioral attributes like affective expression, engagement and social and cognitive response processes

So, the question here is can we explore the feasibility of social media in quantifying attributes of communication, which is how the communication happens between Citizen to Citizen, Police toCitizen, Police to Police and Citizens to Police, in the frame workof Facebook, in the frame work of a Facebook page of Police Organization. So that is the question which is, can we actually find out what kind of attributes, how the communication happens within these four sets of interactions.

Then, identifying behavioral attributes like affective expression, engagement social and cognitiveresponseprocess;insimpletermitisbasicallylookingatfindingouthow positive or negative the interactions are, how frequentlythe engagement happens, how is the response between the parties involved in the interaction.

(ReferSlideTime:01:20)



Research Questions

- *RQ 1: Topical Characteristics*
  - Nature of content and topics that characterize social media discussion threads
- *RQ 2: Engagement Characteristics*
  - How do citizens and police engage in social media discussion threads?
- *RQ 3: Emotional Exchanges*
  - Nature of emotions and affective expression that manifest on social media
- *RQ 4: Cognitive and Social Orientation*
  - What are the linguistic attributes that characterize cognitive and social response processes?

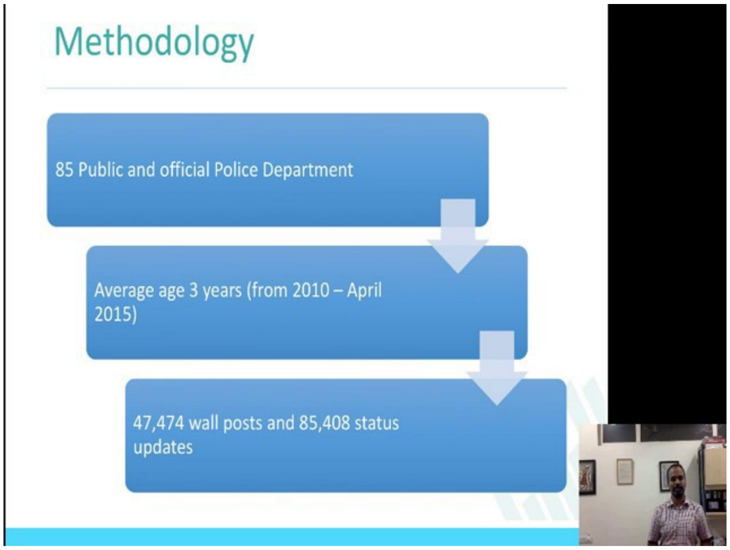
Since, specific questions that were look at are; Topical Characteristics, Engagement Characteristics, Emotional Exchanges, and Cognitive and Social Orientation. Let me walk you through little bit about what these are, and then show you some data and try to finds some results with the data that was collected, so nature of content and topics that characterize social media discussion threads. That is the topical characteristics which is what kind of content and topics people talking about.

And how the citizens and police engaged in social media discussions, how the engagement happens, what level of the engagement is it, how much engagement is there in the interaction. The third one the emotional exchanges, is a nature of emotions and affective expression that manifest in social media, which is more like how much emotions positive, negative, some concepts that are actually talk about later is violence, arousal and things like that. Cognitive and social orientation, what are the linguistic attributes that characterize cognitive and social response processes in the context of the policing Facebook pages.

These arethe fourthedifferentaspects that wewilllook atin terms ofanalyzing thedata and seeing what we can draw from the content that is getting generated. Again please keep in mind the way that I am driving this whole course, whole content is to actually take a problem try to ask some questions, look at some data, make some, do some analysis, make some inferences, and come back to the question that we have asked and say what did we learn from those questions.

If we see from the starting point that is how we have been doing credibility and trust, Boston Marathon, Hurricane Sandy kind of topics and then on privacy to doing some analysis of the face recognition and even in policing earlier part of this week we lookedat the content how the police and citizen, only Bangalore City Police interactions were going on.

(ReferSlideTime:03:22)



Methodology

85 Public and official Police Department

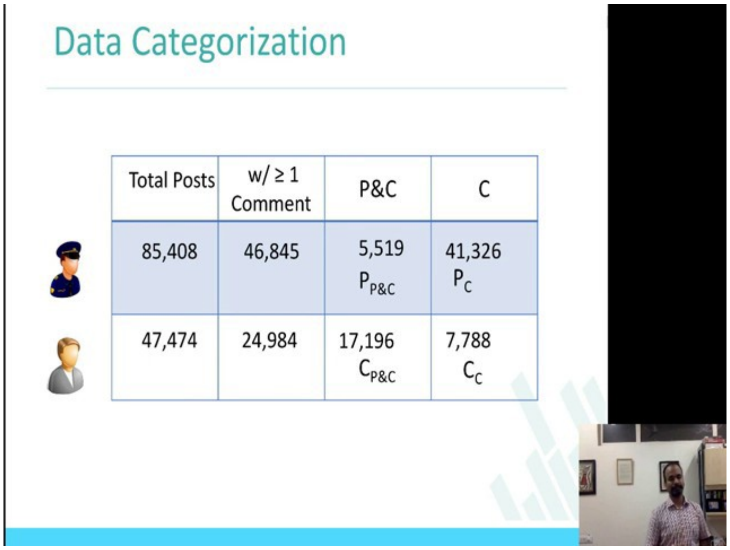
Average age 3 years (from 2010 – April 2015)

47,474 wall posts and 85,408 status updates

If you remember in the second part of this week I actually showed you a page where we were collecting, where they were pages of different Police Organization that was listed. It's the same list that wetook which is 85 Police Organizations list, ofthere are 85public and official police departments of course some of them are verified some of them are not verified, so we took this data and we took the data for the pages that were at least about averageageof3yearsbetween2010andApril2015,wherethedatacollectedandthere were47,474wallpostsandstatusupdates.

In the wall post itself, wall post or the status updates I actually do status update from the Facebook page in my account. Wall posts are the once that are showing upon my wall from others who post on the page. And status updates from me, wall posts from the people interacting with me on the page.

(ReferSlideTime:04:26)



Data Categorization

	Total Posts	w/ ≥ 1 Comment	P&C	C
	85,408	46,845	5,519 $P_{P\&C}$	41,326 P_C
	47,474	24,984	17,196 $C_{P\&C}$	7,788 C_C

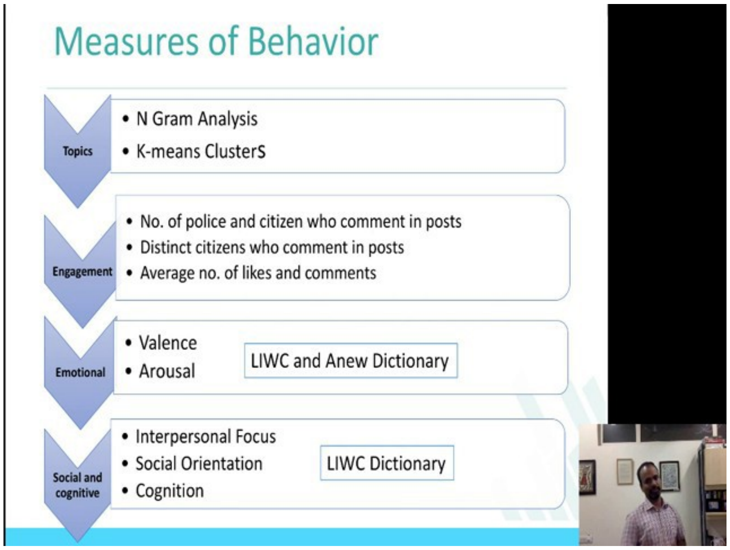
So,totalpoststhatweredonebyapolice.Theicons thatwillconsistentlygothroughthis police is for one that on the second row in this table and citizens is for the third row on this table. So, the total posts done by police were about 85,408, by the citizenswere about 47,474.This number would be same as in the last slide which is citizen walls posts people are posting on the page and the status updates which is police themselves the posting on the page. And if you looked at the post which had one or more comment is about 46,000. The posts which had one more comment from the citizens among the 47,000 is about 24,000.

So, if we see the interactions where police and citizens are interacting, the comments which is Pand C is when the police post, the way to read this table is row two which is thepolicepostandthecolumnnumber3whichisPandCmeansthatpoliceisposting thecontentandthepoliceandcitizensbothareinteractingwhichisP,PandC.AndtheC is only citizens are interacting, which shows that the police posts, citizens are actually interacting with police more.

And when you look at total posts in terms of citizens which is 47,000 which has posts, comment, number of post which are more than one comment, one more comment is about 24,000, and the wayto read this is citizens are posting and then police and citizens arecommenting onitwhichisabout17,000andonlycitizenscommentingisabout 7000.

This just gives you sense of what the data set is, which is how much of interactions are happening, when police post with citizens and police, and how much of interactions are happening in citizen post, but police and citizens and citizens are interacting.

(ReferSlideTime:06:32)



Measures of Behavior

Topics
- N Gram Analysis
- K-means ClusterS

Engagement
- No. of police and citizen who comment in posts
- Distinct citizens who comment in posts
- Average no. of likes and comments

Emotional
- Valence
- Arousal

LIWC and Anew Dictionary

Social and cognitive
- Interpersonal Focus
- Social Orientation
- Cognition

LIWC Dictionary

Also the things that, the measures were done in terms looking at what analysis could be done were the topics engagement, emotional, social cognitive evaluation. These are the four things that I showed you guys research questions can be started this lecture. Topics, N Gram Analysis was done which is how frequently the words are actually appear in 1 gram would be the single word analysis, bigram would be 2 words appearing together.

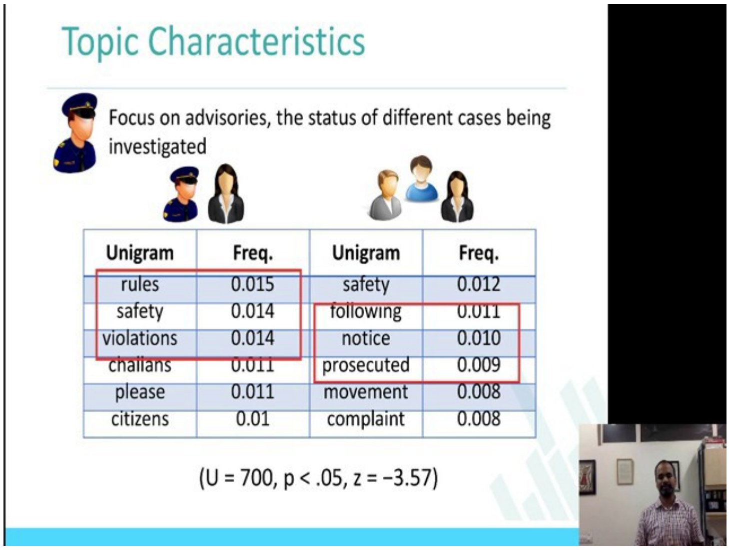
K-means Clusters, clusters are basically way in when we see users together interactingon the same post more number of people, there is a cluster there are. We can actually look in at how much of clusters happen. Clusters for example, the students who are taking this class from anywhere in the country and let us take if you look at only the students who are taking it from one particular state that would be a cluster. Students taking this course from one state could be treated as a cluster, within the network of all the students taking this course from all around the country,that is the way I havelookedat clusters.

Number of police in citizen who comment in posts: distinct citizens who comment in posts, average number of likes and comments. These are the things we look at, next of couple of slides, we'll just go through in only weeks. So, valence is another way ofsaying about the positive and negative sentiment. Arousal is about intensity of the posts itself, how strong the post is. And then answers could be done in many ways.

In thecoursealsowewillgethereintroduceinthetutorial, wewillgetyou introduced toNLTK, which is a tool kit for doing some of these text, word analysis also. Here LIWC which is again, a tool, which could be used, takes input as text paragraph let us assumeor a post from Facebook, given this input to LIWC, LIWC will produce different categories of concepts that are appearing in the content that was given to the LIWC. Which means if you get a paragraph which as a lot of positive words, lot of happywords it will produce a category as happy which will be what we have.

So, it has being very frequently used in analyzing text content, but specifically within these categories that LIWC, I think that it has about 35 or 37 categories that has if youare interested in looking at the categories in the text which is posted, LIWC could actually give you that. And again similar there are ANEW dictionary also. Interpersonal focus, social orientation and cognition, again this is analyzed through LIWC content we shall walk you through now.

(ReferSlideTime:09:27)



Topic Characteristics

Focus on advisories, the status of different cases being investigated

Unigram	Freq.	Unigram	Freq.
rules	0.015	safety	0.012
safety	0.014	following	0.011
violations	0.014	notice	0.010
challans	0.011	prosecuted	0.009
please	0.011	movement	0.008
citizens	0.01	complaint	0.008

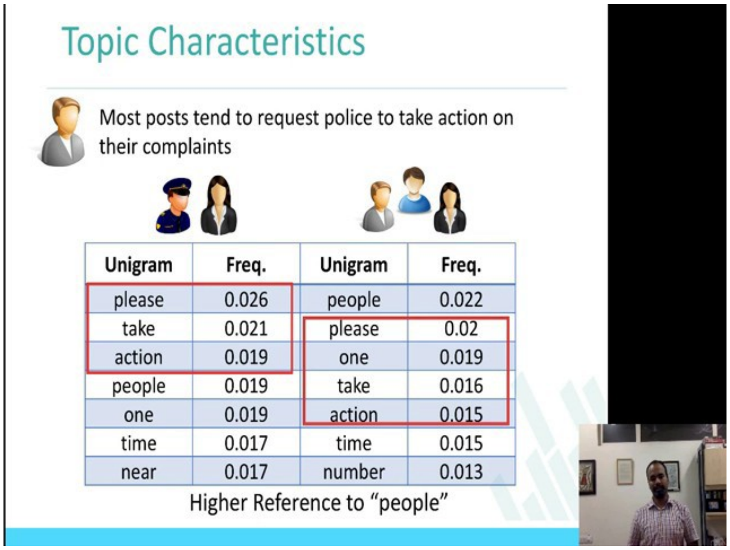
$(U = 700, p < .05, z = -3.57)$

First let us look at the topic characteristics. Four things we said, the first one is the topic that were being discussed. So focus on, so when police post again if you look at icon on the top it is police, when the police post are focusing on advisories and the status of different cases being investigated and information that the citizens will be interested in, that is what police is posting about.

If you look at the content analysis which is in police post interactions in police and citizens there is always discussions about rules, safety, violations, these are looking at only unigrams now. So you look at the concept of rules people, police posting aboutthese are the rules do not cut the signal, safety rules, wear helmet, violations and topics like that.

In terms of only citizens talking about it is actually following the content, notice, prosecuted, the interactions that citizens have, is mostly focused on those topics. Thisjustgivesyouasenseandthehighlightedisbasically totellyou howthedifferencesare, where in terms of the topics. Following, notice and prosecution versus rules, safety and violations; this gives you a sense of what kind of topics are discussed when police post and the interactions happen between police and citizens and between only citizens.

(ReferSlideTime:10:50)



Topic Characteristics

Most posts tend to request police to take action on their complaints

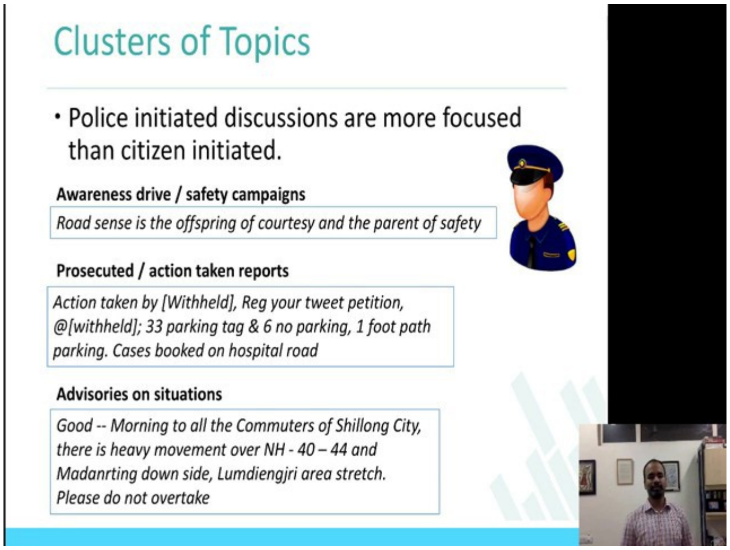
Unigram	Freq.	Unigram	Freq.
please	0.026	people	0.022
take	0.021	please	0.02
action	0.019	one	0.019
people	0.019	take	0.016
one	0.019	action	0.015
time	0.017	time	0.015
near	0.017	number	0.013

Higher Reference to “people”

Nowifyouflipthequestionandsee,whencitizensdothepostwhathappens.Mostposts request the police to a take action, which is why probably citizens are using these pages, whichiscitizenspostonthepageforaskingsomeaction.Ifyoulookatonlytheunigram analysis,weseethatpleasetakeaction, veryevidently,pleasetaketheactionseemstobe the most frequent thing which will come when citizens post and the interactions between citizens and police.

In the citizens again if you take at least one, please, one, take, action, people consist unigrams that are done within the citizens. It is basically showing you that the interactions happens between police and citizens, when citizens posts for action, police and citizens are talking about, take actions on these and citizens among themselves are also talking about take action and please help and things like that. These two topic analysis just helps us to understand how the interactions or what type of topics are being discussed when citizens and police post on content, post on these Facebook pages.

(ReferSlideTime:12:11)



Clusters of Topics

- Police initiated discussions are more focused than citizen initiated.

Awareness drive / safety campaigns

Road sense is the offspring of courtesy and the parent of safety

Prosecuted / action taken reports

Action taken by [Withheld], Reg your tweet petition, @[withheld]; 33 parking tag & 6 no parking, 1 foot path parking. Cases booked on hospital road

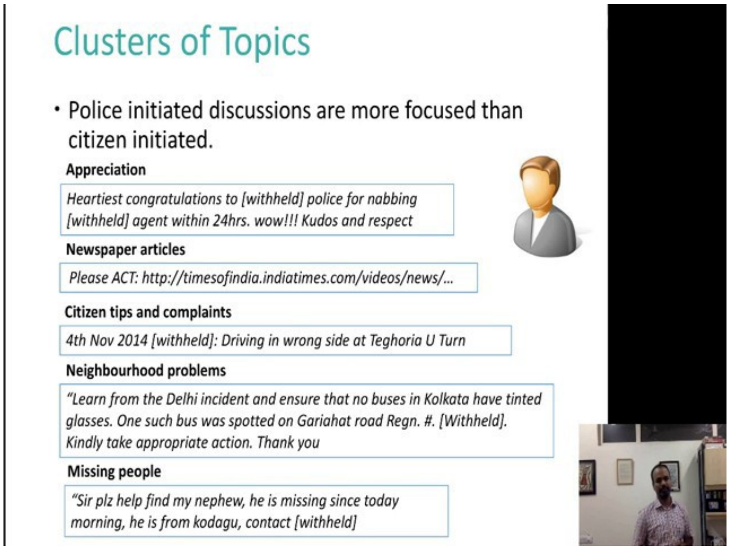
Advisories on situations

Good -- Morning to all the Commuters of Shillong City, there is heavy movement over NH - 40 – 44 and Madanrting down side, Lumdiengjri area stretch. Please do not overtake

Now,when you look at the Clusters of Topic, here when police initiated discussions are more focused than citizen initiated, let us see how; and why we are actually making this conclusion which is, when police starts the discussion it is more specific about topics, awareness drive, safety campaigns which is specifically talking about. If you look at the categories that are going to be shown up for the police created discussions, it's actually very small where as if you look at citizens it is actually quite disperse.

Here in terms of police, awareness drive and safety campaigns, road senses, the offering of courtesy, and the parent of safety, things the posts that police do. Prosecuted and action report, action taken blah blah blah regarding your tweet petition and the details about it.Advisories on situations, good morning to all commuters of Shillong City,there is heavy movement over NH-40 to 44 and blah blah blah. So, the topics are pretty small.

(ReferSlideTime:13:24)



Clusters of Topics

- Police initiated discussions are more focused than citizen initiated.

Appreciation

Heartiest congratulations to [withheld] police for nabbing [withheld] agent within 24hrs. wow!!! Kudos and respect

Newspaper articles

Please ACT: http://timesofindia.indiatimes.com/videos/news/...

Citizen tips and complaints

4th Nov 2014 [withheld]: Driving in wrong side at Teghoria U Turn

Neighbourhood problems

"Learn from the Delhi incident and ensure that no buses in Kolkata have tinted glasses. One such bus was spotted on Gariahat road Regn. #. [Withheld]. Kindly take appropriate action. Thank you

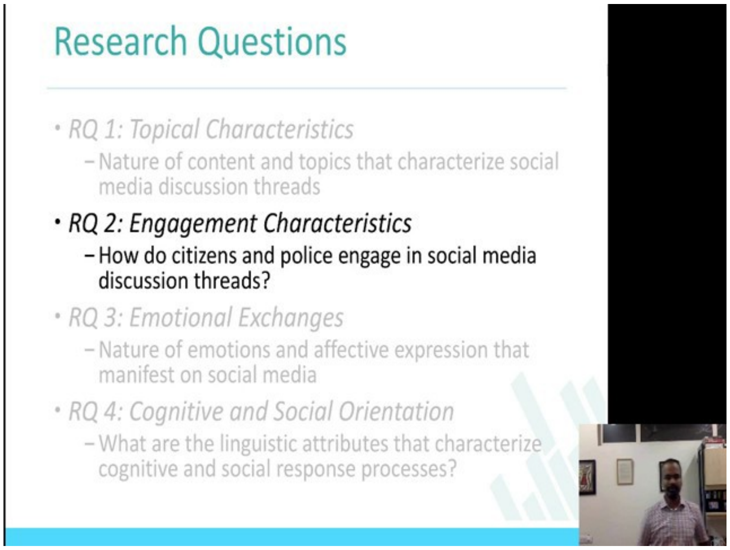
Missing people

"Sir plz help find my nephew, he is missing since today morning, he is from kodagu, contact [withheld]

If you look at the citizens one, the topics are quiet diverse. Understandably again, it is very intuitive than the topics that are coming from citizens are much wider than the police itself. But the police initiated discussions are more focused than citizen initiated.

In the conclusion here is an appreciation, newspaper articles, citizen tips and complaints, neighborhood problems, missing people and this list goes up. We are appreciating the police, sharing some news articles that they would like to share it on these pages, citizen tips, complaints, driving in wrong side at blah blah blah, neighborhood problems,missing people. These are the buckets of topics that we saw in the post that were coming from citizens. So, you can clearly see the topics have been very diverse and the police discussions are much more focused.

(ReferSlideTime:14:24)



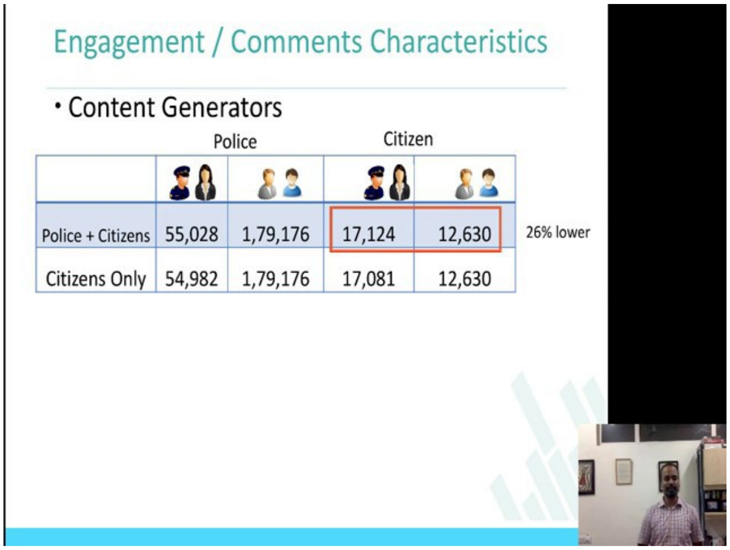
Research Questions

- *RQ 1: Topical Characteristics*
  - Nature of content and topics that characterize social media discussion threads
- *RQ 2: Engagement Characteristics*
  - How do citizens and police engage in social media discussion threads?
- *RQ 3: Emotional Exchanges*
  - Nature of emotions and affective expression that manifest on social media
- *RQ 4: Cognitive and Social Orientation*
  - What are the linguistic attributes that characterize cognitive and social response processes?

Now let us move on to the next question. So the first problem, first saying to look forward is the topics that are being discussed on these Facebook pages between citizens and between police. This is very different from the 5.1 and 5.2 topics that our week lecture that we saw, which were we were interested in more like within the Bangalore City Police can we actually get some actionable information.

So with the questions, the intention for looking at these kinds of data is very different than these two, one is looking for content, actionable information here we're lookingwhat the interaction says, can be actually learn something specific. Engagement characteristics;howdothecitizensandpoliceengageinsocialmediadiscussionthreads?

(ReferSlideTime:15:15).



Engagement / Comments Characteristics

- Content Generators

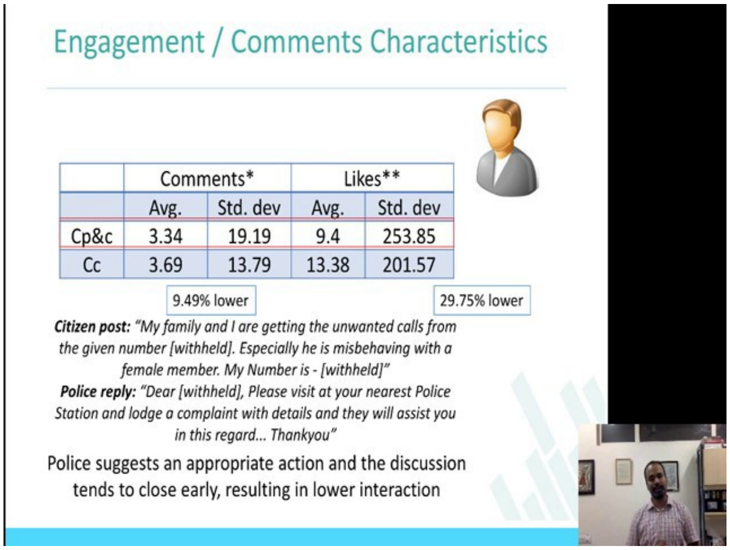
	Police		Citizen		
	👮👩	👤👤	👮👩	👤👤	
Police + Citizens	55,028	1,79,176	17,124	12,630	26% lower
Citizens Only	54,982	1,79,176	17,081	12,630	

Now let us look at the engagement or comments characteristics which is, when police do the post how the comments are discussed, when citizens do a post how comments are being spread on Facebook. Ifyou lookat the firstset of columns which is the second and the third column, this is basically showing that when police does the post what happensto the comments if police and citizens are involved in the comments. So, it shows it's about 55,000 within the police and citizens group.

As you look at citizens, if citizens' post and then the comments are from police and citizens it is seems to be much much lower. It is about 26 percent lower in terms of the commentsontheFacebookpages.Ifthepostarecomingfromcitizensandthecomments are coming from police and citizens.

This probably is because when the police does the post there is much more interactions because the information that they are giving is probably more useful, probably more engaging in terms of actually the interactions on this platform. When police does and only citizens, citizens only interact it's higher and when citizens does only citizens interactions of the comments are also lower than the police side.

(ReferSlideTime:16:46)



Engagement / Comments Characteristics

	Comments*		Likes**	
	Avg.	Std. dev	Avg.	Std. dev
Cp&c	3.34	19.19	9.4	253.85
Cc	3.69	13.79	13.38	201.57

9.49% lower

29.75% lower

Citizen post: *"My family and I are getting the unwanted calls from the given number [withheld]. Especially he is misbehaving with a female member. My Number is - [withheld]"*

Police reply: *"Dear [withheld], Please visit at your nearest Police Station and lodge a complaint with details and they will assist you in this regard... Thankyou"*

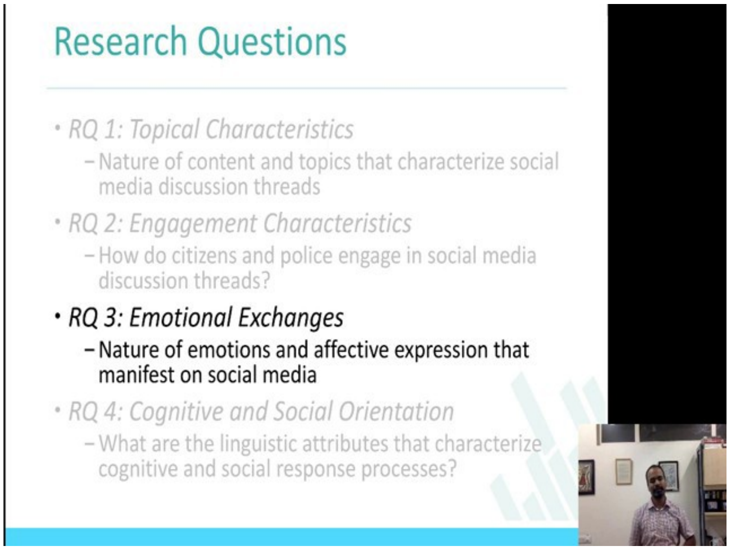
Police suggests an appropriate action and the discussion tends to close early, resulting in lower interaction

This slide is just a inside view of the data of the last slide. Last slide was this. This istotal number where as this is average number. This is average number of comments for the posts that we saw in the last slide, which is if the police and citizens are interacting there the average 3.3 comments are there for post, where as likes is about 9 likes, only citizens is about3.69 and about13.38 is on the average. So, the total gives you apicture, whereas when you look at an average it is actually the picture is slightly different which is C PC is actually 9.49 percent lower than the Cc which is the comments by police and the citizens versus comments by only citizens.

In the same way in the likes also the Cc which is comments by citizens only if the post comes from police is actually higher compared to the comments by police and citizens. This is per post average number of likes and average number of comments. In example which we kept there, a Citizen post: my family and I are getting the unwanted calls from blah blah blah blah. Police replied: dear please visit at your nearest police stationand give the details. Police suggest an appropriate action and the discussion tends to close early when the police are starting to interact. That is the reason why C P C is actually lower is because when policestarts looking at the post, police interacts they can actually easily or quickly close the loop of the discussion on the post.

Hope that make sense which is, the second question that we saw is engagement characteristics how police and citizens are engaging. So the kind of conclusions you can see is average number of comments and average number of likes when police and citizensarethereisactuallylower thanonlyacitizensbeingthere.Asimplereasoncould be police can actually quickly close in the loop of the discussion.

(ReferSlideTime:19:07)

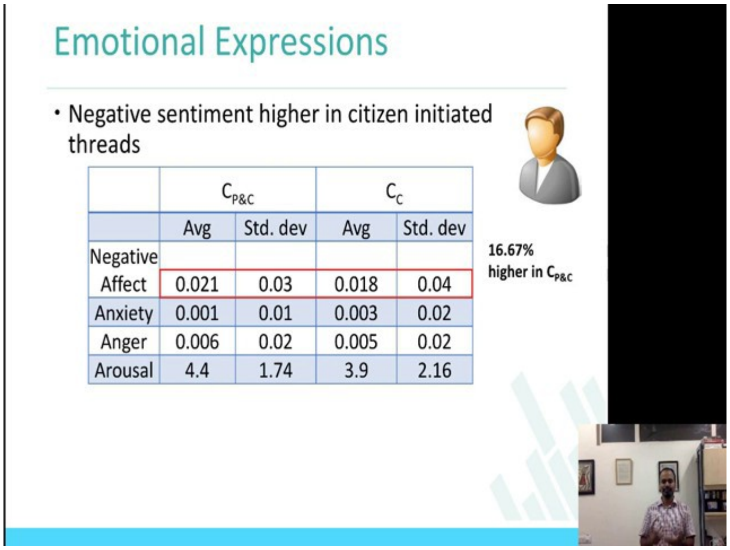


Research Questions

- *RQ 1: Topical Characteristics*
  - Nature of content and topics that characterize social media discussion threads
- *RQ 2: Engagement Characteristics*
  - How do citizens and police engage in social media discussion threads?
- *RQ 3: Emotional Exchanges*
  - Nature of emotions and affective expression that manifest on social media
- *RQ 4: Cognitive and Social Orientation*
  - What are the linguistic attributes that characterize cognitive and social response processes?

Thenextonewelookatisemotionalexchange,whichisnatureofemotionsandaffective expression that are being discussed on social media.

(ReferSlideTime:19:18)



Emotional Expressions

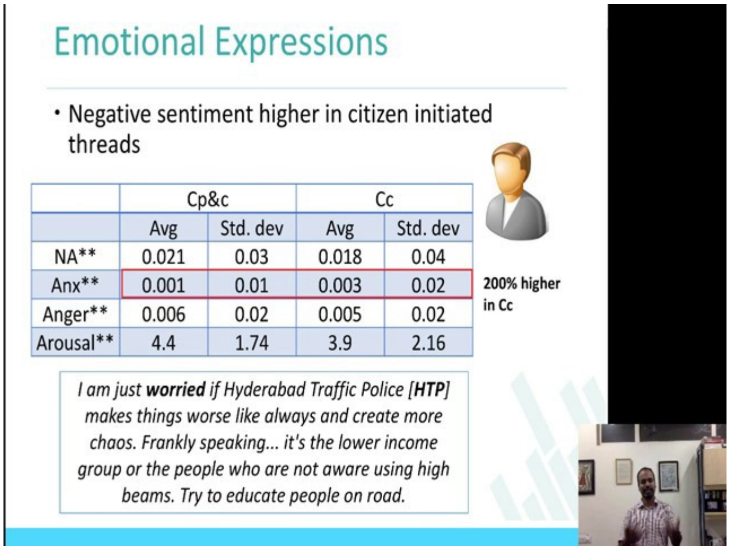
- Negative sentiment higher in citizen initiated threads

	$C_{P\&C}$		C_C	
	Avg	Std. dev	Avg	Std. dev
Negative Affect	0.021	0.03	0.018	0.04
Anxiety	0.001	0.01	0.003	0.02
Anger	0.006	0.02	0.005	0.02
Arousal	4.4	1.74	3.9	2.16

16.67% higher in $C_{P\&C}$

As a one conclusion that we can look at is when citizens initiated threads there is always highernegativesentiment.HereisatablewhichactuallyconnectstothetopicsthatIsaid earlier which is negative effect, anxiety, anger and arousal. In the same one you can see thatnegativeaffectishigherwhenpoliceisactuallyinvolvedintheinteraction,whichis 0.021 versus 0.018. The reason for this would be that citizens are expressing their views about crimes, expressing the views about things that are going on and around them, even in the some of the graph seen before like neighborhood problems and they want to actually express the views so therefore the negative affect is higher.

(ReferSlideTime:20:04)



Emotional Expressions

- Negative sentiment higher in citizen initiated threads

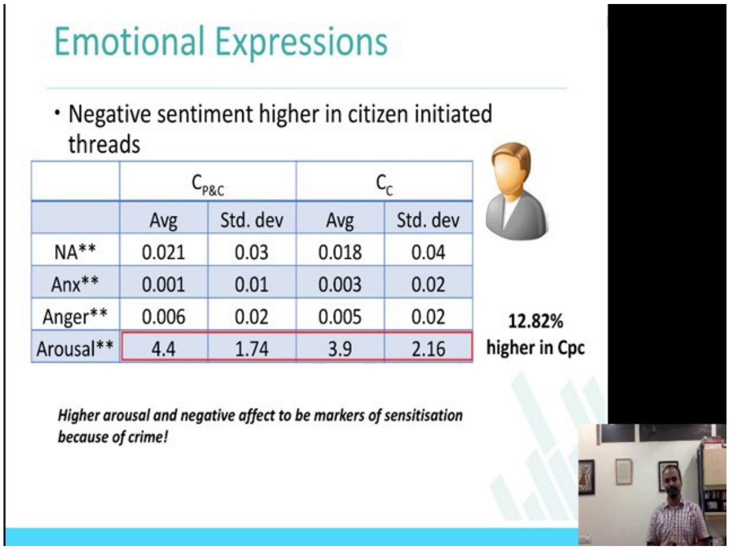
	Cp&c		Cc	
	Avg	Std. dev	Avg	Std. dev
NA**	0.021	0.03	0.018	0.04
Anx**	0.001	0.01	0.003	0.02
Anger**	0.006	0.02	0.005	0.02
Arousal**	4.4	1.74	3.9	2.16

200% higher in Cc

*I am just **worried** if Hyderabad Traffic Police [**HTP**] makes things worse like always and create more chaos. Frankly speaking... it's the lower income group or the people who are not aware using high beams. Try to educate people on road.*

If you look at the next one where the anxiety is actually lower when police starts interactinginthesepostswhichareinitiatedbythecitizens,youcanseethevaluetobe 0.001 versus Cc to be 0.003.And the reason could be is that police starting to interact in these post, police could actually complete the loop, as I said few slides before also completing the discussion and making the discussion closure which lets the anxiety ofthe citizens to go low.

(ReferSlideTime:20:37)



Emotional Expressions

- Negative sentiment higher in citizen initiated threads

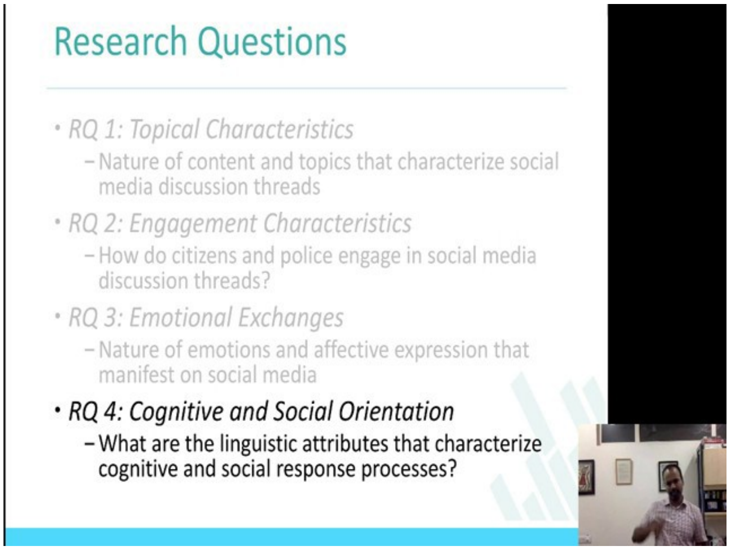
	$C_{P\&C}$		C_C	
	Avg	Std. dev	Avg	Std. dev
NA**	0.021	0.03	0.018	0.04
Anx**	0.001	0.01	0.003	0.02
Anger**	0.006	0.02	0.005	0.02
Arousal**	4.4	1.74	3.9	2.16

12.82% higher in Cpc

Higher arousal and negative affect to be markers of sensitisation because of crime!

And in the third row, we actually show you can anger is also higher in terms of C P C versus C C which is 0.006 to 0.005, where the anger is higher is again could be that citizens are actually expressing the views and they really want to get things done andthey are approaching the police to actually get things done as quickly as possible. And arousal is of course higher in C PC. Again when the citizens initiated threads, the C PC is higher, the arousal because the intensive discussion is higher because you want actually get things done from police again.

(ReferSlideTime:21:14)

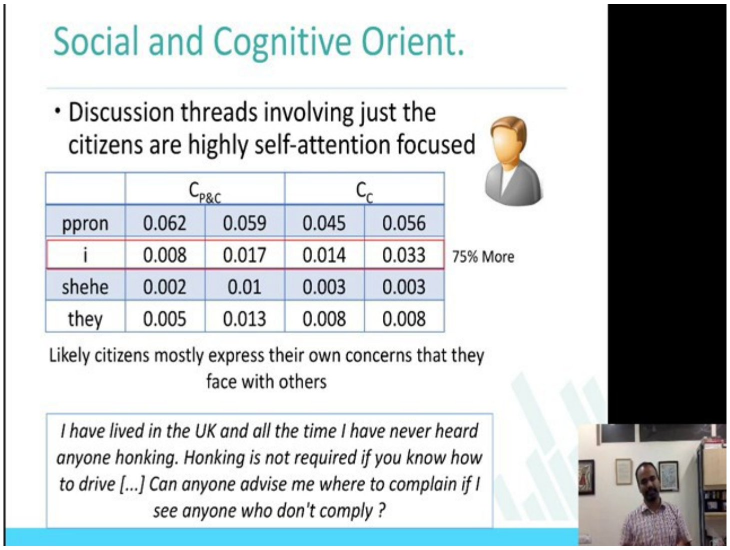


Research Questions

- *RQ 1: Topical Characteristics*
  - Nature of content and topics that characterize social media discussion threads
- *RQ 2: Engagement Characteristics*
  - How do citizens and police engage in social media discussion threads?
- *RQ 3: Emotional Exchanges*
  - Nature of emotions and affective expression that manifest on social media
- *RQ 4: Cognitive and Social Orientation*
  - What are the linguistic attributes that characterize cognitive and social response processes?

Now let us look at the last part of this deck of slides in terms of actually the questions thatwestartedwith.Westartedthefirstquestiontobetopicalcharacteristics,secondone to betheengagementcharacteristics, third oneto beemotional exchanges, and thefourth one to be cognitive and social orientation. Here the question that we are trying to post was; what are the linguistic attributes or characterize cognitive in social response process? I think this is a very broad and question that people could study in multiple ways.

(ReferSlideTime:21:49)



Social and Cognitive Orient.

- Discussion threads involving just the citizens are highly self-attention focused

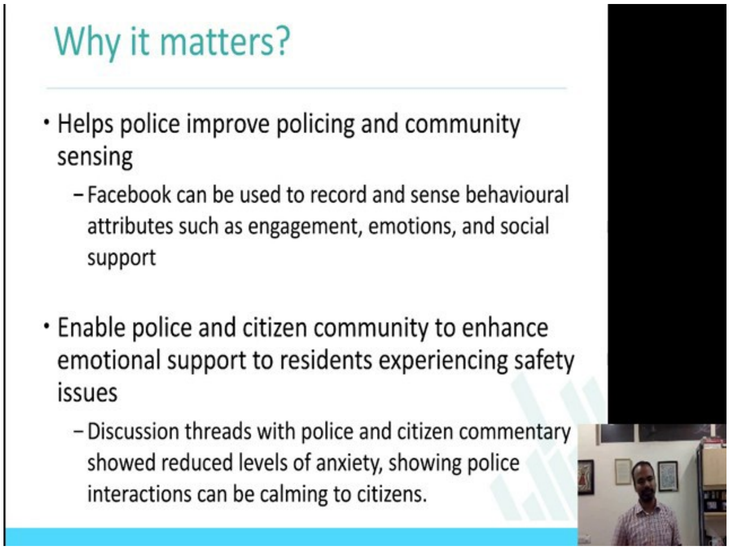
	$C_{P\&C}$		C_C		
ppron	0.062	0.059	0.045	0.056	
i	0.008	0.017	0.014	0.033	75% More
shehe	0.002	0.01	0.003	0.003	
they	0.005	0.013	0.008	0.008	

Likely citizens mostly express their own concerns that they face with others

I have lived in the UK and all the time I have never heard anyone honking. Honking is not required if you know how to drive [...] Can anyone advise me where to complain if I see anyone who don't comply ?

Theoneapproachthatwecoulditwaslookingatthepropernounsi’sandhe'sandthey's, words used in the post that were initiated by the citizens. And we see that most of the post that the citizens initiate, are actually self driven or mentions more of themselves. You can clearly see 'i' being the highest in this table. An example for the post is, 'I have lived in the UK and all the time I have never heard anyone honking blah blah blah'. So, these kinds of post are mostly presented in the Facebook data that we collect.

(ReferSlideTime:22:28)



Why it matters?

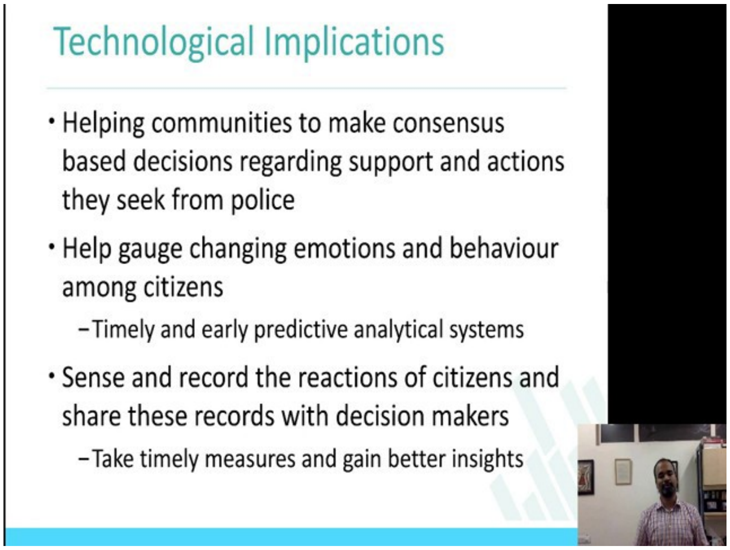
- Helps police improve policing and community sensing
  - Facebook can be used to record and sense behavioural attributes such as engagement, emotions, and social support
- Enable police and citizen community to enhance emotional support to residents experiencing safety issues
  - Discussion threads with police and citizen commentary showed reduced levels of anxiety, showing police interactions can be calming to citizens.

Now let us look at why all this matters. For the entire week for this course, this week I kind of took you around the data that are created on Facebook, that are created on social media for Police Organizations doing some analysis, seeing what kind of interactionsthat are going on between police and citizens, why does it all matter. One it definitely tells police improve policing and community sensing, which is to collect data from these kind of social networks can be used to record and sense behavioral attributes, such as engagement,emotions,socialsupport.Whichiswhatwearetryingtocapturein5.3.And it will enable police incidence community to enhance emotional support to residents experiencing safety issues also.

These are the thing that we talked about again in the analysis. Discussion thread with police and citizens where gives the level of anxiety when police talks interacting right, a couples of slides back showing you about C P C, in C P C the value for anxiety is actuallylower is just because that they,when police staffs interacting the citizens tend to be more, anxious level being it is lower or it tends to be lower because police starts actually answering some of the questions that people are asking.

For example, I say that there is some problem here by in my neighborhood and police comesandsays,Ohgood,thankyouforlettingusknowaboutitandwewillactuallydo the appropriate thing forwarding this to the right police officer all that. So, there is some level of satisfaction to the citizens while the CP responses.

(ReferSlideTime:24:02)



Technological Implications

- Helping communities to make consensus based decisions regarding support and actions they seek from police
- Help gauge changing emotions and behaviour among citizens
  - Timely and early predictive analytical systems
- Sense and record the reactions of citizens and share these records with decision makers
  - Take timely measures and gain better insights

Interestingly we can actually develop some technologies also, helping communities to make consensus based decisions regarding support and actions they seek from police. Theycan actuallybuild technologies which will help these kinds of interactions and help decisions also to be made which can help have more safer cities, help our more interactions between police and citizens. It can help actually understand the change in emotions of people also.

For example, I manage the police page today, I see what kind of views that people have about a particular problem today and what kind of content they are generating today which is from my jurisdiction and month later what happens, a year later what happens, understanding this will actually help in terms of making some predictive analytics solutions also. Predictive analytics around this data and actually helping Police Organizations make their job better, make cities more safer.

Itcanalsohelpsenseandtherecordthereactionsofcitizensandsharetheserecordswith decisionmakers.Itcanactuallyget,meaningsomeorganizationsalsocandothis.

Publicly saying that these are the number of posts that we got this month this week and this is what we get. It can actually act as an early warning system; it can act as a predictive usage of this data, it can use for predictive analysis also. Essentially all of this can help both societal impact in terms of police and citizens interacting better. Technologies can be built to help police make their decision better; it can help citizens have safer lives.

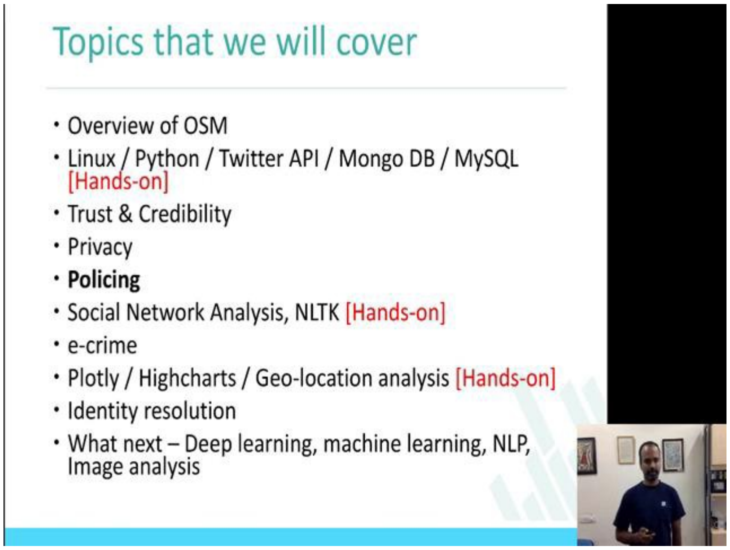
With that I will stop the content for week 5. So, essentially week 5 is totally about policing. And we looked at initially how these organizations create, what kind of pages do they have, and what content are getting generated on this pages, from therewe wenton to ask some research questions are on that. And these kind of research is not onlydone in terms of just Indianwebpages, Indian Police Organizations, this is done allacross the world.

Thankyou.

eCrime on Online Social Media Part-I

WelcomebacktothecoursePrivacyandSecurityinOnlineSocialMedia,thisisweek6.

(ReferSlideTime:00:17)

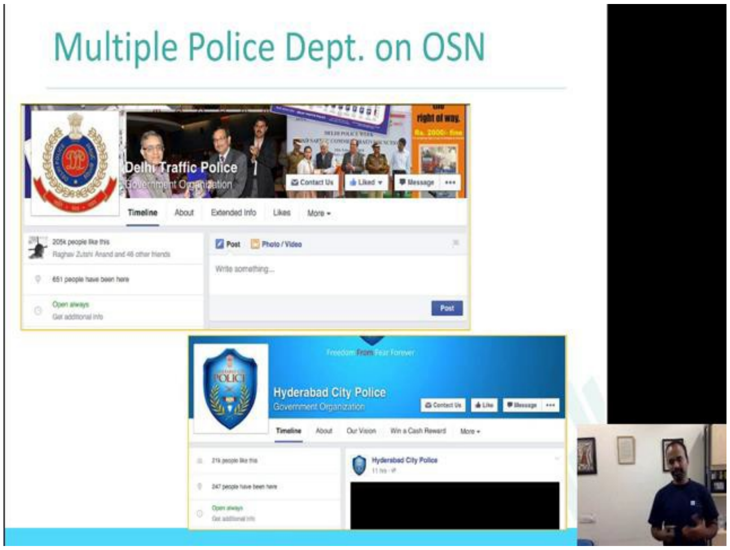


Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- Trust & Credibility
- Privacy
- **Policing**
- Social Network Analysis, NLTK [Hands-on]
- e-crime
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

So, what we have seen until now is generally, overview of online socialmedia. We have had a lot of hands on tutorials about Linux, Python, Twitter API, Mongo DB, MySQL and then I went into topics like Trust and Credibility, then we saw Privacy, last week we saw what is Policing how online social media is being used by police organizations specifically in India and what research problems, what questions that you can actually studyfromthe data that you collect fromthese socialmedia services.

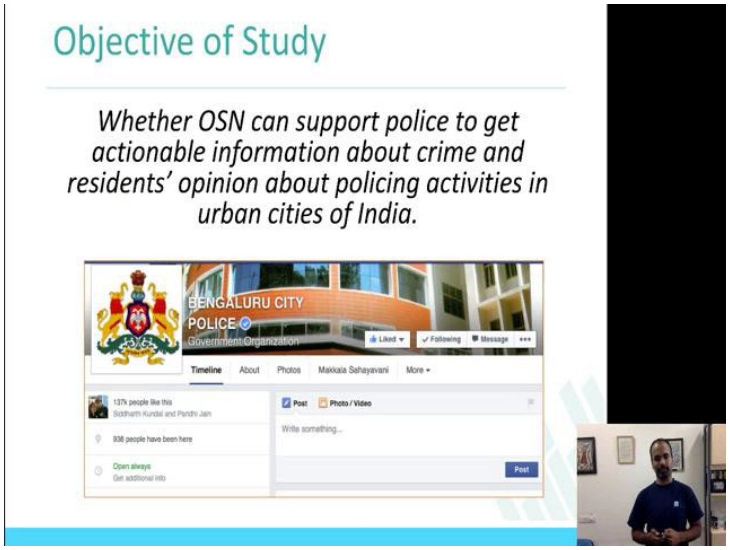
(ReferSlideTime:00:56)



Multiple Police Dept. on OSN

Let me just quickly tell you what we saw. Multiple police organizations have actually adopted using Facebook, Twitter, for sharing for interacting with the citizens and that is the topic that we saw inthe context ofpolicing.Aspecific questionthat we saw washow we canactuallyuse this data fromsocialmedia to collect actionable information.

(ReferSlideTime:01:19)

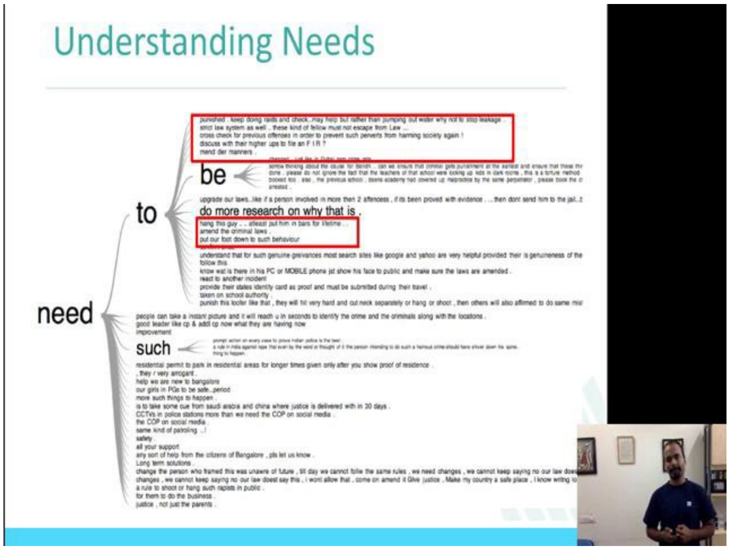


Objective of Study

Whether OSN can support police to get actionable information about crime and residents' opinion about policing activities in urban cities of India.

Isitpossiblewecancollectsomeactionableinformation?Isitpossibletousethis information for making any interesting judgments?

(ReferSlideTime:01:33)

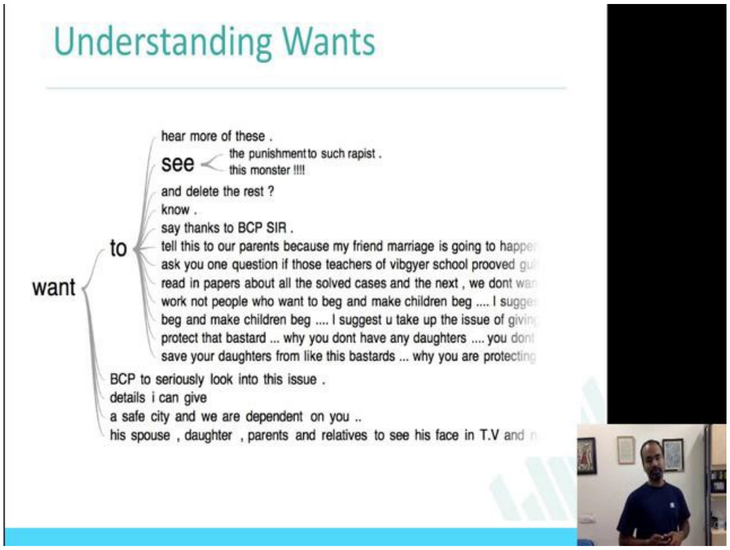


Understanding Needs

need
to
be
punished . keep doing raids and check..may help but rather than pumping out water why not to stop leakage .
strict law system as well .. these kind of fellow must not escape from Law ...
cross check for previous offenses in order to prevent such perverts from harming society again !
discuss with their higher ups to file an F I R ?
mend der manners .
upgrade our laws..like if a person involved in more then 2 affencess , if its been proved with evidence . ...then dont send him to the jail..
do more research on why that is .
hang this guy ... atleast put him in bars for lifetime ...
amend the criminal laws .
put our foot down to such behaviour
understand that for such genuine grievances most search sites like google and yahoo are very helpful provided their is genuineness of the
follow this
know wat is there in his PC or MOBILE phone jst show his face to public and make sure the laws are amended .
react to another incident
provide their states identity card as proof and must be submited during their travel .
taken on school authority .
punish this looler like that , they will hit very hard and cut neck separately or hang or shoot , then others will also affimed to do same mis
people can take a instant picture and it will reach u in seconds to identify the crime and the criminals along with the locations .
good leader like cp &amp; addl cp now what they are having now
improvement
such
prompt action on every case to prove Indian police is the best .
thing to happen .
residential permit to park in residential areas for longer times given only after you show proof of residence .
, they r very arrogant .
help we are new to bangalore
our girls in PGs to be safe...period
more such things to happen .
is to take some cue from saudi arabia and china where justice is delivered with in 30 days .
CCTVs in police stations more than we need the COP on social media .
the COP on social media .
same kind of patroling ...!
safety .
all your support
any sort of help from the citizens of Bangalore , pls let us know .
Long term solutions .
change the person who framed this was unaware of future , till day we cannot follw the same rules , we need changes , we cannot keep saying no our law does
changes , we cannot keep saying no our law doest say this , i wont allow that , come on amend it Give justice , Make my country a safe place , I know writing
a rule to shoot or hang such rapists in public .
for them to do the business .
justice , not just the parents .

In this context we also saw that how we can use the text content that is posted on these social media services to take some actionable information. For example, this tree shows how you can understand the needs of citizens who are posting on these networks. Likefor example it says, need to be punished, need to hang this guy. So, these are the needs from the citizens who are posting this content on Facebook or Twitter or other social media services.

(ReferSlideTime:02:10)

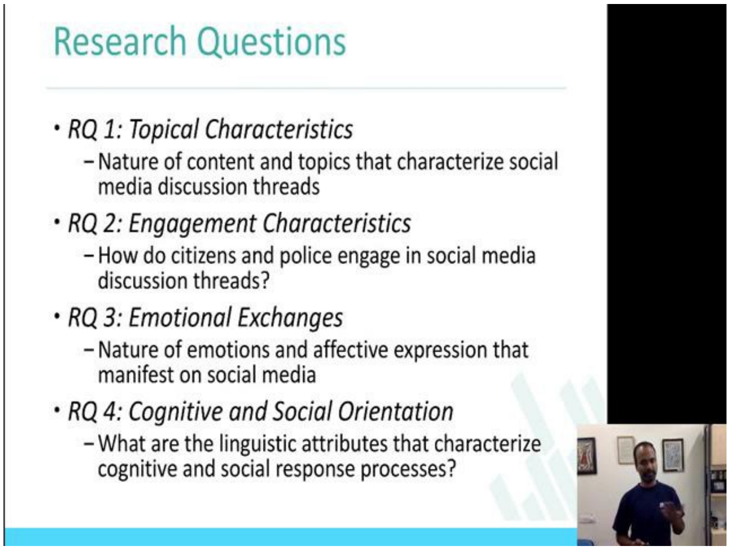


Understanding Wants

want
to
hear more of these .
see
the punishment to such rapist .
this monster !!!!
and delete the rest ?
know .
say thanks to BCP SIR .
tell this to our parents because my friend marriage is going to happe
ask you one question if those teachers of vibgyer school prooved gu
read in papers about all the solved cases and the next , we dont wa
work not people who want to beg and make children beg .... I sugge
beg and make children beg .... I suggest u take up the issue of givin
protect that bastard ... why you dont have any daughters .... you don
save your daughters from like this bastards ... why you are protecting
BCP to seriously look into this issue .
details i can give
a safe city and we are dependent on you ..
his spouse , daughter , parents and relatives to see his face in T.V and

Then you can also look at understanding wants, which is what is that citizens are interested in wanting from police, this we like want to hear more of these, here want to see the punishment of such people, want to saythanks to BCPSir. This is something we saw earlier and we are just going to quicklybrush it only.

(ReferSlideTime:02:30)

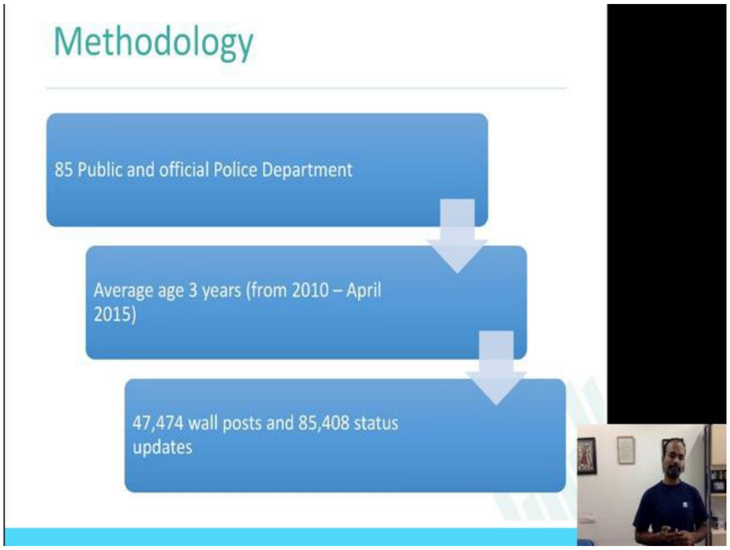


Research Questions

- *RQ 1: Topical Characteristics*
  - Nature of content and topics that characterize social media discussion threads
- *RQ 2: Engagement Characteristics*
  - How do citizens and police engage in social media discussion threads?
- *RQ 3: Emotional Exchanges*
  - Nature of emotions and affective expression that manifest on social media
- *RQ 4: Cognitive and Social Orientation*
  - What are the linguistic attributes that characterize cognitive and social response processes?

Thefour questionsthat wesaw specificallywhere,topicalcharacteristics,what topicsare being discussed, how the engagement between police and citizens are happening, what emotional exchanges are happening between citizens and police, specifically we also looked at arousal, violence and topics around that. Finally, we looked at cognitive and socialorientation, linguistic attributes, unigram, and bigram, and topics around that.

(ReferSlideTime:03:04)



Methodology

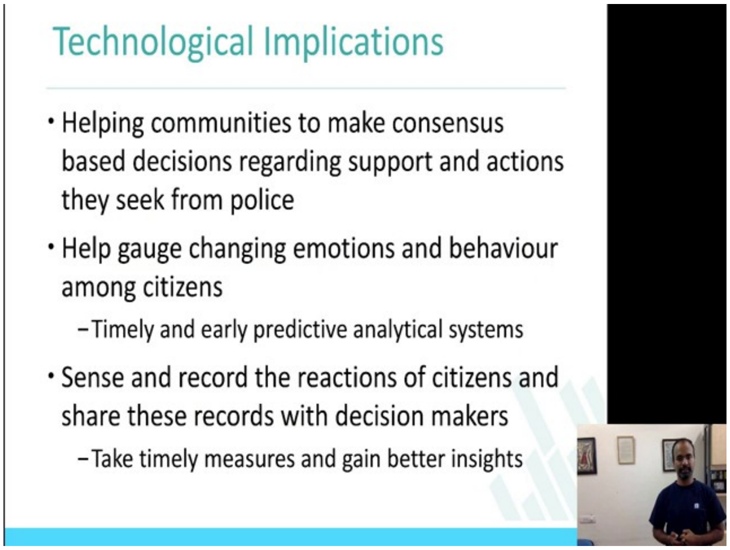
85 Public and official Police Department

Average age 3 years (from 2010 – April 2015)

47,474 wall posts and 85,408 status updates

In this data specifically what we saw was collecting data from 85 publicly and official departments between this period of a 2010 and 2015, the analysis was done on 47,474 wall posts and 85,000 status updates.

(ReferSlideTime:03:22)



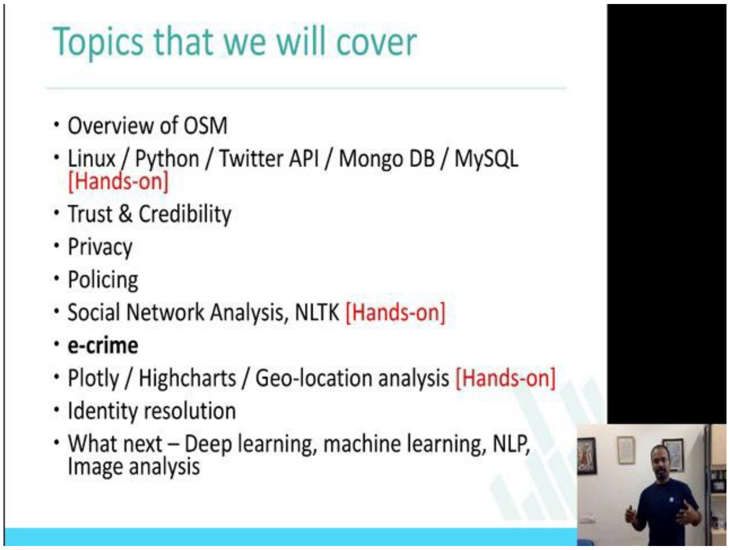
Technological Implications

- Helping communities to make consensus based decisions regarding support and actions they seek from police
- Help gauge changing emotions and behaviour among citizens
  - Timely and early predictive analytical systems
- Sense and record the reactions of citizens and share these records with decision makers
  - Take timely measures and gain better insights

And of course, the technicalimplications ofdoing allthis is helping communities to help thepoliceorganizations, buildtechnologieswhichcanbeusedbycitizenstointeractwith policebetter,buildtechnologiesthatpolicecanuseforinteractingwithcitizensbetterand making the society a safer place to live.

So, that is the broader goal of studying these concepts on social media. As I said in the last week also I would really like to see people talk about their citypolice organizations andinteractionsifany,ontheforum,I havenot seenanythingmuchuntilnow,butIthink fornowmanyofyoumayjust understandingthecontentthatisjustthecontentitself.But it is actually great to see some going to see some interesting questions, students are asking in the forum.

(ReferSlideTime:04:19)



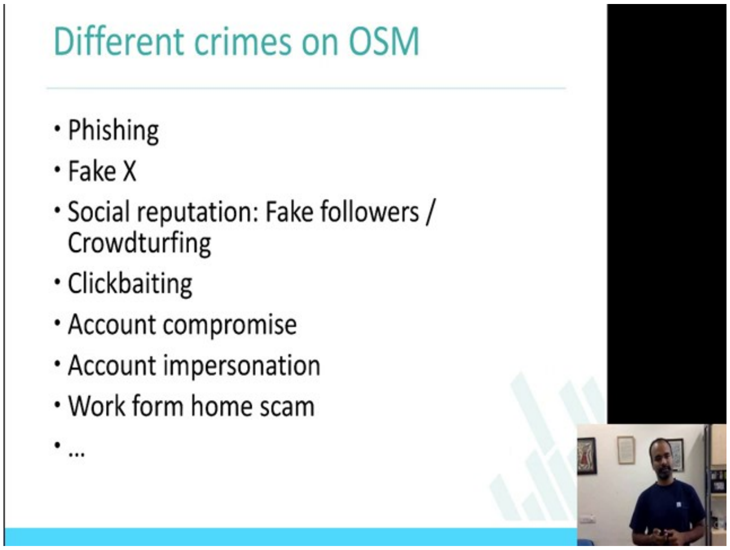
Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- Trust & Credibility
- Privacy
- Policing
- Social Network Analysis, NLTK [Hands-on]
- **e-crime**
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

So, what we willdo now is we willmove onto another topic fromhere. The topic now I want to look at is e-crime; e-crime, cyber crime anything that isaround electroniccrime, but focus it only on the social media context. Crimes happen allaround the places usingthe internet, using the web, but people will focus on these kind of crimes only that is happening on social media.

And asthe patternonthecourse,wewilldo somebasicsnowinthefirst part ofthisweek then I willget into some research questions or questions that one could answer using the data that is been collected on crimes from these social media services itself. We will do these hands on tutorials also, which is looking at social network analysis tools and then NLTK. And there are other hands on tutorials also that are we have planned over the course of next few weeks.

(ReferSlideTime:05:21)

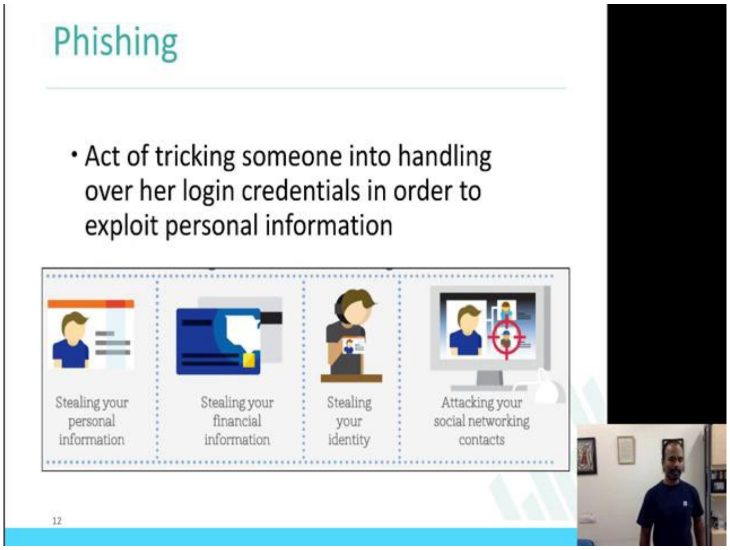


Different crimes on OSM

- Phishing
- Fake X
- Social reputation: Fake followers / Crowdturfing
- Clickbaiting
- Account compromise
- Account impersonation
- Work form home scam
- ...

So here is a list of not a comprehensive list, here is a list of crimes that I thought I will cover before getting into details of any one particular topic. We are going to look at one or two topics in detail, but before that let me just walk you through some crimes that happenononlinesocialmedia.Iamsuresomeofthiswearealreadyawarebutletmejust brush it to get your sense of what the crimes that are going on in these social media services.

(ReferSlideTime:05:51)



Phishing

- Act of tricking someone into handling over her login credentials in order to exploit personal information

Stealing your personal information
Stealing your financial information
Stealing your identity
Attacking your social networking contacts
12

The first one which is phishing, and again these are not arranged in any particular order andtheyarenot comprehensiveatall.Thephishingproblemonsocialmediaservices,the act oftrickingsomeoneto into handling orlogging detailswhichisbasicallythereisa,in traditionalwaysinemails,inemaildomainsyougetemailswhichsayspleaseclickonthislink or please click onthe links to update a password or your account is expired click on this link to activate your account.

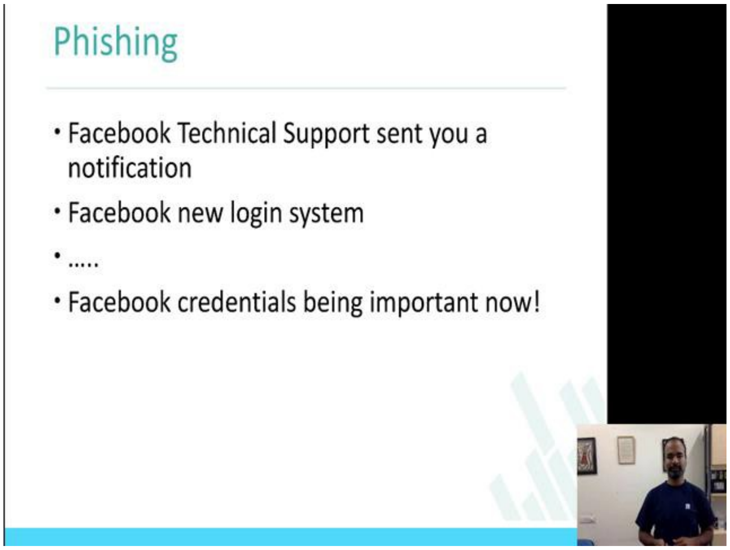
Whenyouclickonthislinkyouare takento afakewebsitewhichsometimeslookslikea legitimatewebsite, but sometimesit doesnot needto belooking likealegitimatewebsite also. And when you go there it is asking for username, password and when you give the username, password, you are basicallysharing the credentials to someone else.

And these kind of emailshave been playing around for a long time and there are many sophisticated attacks that has happened using these emails; phishing itself, just phishing, sphere phishing. Sphere phishing is a way by which you target a set of people. For example, in this course I could just target only the people who are taking this course saying as though it is email coming from PK at IIIT, saying please click on this link to know further links that I have actually posted on the web about the course.And then, of course,someof youmay beinterestedin whatIam speakingaboutthecourseandyou willclickonthelink,butitisnotactuallyalegitimateemailor alegitimatelink.Sothatis about sphere phishing, but then there are other types of phishing also, which is whaling where the specific CEO's of a company are targeted while sending out these phishing emails. There are many different types of phishing attacks that have been going on. So, that istraditional.

But now when you move on to the socialnetwork, these attacks have also calculated the socialmediaservicesalso. For example, alinkontotheTwitter timelinewilltellyouthat please click onthislinkto get somemoneyandthenwhenyouclickonthislinkorplease clickonthislinktochangetheFacebookpasswordthatyou havecreated,therewassome problemin your access with Facebook; click on this link to update the password.

As in the traditionalwayifyou click onthis link you willend up actuallygoing to a fake websiteand giving awaythecredentials,that isphishingandImeanyoucanthinkofit as a phishing as in the traditional ways in emails itself, but spreading on the social media services. There could be a linkonFacebook, there couldbealinkonTwitter;therecould be anemailto say, please click onthis image to get some more informationabout atopic and it could actuallytake you to a fake website. So that is phishing.

(ReferSlideTime:08:44)



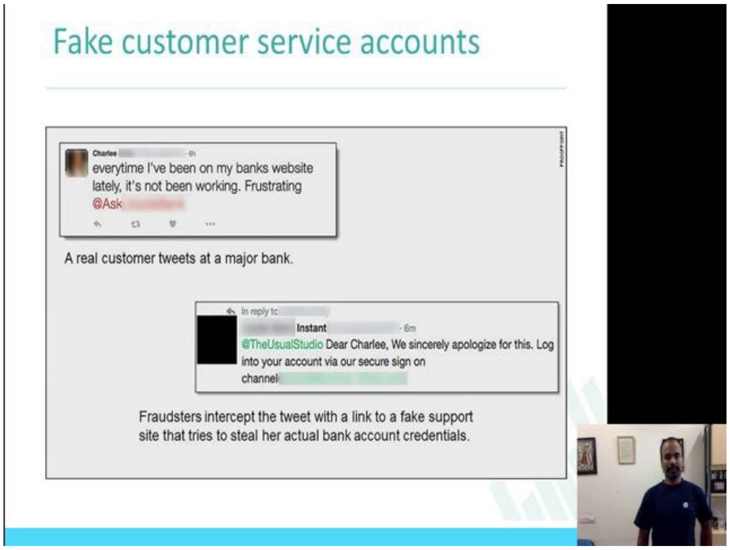
Phishing

- Facebook Technical Support sent you a notification
- Facebook new login system
- .....
- Facebook credentials being important now!

So, specifically the examples in phishing that are going on now or have been around for sometime is Facebook technicalsupport sent you a notificationsaying that, there issome probleminyouraccountpleasego verify.Facebooknewloginsystemthatisemailsgoing aroundwhichsaysthatFacebookhasinventedanewloginsystemandclickonthislinkto create your account on this new login system or merge this account to the Facebook account and things like that, these emails have been going around.

Andifyoureallylookat it,Facebookcredentialsarebecomingmoreandmoreimportant, because if I know your Facebook credentials I actually get to know your friends, I actuallyget to know your pattern ofusage, interest and topicsthat you maybe interested in spending time. These things can be used against you. So that is the reason why Facebook credentials are also becoming more and more popular, compared to the email address,compareto thefinancialaccount detailsthat one wasalso chasingbefore.Thatis phishing.

(ReferSlideTime:09:50)



Fake customer service accounts

Charlee [redacted] · 6h
everytime I've been on my banks website lately, it's not been working. Frustrating @Ask[redacted]

A real customer tweets at a major bank.

In reply to [redacted]
[redacted] Instant [redacted] · 6m
@TheUsualStudio Dear Charlee, We sincerely apologize for this. Log into your account via our secure sign on channel [redacted]

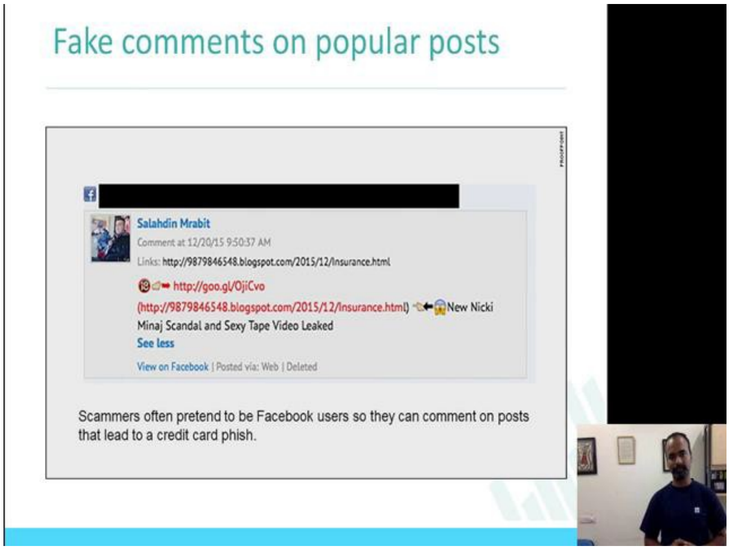
Fraudsters intercept the tweet with a link to a fake support site that tries to steal her actual bank account credentials.

Let me walk you through some fake things that are going on online social media also. Here isone whichisfake customer service accountswhichis, IhaveaproblemIactually post a tweet saying I have problem with this bank. For example in this case everytime I havebeenonmybank'swebsitelately,ithasnotbeenworking,frustrating.Theyare actually tagging the right bank; they are tagging the right organization. For example, we could actually think of the same thing tagging HDFC Bank, ICICI Bank. These kinds of organizations have legitimate accounts and people using Twitter can tag them.

So, what happens now? This is a real tweet and real customer asking for real problem. What the fraudsters do is they look at this tweet, they have mechanisms to figure out these kind of tweets are going on. They actually reply to these tweets as though it is the bankwhichisreplyingto thistweet.And theywillcreateaccountswhichareverysimilar to the real account and reply to the post as though the real account, real organization is actually talking to you. In this example the Usual Studio Dear Charlee, We sincerely apologize for this - loginto your account via secure signonchannelblah blahblah.

This is the customer service account; fake things that is going on, whichis realcustomer tagging or connecting to a real bank organization or an organization. The fraudsters createaccountswhicharevery closeto therealaccount andactuallystart interactingwith the customer.Theseisfakecustomerserviceaccount problem.Thisparticularexampleis on Twitter, but one could think of such problems being on all social networks also. Because all of these legitimate organizations are actually using social media to interact with their customers.

(ReferSlideTime:11:48)



Fake comments on popular posts

[redacted]

Salahdin Mrabit
Comment at 12/20/15 9:50:37 AM
Links: http://9879846548.blogspot.com/2015/12/Insurance.html

⑱👉➡ http://goo.gl/OjiCvo (http://9879846548.blogspot.com/2015/12/Insurance.html) 👈⬅😱New Nicki Minaj Scandal and Sexy Tape Video Leaked
See less

View on Facebook | Posted via: Web | Deleted

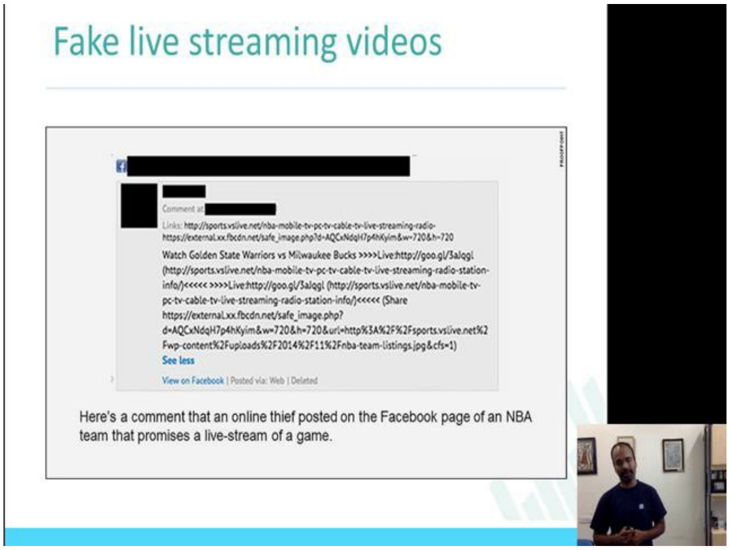
Scammers often pretend to be Facebook users so they can comment on posts that lead to a credit card phish.

Here is the second one, fake comments on popular post. I think some of the posts thatthat become very popular. Let us take the prime minister was talking about it, if it was Obamawho istalkingabout somethingthesepostsbecomeverypopular.Andwhenthese posts become verypopular there arealso lot ofcomments. For example, nowI amsureif you look at the Olympics Facebook page or the twitter handle or the hashtag, people are actually talking a lot about things that are going on in the Olympics in the context of Facebook page and Twitter accounts also.

Therefore, what scammers do is that they actually pretend to be Facebook users so they can comment on this. For example, I could create an account, I could create Facebook account which looks very legitimate, I can start posting on these Olympics relevant post which are very popular and I will kick you from that to a fake website, and get your information.And if you too click on this link I will give you also down somemalware inyour code and things like that.

So that is a second type of a fake thing that is going on, which is fake comments on popular posts because the reason why it is popular post is that it gets in more and more fashion, and more people actuallyget to see it, it isconnected to the topic that people are more interested on.

(ReferSlideTime:13:12)



Fake live streaming videos

[redacted]

[redacted]

[redacted]

Comment at [redacted]

Links: http://sports.vslive.net/nba-mobile-tv-pc-tv-cable-tv-live-streaming-radio-
https://external.xx.fbcdn.net/safe_image.php?d=AQCxNdqH7p4hKyim&w=720&h=720

Watch Golden State Warriors vs Milwaukee Bucks >>>>Live:http://goo.gl/3aJqgl (http://sports.vslive.net/nba-mobile-tv-pc-tv-cable-tv-live-streaming-radio-station-info/)<<<<< >>>>Live:http://goo.gl/3aJqgl (http://sports.vslive.net/nba-mobile-tv-pc-tv-cable-tv-live-streaming-radio-station-info/)<<<<< (Share https://external.xx.fbcdn.net/safe_image.php?d=AQCxNdqH7p4hKyim&w=720&h=720&url=http%3A%2F%2Fsports.vslive.net%2Fwp-content%2Fuploads%2F2014%2F11%2Fnba-team-listings.jpg&cfs=1)

See less

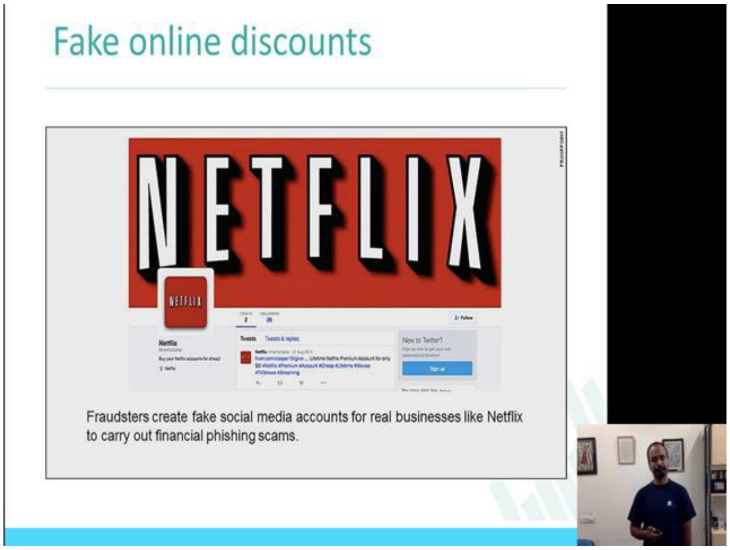
View on Facebook | Posted via: Web | Deleted

Here's a comment that an online thief posted on the Facebook page of an NBA team that promises a live-stream of a game.

The third one is fake live streaming videos, which is particularly in the context of Olympics and cricket matches, world cups and things like that, there is tendency of actually looking for these matches in live. Here is an example where this post is actually saying live video for thismatch, right?Ifyou are interested inwatching it inyour laptop, in your phone you tend to actually look at these pages, look at these links which talks about this game and tend to actuallytaking into a fake website.

Fake live streaming videos, which is there is no video, there is no real video which is connected, but the scammers actually tend to take the users to fake things. And they do this in the context ofsome games that are going on, some eventsthat are going on, some shows that are going on. For example, currently in terms of Rio somebody says that currently India has won medal and here is the video of the match. So, that is the kind of scamthat is going on in the context oflive streaming videos.

(ReferSlideTime:14:27)

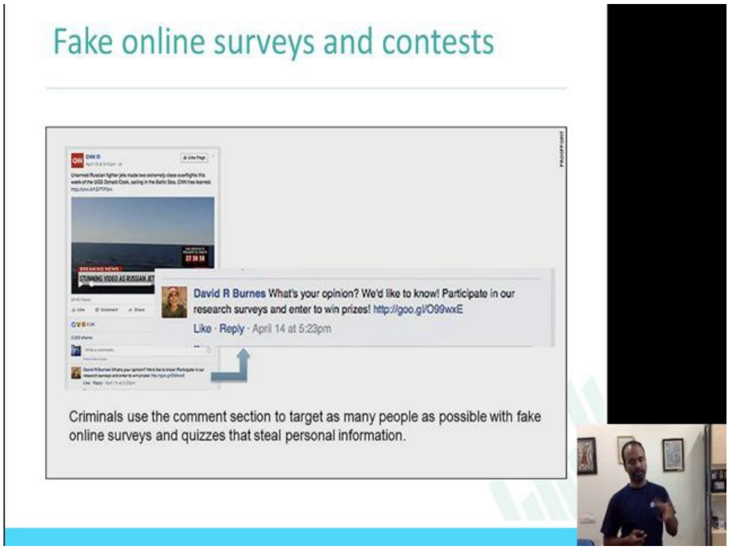


Fake online discounts

Fraudsters create fake social media accounts for real businesses like Netflix to carry out financial phishing scams.

The next one is fake online discounts which is, scammers take the real account, real organization in this case - Netflix, it could be anything Facebook, it could be Flipkart, it couldbeanyrealorganization.Theycreatefakeaccountsthat lookslike realbusinessand they are actually carry out business using these fake list, but giving you discounts. Like forexampleNetflixcouldsaythat,thispagewhichisafakepage,itcouldsaythatthereis a 10 percent discount in Netflex account that you open now. 40 percent discount for the next 6 months, if you open the account right now. These kind of posts can actually lure people into using these fake accounts, fake pages, fake services. So, that is the next fake crime that I thought we will talk about.

(ReferSlideTime:15:25)



Fake online surveys and contests

David R Burnes What's your opinion? We'd like to know! Participate in our research surveys and enter to win prizes! http://goo.gl/O99wxE

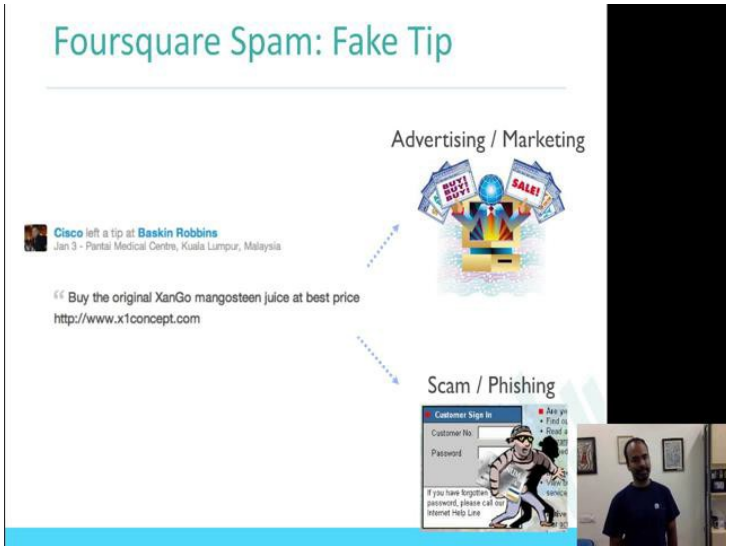
Like · Reply · April 14 at 5:23pm

Criminals use the comment section to target as many people as possible with fake online surveys and quizzes that steal personal information.

Next type is, Fake Online Surveys and Contests. These kind of scams have been around foralong,longtime,wherethecriminalsofthesescammersget youto get survey,fillthe surveyto get somemoney, to getsomeinformation.Forexample,howdoyouknowyour personality? Personality test and find out other people who are bonding your date, who hasthesamepersonalityand thingslikethat, whilethesekindofthingshavebeenaround for a long time.And there were also contests, win1000 Rupees for filling onthis survey. So these have being there in traditional ways now these have moved on to the social media services. Here is an example where, what is your opinion, we would liketo know, participate in our research surveys and enter to win prizes, here is the link.

Again this is a fake claim, this could actually be malicious, and this could actually be collecting personal information. But the source of starting this is getting you to click on the link a survey or contest. So, that is the last cyber fake version that I thought I would actuallymentionittoyou.Quicklyafewfakecrimesthatyoucanthinkof-fake customer service account, fake comments on popular post, fake live streaming videos, fake online discounts, and fake online surveys and contests. So, these are the different types ofcrimes that can be go scams that can be happen on socialmedia services.

(ReferSlideTime:17:01)


Foursquare Spam: Fake Tip

Cisco left a tip at Baskin Robbins
Jan 3 - Pantai Medical Centre, Kuala Lumpur, Malaysia

“ Buy the original XanGo mangosteen juice at best price
http://www.x1concept.com

Advertising / Marketing

Scam / Phishing

BUY! BUY! BUY!
SALE!
Customer Sign In
Customer No.
Password
If you have forgotten password, please call our Internet Help Line

HereisfewmoreI,ImeanIthinkifyoulookatmyfirstslidewhereIshowedyoudifferent types of fake crime things I was going to talk about, fake was that part. Now here is another one which is a Fake Tip: Foursquare is the most popular location based social network. Inthisfoursquarefor exampleI couldactuallywalkinto IIITandthensaythat I have checked into IIITDelhi. So that isthecheckin, andyoucanalso leaveatip,I go to Saravana Bhavan. I eat food at Saravana Bhavan and I say that the food is pretty good. So, in that tip, people actually, the scammers and the criminals actuallypost information that can take you to a fake website.

For example, here by the original XanGo and mangosteen juice at best price, this link. This is the tip posted on particular location, so it is taking you to link which could be actually phishing so, it is also studying, giving you information, advertising about such certainproductd. So that is the fake tip, the informationthat are posted ina tip that isnot relevant to that particular venue, andtaking youto afakecontent istheproblemhere.So, that is the fake account foursquare.

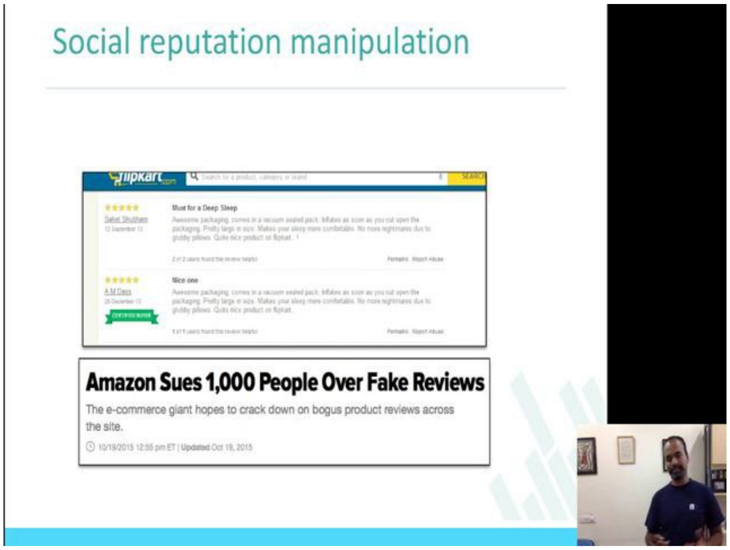
(ReferSlideTime:18:23)

Social reputation

Social reputation has become such a big deal now, everybody talks about I have 2.5 million users and then 2.5 million followers, then the number of likes that you have on your page is becoming the way that people measure your influence in the society. Even among friends, it does not have to be the celebrities, politicians, evenamong friends you are more, more curious about how many friends other have. The social status is now being measured by the presence in social media; by the number of likes that you get on posts, number offriends that you have, it is becoming more and more popular.

For example, Facebook likes andAmazon reviews,YouTube likes, the endorsement that happens on LinkedIn where you are endorsed for a particular topic, how many people have endorsed you, what topics have you been endorsed. These are becoming a measure ofinfluenceinthesociety,numberoftweets,numberoffollowers,numberoffollowings, all of them become a measure by which people think of your social reputation. But the problemisallofthemalsohaveproblems,becauseofthesewaysbywhichcreatingsocial reputation has happened, you can actuallymanipulate these socialreputation also.

(ReferSlideTime:19:48)



Social reputation manipulation

Saket Shubham
12 September 13

Must for a Deep Sleep

Awesome packaging, comes in a vacuum sealed pack. Inflates as soon as you cut open the packaging. Pretty large in size. Makes your sleep more comfortable. No more nightmares due to grubby pillows. Quite nice product on flipkart.

2 of 2 users found this review helpful

A M Dass
28 December 13

Nice one

Awesome packaging, comes in a vacuum sealed pack. Inflates as soon as you cut open the packaging. Pretty large in size. Makes your sleep more comfortable. No more nightmares due to grubby pillows. Quite nice product on flipkart.

1 of 1 users found this review helpful

Amazon Sues 1,000 People Over Fake Reviews

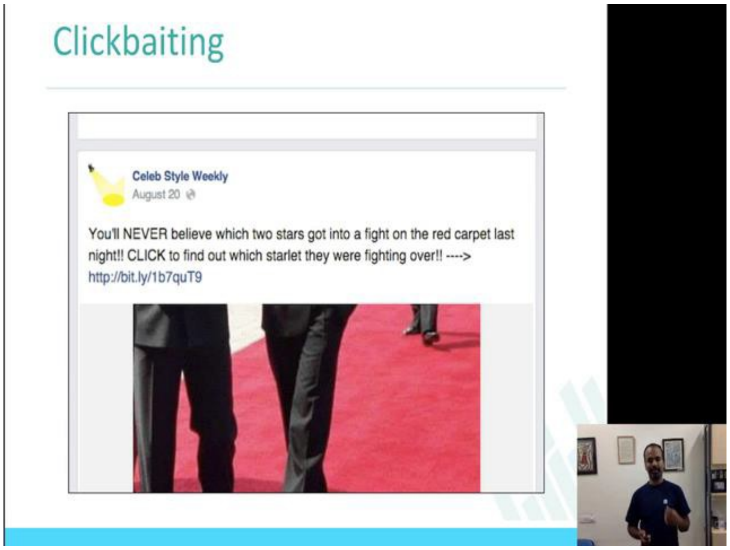
The e-commerce giant hopes to crack down on bogus product reviews across the site.

10/19/2015 12:55 pm ET | Updated Oct 19, 2015

In this case, some examples that I had put in here is Flipkart, social reputation can be manipulated by actually writing good reviews about product. So, reviews become a big way by which you can actually manipulate the social reputation of the product, of the company, of the seller, all of them can actually be manipulated. It is actually a very big problemintermsofstudyingAmazon's reviewsor Flipkart'sreviewsalso for products.

Here is a case;Amazon sues 1000 people over fake reviews. People have been studying withreviewsproblemforalong,longtime.Itisnot onlyreviews,itisalsoaboutstudying the fake followers, studying the fake endorsements, all of them are actually relevant problems. If anybody is interested in taking up some of these, these are actually very interestingproblems,verychallengingproblemsalso andveryrealworldproblemswhich is, you canactuallylook at the solutions that you build, becoming /influencingpeople's thinking.

(ReferSlideTime:20:52)



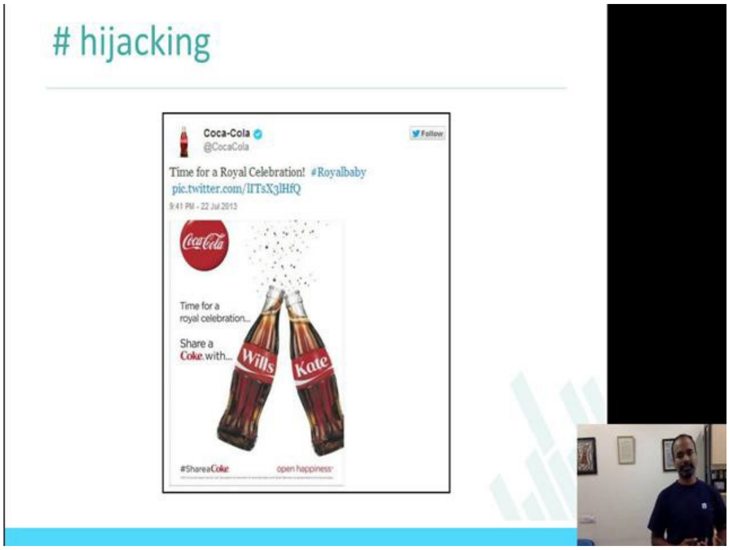
Clickbaiting

Celeb Style Weekly
August 20

You'll NEVER believe which two stars got into a fight on the red carpet last night!! CLICK to find out which starlet they were fighting over!! ---->
http://bit.ly/1b7quT9

Here is another problem in terms of crimes on socialmedia. Clickbaiting, where you are actual director your keeping the website, so you go read a particular page of news or something,theretheypresentyouwithinformationwhichissometimesrelevantsometime not relevant and they take you to a fake website. So, here in this case also, the link here, this information was actually presented in one of the social media services where it was taking it to a fake website. Clickbaiting - getting you to click on links which are not legitimate.

(ReferSlideTime:21:31)



hijacking

Coca-Cola
@CocaCola

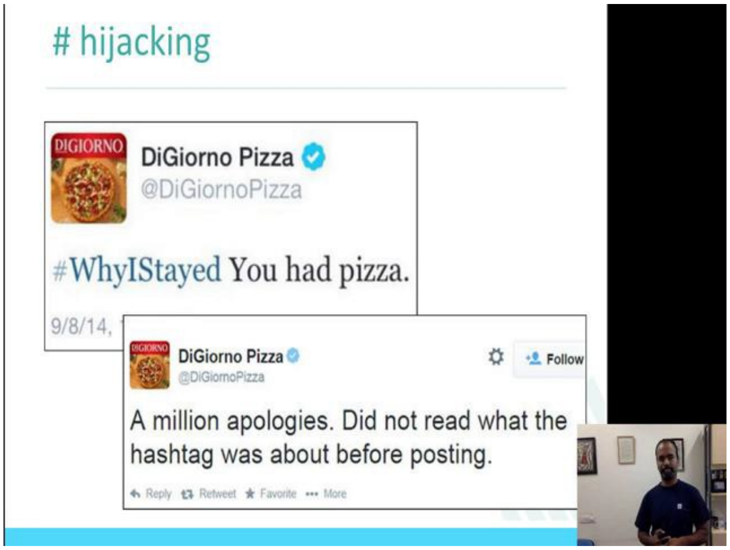
Time for a Royal Celebration! #Royalbaby
pic.twitter.com/lITsX3lHfQ

9:41 PM - 22 Jul 2013

Hashtaghijacking;hashtaghijackingisalsobecomingabigissuethesedays,Iassumeall of you know what a hashtag is. Hashtag is the waybywhich a particular set oftweets, if you want to talk about now Olympics you use hashtag Rio 2016. So that is the way of using hashtag Rio 2016, you are saying that the content that I amposting isconnected to this topic, so Twitter canactuallybring inallthesepostswhichhashashtagRio 2016and show it to people who are interested init. So, that is the logic behind using a hashtag.

InthisexamplewhereCocoColahasactuallyposted tweet whichsays, ‘TimeforaRoyal Celebration hashtag Royalbaby’. Here what coke is doing is, coke is actually using a hashtag which is very popular otherwise for actually selling their product. That is hijacking right, royal baby is nothing relevant to coke. They are kind of using it to promote their products. So that is one wayof hijacking the hashtag.

(ReferSlideTime:22:40)



hijacking

DiGiorno Pizza
@DiGiornoPizza

#WhyIStayed You had pizza.

9/8/14,

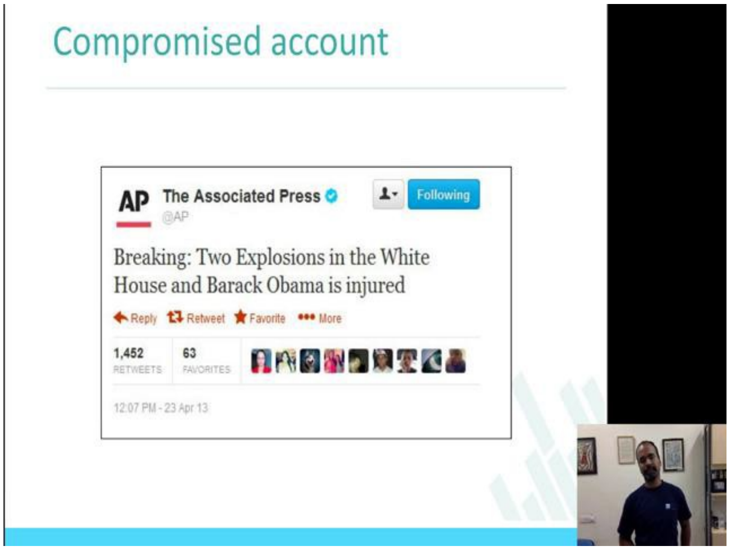
DiGiorno Pizza
@DiGiornoPizza

A million apologies. Did not read what the hashtag was about before posting.

Here is another example also. WhyI stayed was a hashtag that was trending, was getting popular so this pizza,DiGiorno Pizza thought of using this popular hash tag actually to sell to mention about pizza. They used this hashtag why I stayed because you had the pizza, but unfortunately this also back fire, here is the post that they had to actually apologize for doing this post. A million apologies, did not read what the hashtag was about. The hashtag was actually used in some of the context where people were actually using thishashtag talk about a particular situation.Therefore, taking thehashtagwhichis not relevant to thistopic, using it for selling a product isactuallyhijacking.

Nowjust furthertalkingabout forexample,youwouldsaysomethingabout whatyouare doing now with the hashtag Rio 2016 which will actually show up on people who are looking at timeline for the posts which has Rio 2016. So that is the problem in with hashtag hijacking.

(ReferSlideTime:23:49)



Compromised account

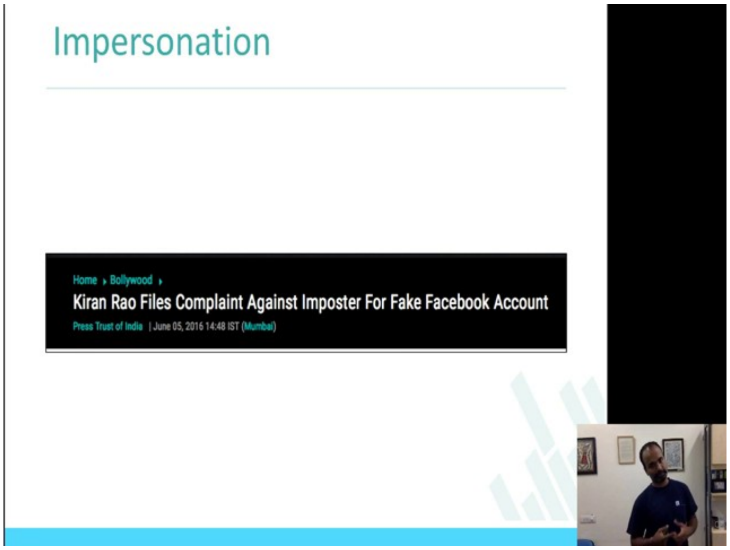
The Associated Press @AP

Breaking: Two Explosions in the White House and Barack Obama is injured

Compromised account, I have actually shown this particular tweet even in my trust and credibility section, but I brought this back just to tell you different problem, I think I explained the problem then but I will explain it in the context of e-crimes also. Compromisedaccount, whereTheAssociatedPressisaverifiedaccount andthisaccount was compromised for sometime which is, somebody else had access to this account and the tweet was, Breaking: Two Explosions in the White House and Barack Obama is injured'I amsure you can allagree that the effect this tweet must have had.

This is account compromised, somebody else getting access to your account because of leak of username password and getting that to misuse, getting the account to be misused also. That is compromised account.

(ReferSlideTime:24:40)



Impersonation

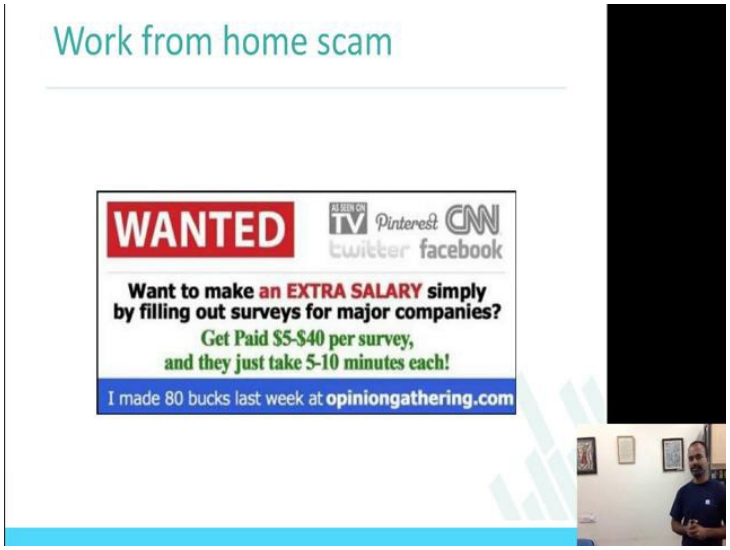
Home ▸ Bollywood ▸

Kiran Rao Files Complaint Against Imposter For Fake Facebook Account

Press Trust of India | June 05, 2016 14:48 IST (Mumbai)

Impersonation: Impersonationisalso another problemwhichisIcantakeanaccount like for example, any of you in the class I can take some details of you that I know pictures, and your city, and the information that I could collate from online sources, use that to actually create an account which as though looks like it is you. Here is a complaint that Kiran Rao has actually filed saying that fake account has been created, and there are many, many fake accounts like this. If you know remember the policing section I also showed you about the fake account of police organizations also.And it is not just about individuals, even organization's accounts are actuallycreated fake.

(ReferSlideTime:25:28)



Work from home scam

WANTED

Want to make an EXTRA SALARY simply by filling out surveys for major companies?

Get Paid $5-$40 per survey, and they just take 5-10 minutes each!

I made 80 bucks last week at opiniongathering.com

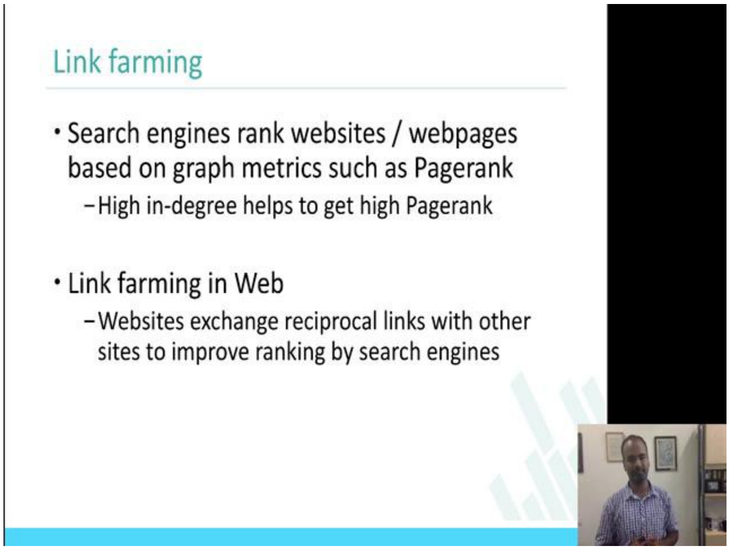
Here is another interesting problem which is, Work from home scam.Againthese things have been in traditional ways for example, if you are driving down somewhere in the signal, you will see a post which says, 'want to won 1000 Rupees a day sitting at home pleasecallthisnumber'thesekindofscamsarebeingthere.Hereisanexampleofascam that went popular in Pinterest where this image was actually floating around, 'want to makeanextrasalarysimplybyfillingoutsurveyformajorcompanies,hereisawebsiteto go to.Youget paid5to 40dollarsper survey.Thisisworkfromhomescam.Againthere isalot ofscamswhicharesimilarto theseworkfromhomescams,different versionsthat are verypopular on socialnetworks. So, this is an important scamalso.

With that I will actually wrap up my first part of the week 6, where I thought I will just introduce you to different scams, different crimes, because we will talk about crimes in this week, looking at different crimes some data was collected, what kind of analysis could be done, what kind of solutions that we could build in reducing these problems of crimes on social networks.

eCrime on Online Social Media Part-II.

Once again welcome back. This is Privacy and Security in Online Social Media, week 6 the second part.

(ReferSlideTime:00:18)



Link farming

- Search engines rank websites / webpages based on graph metrics such as Pagerank
  - High in-degree helps to get high Pagerank
- Link farming in Web
  - Websites exchange reciprocal links with other sites to improve ranking by search engines

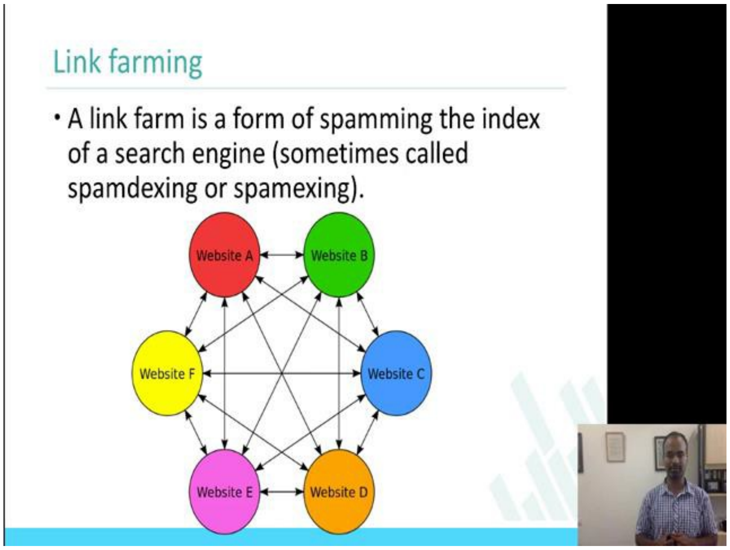
Also inthe first part, wegenerallysawabout whate-crimesare, different typesofcrimes on social networks, specific examples about different crimes; how that affects yor social reputation? How malicious users are actually making use of these social network interactions and topics are answered.

So, what we will see now is some specific problem in socialnetworks, crimes and issues on social networks and we will takesome one data set and answer some questions with that data set. Also search engines rank websites basically the Pagerank idea where every page is linked to everypage and Pagerank ofthe rank ofeverypage increases depending on the links that it has with the pages and. So, essentially if you have more of high in-degree helps in increasing the Pagerank.

So, Googleworksonthistechnology, where youactuallyhave;youcreateawebsite, you link it to, let us take to IIITD’s website and IIITD links it back to you, then I think your Pagerank increases heavily, so that is simple idea for Pagerank, but link forming in onthe web is basically an idea where websites exchange reciprocal lengths with othersites to improve the rank. So, the idea of making the links between websites which is not otherwise there; creating links or increasing the links ofthe websites to other websites is actually link farming.

Same ideas are connected to Pagerank. Pagerank is benign or legitimate links that youcreate. link farming is the idea in which these links are created which are not benignones.

(ReferSlideTime:02:12)

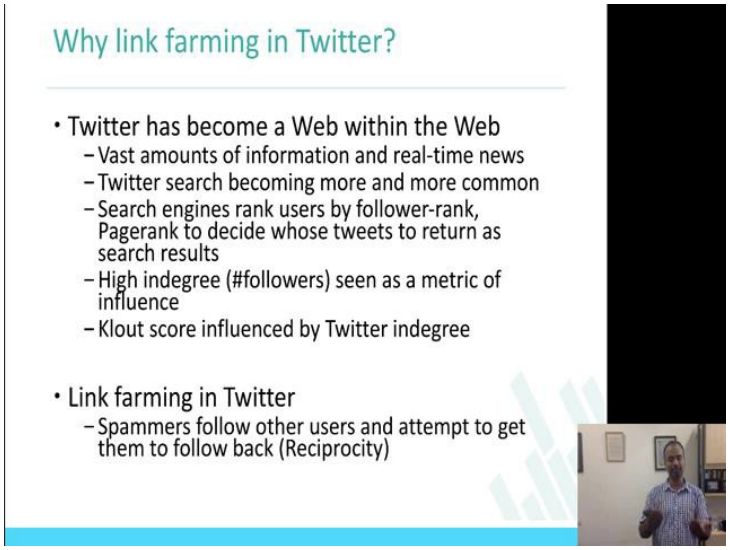


Link farming

- A link farm is a form of spamming the index of a search engine (sometimes called spamdexing or spamexing).

So,hereisasimplediagramto showthat what,how link farmingorwhat link farming is. A link farm is a form of spamming the index of the search engine which is essentially increasing the links between different websites like, for example, website and website A and B. Allthe, if you start creating links betweenthese websites, iftheydo not exist and that is called actually link farming. Sometimes, it is also called spamdexing and spamexing. So, that is the idea for link farming. Link farming is a way which non legitimate links are created betweenthe websites. The idea for doing this is when you do this and when you increase the in-degree which is the links that are coming into the website increases then the Pagerank of the website automatically increases.

(ReferSlideTime:03:04)



Why link farming in Twitter?

- Twitter has become a Web within the Web
  - Vast amounts of information and real-time news
  - Twitter search becoming more and more common
  - Search engines rank users by follower-rank, Pagerank to decide whose tweets to return as search results
  - High indegree (#followers) seen as a metric of influence
  - Klout score influenced by Twitter indegree
- Link farming in Twitter
  - Spammers follow other users and attempt to get them to follow back (Reciprocity)

So, whylink farming inTwitter?So, whatwearegoingto studyis, wearegoingtostudy the ideaoflink farmingspecificallyonlyinthecontext ofTwitter.So,whylink farming? What does it help? Who benefits because of actually link farming on Twitter? So, essentially Twitter, I mean given its nature, given amount of data that is gettinggenerated on Twitter, it is basically a web within the web.

If you want to go to, go and look at the out breaking news, Twitter is the place to start with. It has large amount of data on real time news. There is multiple research done to show that Twitter is where news breaks. If you want to look at the latest in now, year 2016, Twitter is probably the best place to look at and people start searchingfor a topicin Twitter actuallymeaning Twitter definitelyis a micro blogging website, where people push content, but if you look at the pattern in which Twitter is being used, it also being used for search for a topic, live search into people.

So, when you search for atopic inTwitter,thewaythat theresultsarepresented depends on many factors. So, search engines in Twitter can rank, actually follow ranks. It can actually present the results depending on the connections that you have with the person who is talking about that topic like, for example, ifI search for a topic like hashtag year 2016. If any of my friends are talking, it could show up on top. I mean itcould show up onthenumber offollowersthat people have. Ifit isa verified account, it should showup on top.

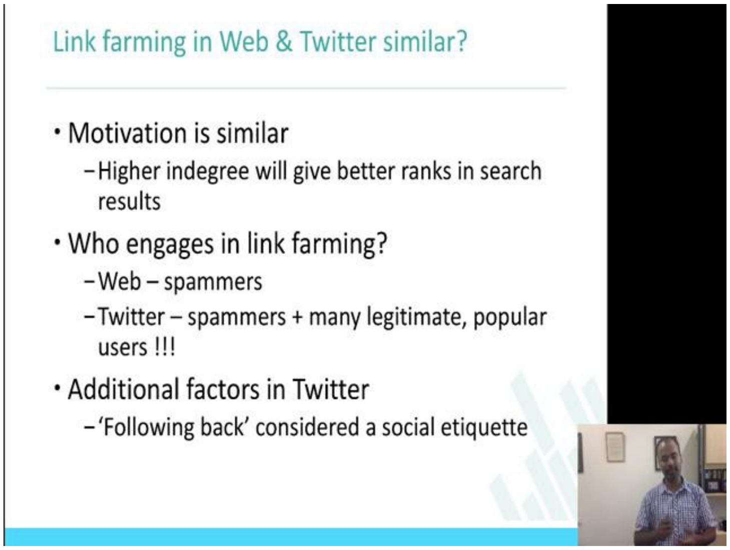
So, the search results can be; search results can be used, these kinds of techniques. PageRank would be one, which is how many people are actually connected to this particular tweet and who are the users, who are connected to this particular tweet. All of this information can be used to actually decide on presenting the search results. Of course, the way that the search results are presented is actually going to bias the users to goto, iftwitter is showing youtheresultsontop5there is more likelythat youaregoing to actually go look at those particular tweets.

And of course, high in-degree which I think when mentioned the part Pagerank, I said, high in-degree which has number of followers seen as a matter of influence on. number of followers that you have is a measure of social reputation. I also mentioned this in the last part of the week 6. There is a score called Klout. So, I mean I would recommend again you people look at what Klout score is, Klout score is essentially a way by which Klout collates all your online presence, particularly in the social media, and gives numbers toit. For example, my score would be 24, which basically says that on scale of1 to 100 what kind of influence are you having on the social media services.

Klout is an interesting mechanism that also, a paper which talks about how Kloutactually finds out these values and researchers have used Klout score as a way tomeasure the influence of the users also. Of course, the topic of influencer by itself is actually hard because you are defining who an influencer is; it is becoming more and more difficult. So, while link farming in twitter is basically a large amount of data is getting generated, real time information is spread there and when users search for topic, the information is actuallypresented depending onthe links, depending onthe Pagerank, depending on the links that follow a rank depending on the links that users are.

And particularly link farming in Twitter is basically, spammers follow other users and attempt to get themto follow back also. Essentially, how do they increase the in-degree, the in-degree is increased if I am a spammer, I start following thousands and thousandsof people and there is a probability that you will actually; the people that I am trying to follow, now will actually follow me back. And again, there is multiple researchers, people who have shown, how the reciprocity can be, there is a high probability that if I follow you, you will follow me back and, giving, with that effect, the link farming actually increases on increases and therefore, twitter can be used to increase the linkfarm.

(ReferSlideTime:07:47)



Link farming in Web & Twitter similar?

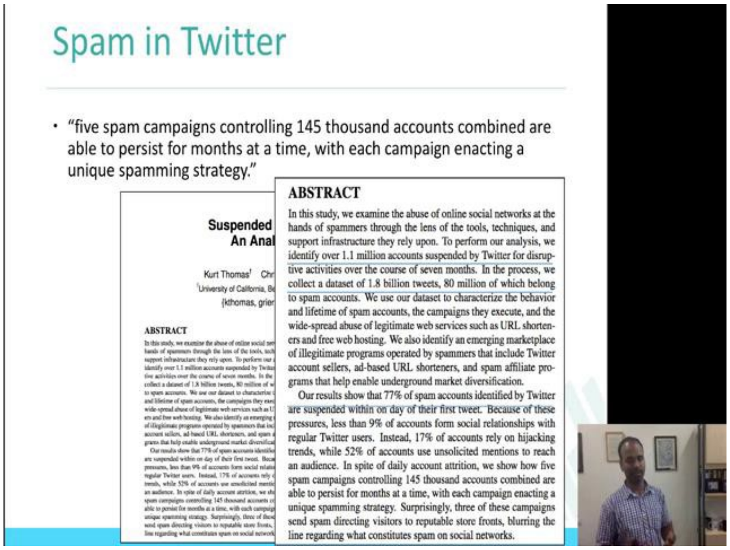
- Motivation is similar
  - Higher indegree will give better ranks in search results
- Who engages in link farming?
  - Web – spammers
  - Twitter – spammers + many legitimate, popular users !!!
- Additional factors in Twitter
  - 'Following back' considered a social etiquette

So, here is a slide to show the differences and similarities of link farming in web and Twitter. In the web increasing my Pagerank, increasing my in-degree, increases my probability of showing up in the search results. In the Twitter space, increasing the in-degree actually increases the gain; similarly, to show on mytweets onthe search results.

In the webs, spammers actually use link farming. In Twitter, spammers do actually link farming, but it is also done bylegitimate and popular users, I think that is the whole idea withwhere youactuallyincreasethe in-degree bymaking your numberoffollowers high and therefore, you can actually your content can actually be presented to a large number largeset ofusers. And ofcourse, inthecontext ofTwitter, inthecontext ofweb, it is not necessary that if I link your website you are probably going to link back to my website;hyperlinks are not created in that way.

Whereas in the context of Twitter there is a high probability that, let us take if I am, I actually follow one of the students who are taking this class, there is a high probability that the studentis going to follow me back again and the same way if I follow aprofessor and the professor probably there is a high probability there the professor will follow me back.

(ReferSlideTime:09:17)



Spam in Twitter

- "five spam campaigns controlling 145 thousand accounts combined are able to persist for months at a time, with each campaign enacting a unique spamming strategy."

ABSTRACT

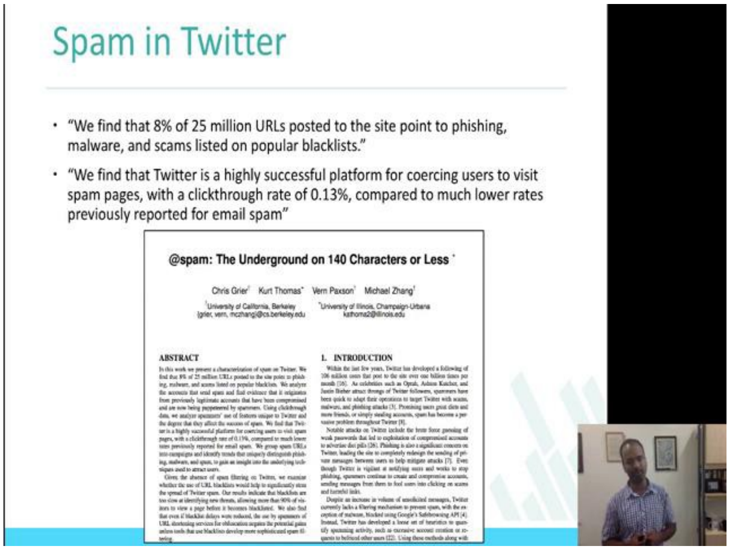
In this study, we examine the abuse of online social networks at the hands of spammers through the lens of the tools, techniques, and support infrastructure they rely upon. To perform our analysis, we identify over 1.1 million accounts suspended by Twitter for disruptive activities over the course of seven months. In the process, we collect a dataset of 1.8 billion tweets, 80 million of which belong to spam accounts. We use our dataset to characterize the behavior and lifetime of spam accounts, the campaigns they execute, and the wide-spread abuse of legitimate web services such as URL shorteners and free web hosting. We also identify an emerging marketplace of illegitimate programs operated by spammers that include Twitter account sellers, ad-based URL shorteners, and spam affiliate programs that help enable underground market diversification.

Our results show that 77% of spam accounts identified by Twitter are suspended within on day of their first tweet. Because of these pressures, less than 9% of accounts form social relationships with regular Twitter users. Instead, 17% of accounts rely on hijacking trends, while 52% of accounts use unsolicited mentions to reach an audience. In spite of daily account attrition, we show how five spam campaigns controlling 145 thousand accounts combined are able to persist for months at a time, with each campaign enacting a unique spamming strategy. Surprisingly, three of these campaigns send spam directing visitors to reputable store fronts, blurring the line regarding what constitutes spam on social networks.

I thought I will walkyou through some literature in the context of spam in Twitter, which isin this case I do not know I think I am going to talk about 2 or 3 research results. Just to tell you, the context of where link farming is going to be kept which is spam in a broader sense. So spam campaigns, here is a paper which is titles 'SuspendedAccounts inRetrospect:AnanalysisofTwitter Spam'. So, 5spamcampaignscontrolling 145 thousand accounts combined are able to persist for months at a time. So, here is a zoomed in version ofthe abstract which reads as, we identify about 1.1 million accounts suspended by Twitter for disruptive activities over the course of 7 months. In theprocess, we collect a dataset of 1.8 billion tweets and 80 million of which belongs to spam accounts.

The problem with that comes with the spam is that it is not only high in these social networks, but thereareactuallytheyalso stayfor longer. So, here it is7 spamcampaigns controlling145000accountsandtheypersists formultiple monthsandtheanotherpartof the abstract reads as, our results show that 77 percent of spam accounts identified by Twitter aresuspended within adayoftheir first tweet.So,thesearesomenumbers, some idea to get a sense of what is happening in Twitter in the context of spam.

(ReferSlideTime:10:51)



Spam in Twitter

- "We find that 8% of 25 million URLs posted to the site point to phishing, malware, and scams listed on popular blacklists."
- "We find that Twitter is a highly successful platform for coercing users to visit spam pages, with a clickthrough rate of 0.13%, compared to much lower rates previously reported for email spam"

@spam: The Underground on 140 Characters or Less *

Chris Grier† Kurt Thomas* Vern Paxson† Michael Zhang†

†University of California, Berkeley
{grier, vern, mczhang}@cs.berkeley.edu

*University of Illinois, Champaign-Urbana
kathoma2@illinois.edu

ABSTRACT

In this work we present a characterization of spam on Twitter. We find that 8% of 25 million URLs posted to the site point to phishing, malware, and scams listed on popular blacklists. We analyze the accounts that send spam and find evidence that it originates from previously legitimate accounts that have been compromised and are now being puppeteered by spammers. Using clickthrough data, we analyze spammers' use of features unique to Twitter and the degree that they affect the success of spam. We find that Twitter is a highly successful platform for coercing users to visit spam pages, with a clickthrough rate of 0.13%, compared to much lower rates previously reported for email spam. We group spam URLs into campaigns and identify trends that uniquely distinguish phishing, malware, and spam, to gain an insight into the underlying techniques used to attract users.

Given the absence of spam filtering on Twitter, we examine whether the use of URL blacklists would help to significantly stem the spread of Twitter spam. Our results indicate that blacklists are too slow at identifying new threats, allowing more than 90% of visitors to view a page before it becomes blacklisted. We also find that even if blacklist delays were reduced, the use by spammers of URL shortening services for obfuscation negates the potential gains unless tools that use blacklists develop more sophisticated spam filtering.

1. INTRODUCTION

Within the last few years, Twitter has developed a following of 106 million users that post to the site over one billion times per month [16]. As celebrities such as Oprah, Ashton Kutcher, and Justin Bieber attract throngs of Twitter followers, spammers have been quick to adapt their operations to target Twitter with scams, malware, and phishing attacks [3]. Promising users great diets and more friends, or simply stealing accounts, spam has become a pervasive problem throughout Twitter [8].

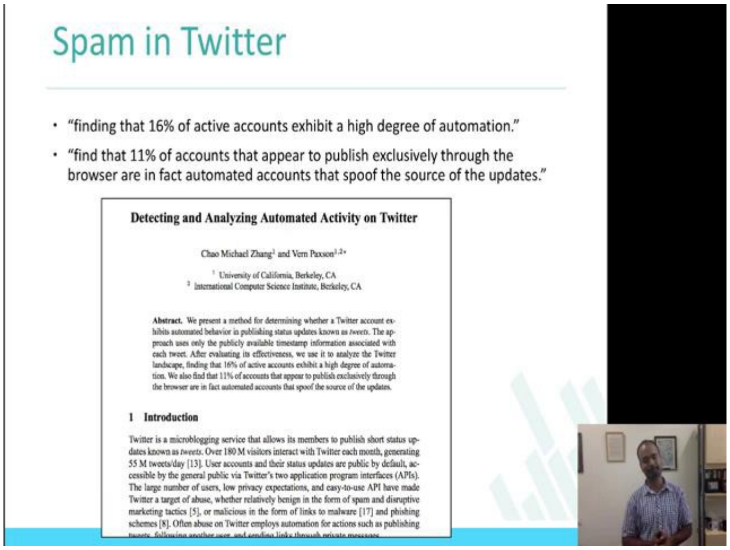
Notable attacks on Twitter include the brute force guessing of weak passwords that led to exploitation of compromised accounts to advertise diet pills [26]. Phishing is also a significant concern on Twitter, leading the site to completely redesign the sending of private messages between users to help mitigate attacks [7]. Even though Twitter is vigilant at notifying users and works to stop phishing, spammers continue to create and compromise accounts, sending messages from them to fool users into clicking on scams and harmful links.

Despite an increase in volume of unsolicited messages, Twitter currently lacks a filtering mechanism to prevent spam, with the exception of malware, blocked using Google's Safebrowsing API [4]. Instead, Twitter has developed a loose set of heuristics to quantify spamming activity, such as excessive account creation or requests to befriend other users [22]. Using these methods along with

Here is another one and this paper is also a popular paper in the context of spam on Twitter. So,this is '@spam: TheUndergroundon140 Charactersor Less'. So, whatthey found is, they found 8 percent of the 25 million URLs posted on the site pointing to phishing, malware and scams. So, that is a lot of URLs which are actually malicious, 8 percent of 25 million URLs, where they are pointing to malware, phishing, scams and malicious ones and they have also found that the click rate is actually higher.

So, they say that, we find that Twitter is a highly successful platform for coercing usersto visit spam pages with a click through rate of 0.13 percent compared to, much lower ratespreviouslyreported for email spam. So, the idea isthat if you get anemailofwhich is which says that please click on this line for buying a product with 10 percent discount that is low probability of you clicking on this link at the email, whereas if the same postis coming fromsomebodywhomyou are following onTwitterthere is a highprobability that you are going to actually check this up. So, that is the probably why this rate is actually high in the context of social networks because it is these posts are coming from your friends.

You remember the Associated Press example that we talked about earlier, where they, where I showed that the post had mentions about White House blast and then they cost actually, why because it is actually coming from Associated Press and there are many peopleactuallyfollowitanditisalsoverifiedaccount.So, thatis why theclickrateis highand astheclickratesare higher, it willactuallybecome moreand moreasuccessful spam campaign.



Spam in Twitter

- "finding that 16% of active accounts exhibit a high degree of automation."
- "find that 11% of accounts that appear to publish exclusively through the browser are in fact automated accounts that spoof the source of the updates."

Detecting and Analyzing Automated Activity on Twitter

Chao Michael Zhang[1] and Vern Paxson[1,2*]

[1] University of California, Berkeley, CA
[2] International Computer Science Institute, Berkeley, CA

Abstract. We present a method for determining whether a Twitter account exhibits automated behavior in publishing status updates known as *tweets*. The approach uses only the publicly available timestamp information associated with each tweet. After evaluating its effectiveness, we use it to analyze the Twitter landscape, finding that 16% of active accounts exhibit a high degree of automation. We also find that 11% of accounts that appear to publish exclusively through the browser are in fact automated accounts that spoof the source of the updates.

1 Introduction

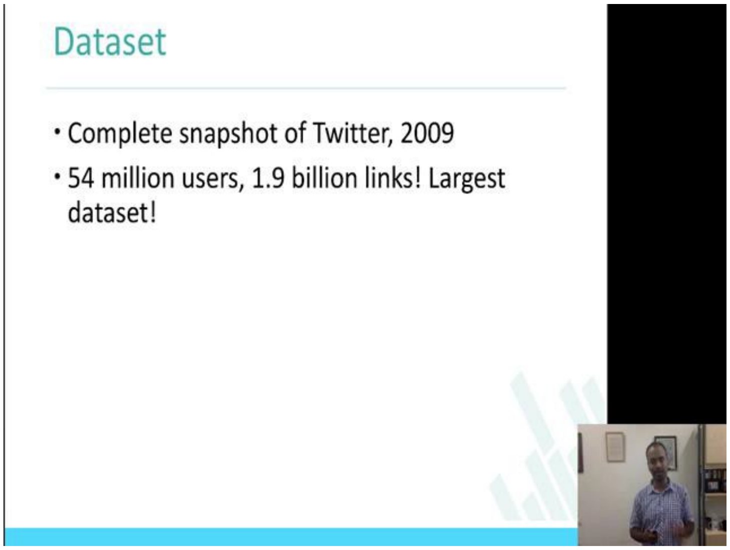
Twitter is a microblogging service that allows its members to publish short status updates known as *tweets*. Over 180 M visitors interact with Twitter each month, generating 55 M tweets/day [13]. User accounts and their status updates are public by default, accessible by the general public via Twitter's two application program interfaces (APIs). The large number of users, low privacy expectations, and easy-to-use API have made Twitter a target of abuse, whether relatively benign in the form of spam and disruptive marketing tactics [5], or malicious in the form of links to malware [17] and phishing schemes [8]. Often abuse on Twitter employs automation for actions such as publishing tweets, following another user, and sending links through private messages.

Here is the third one; third research which shows the detecting, which is titled as, ‘Detecting and Analyzing Automated Activity on Twitter’. So, what it shows is that 16 percent ofactive accountsexhibit a high degree of automation. There are again, there are multiple people working on the space in terms of actually identifying automated post on social networks, particularly on Twitter. One simple technique that people try and researchers have tried and it is also being used in some other products is to actually look at the frequency of the post that somebody does, as a human being, you and I probably will not be posting 50 tweets or 45 tweets and, whereas a bot, an automated service would actually do that.

So, people do the graph of hour of the day and minute of the hour; x axis can be the minute ofthe hour, yaxis can be the hour ofthe dayand if you draw the plot, iftheyare very, very close to each other then there is a high probability that it is actually aautomated service that is did this post.

16 percent ofthe active accounts exhibit a highdegree ofco-ordination. Theyalso found that the11 percent ofaccountsthat appearto publishexclusivelythroughthebrowser are in fact, they are automated accounts that spoof the source of the updates, that is also interesting right. Now, it is not only just the content which is spammed, it is also the spoofing of the source, how the post was done.

(ReferSlideTime:14:16)



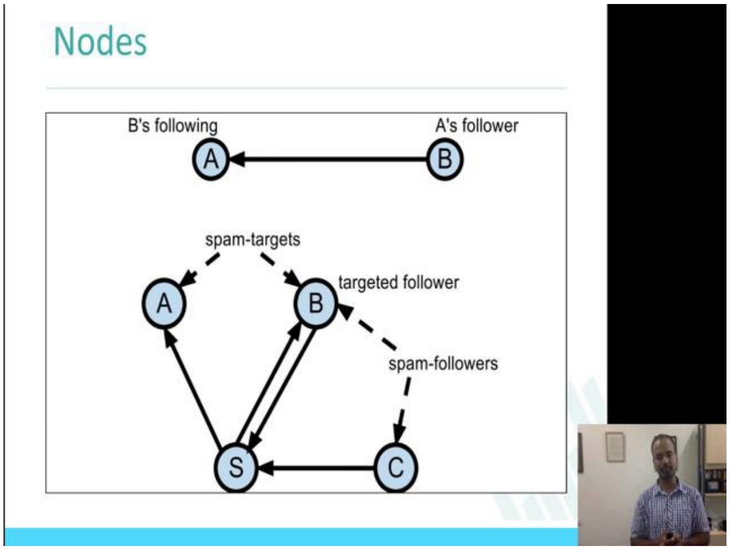
Dataset

- Complete snapshot of Twitter, 2009
- 54 million users, 1.9 billion links! Largest dataset!

So, what we are going to look at inthe specific questions that we are going to answer is, we are going to actually look at research that was done which is using the entire data set of Twitter, which was collected in 2009. It has 54 million users, I think it will be extremely hard to collect its data today because of the number of users, the connections and probably also the infrastructurethat you may need to collect this data.

So, this 2009, 54 million users, 1.9 billion links between the users and it is probably one of the largest set on data largest data set on Twitter.

(ReferSlideTime:14:54)

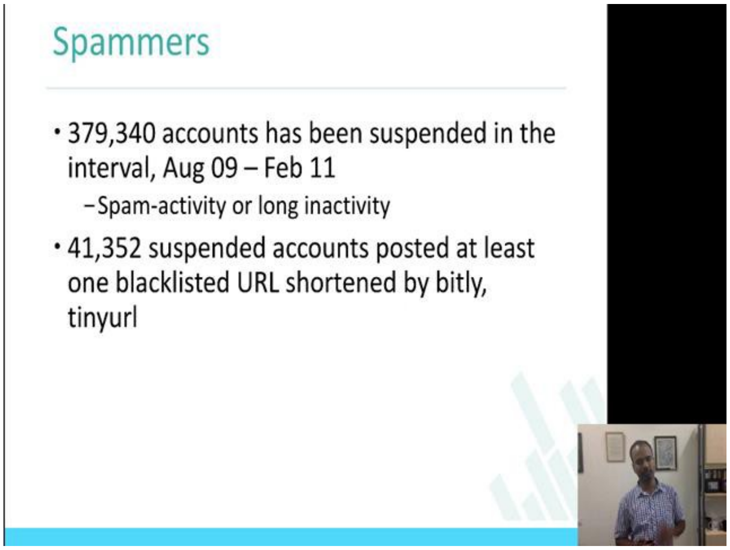


Nodes

So, how the definition of the follower and followings is used in this context is, A and B, if there is a link between A and B and going from B to A, they are marked as towards that, then B is, A is B’s following and B is A’s follower. The error marks towards it which isontheside. So,this isB, which isthe followerofAand Aisthe followingofB, I think wetalked about what Twitter is, basic terminologies, very early in the course. So, thatisfollowerand following for you. Andoverthegraphatthebottomtalksabout spam targets which is the targets where spam is going to be sent, targeted follower and spam follower.

So, B and C is basically showing you the spam followers which are which are the ones that are going to be following S, and A and B are the spam targets they are going to be getting the spam from S. So, the terms are going to be follower, following, spam targets wheretime is going to bespent, sent, spamfollowers, thosearetheoneswhich aregoing to be following. The base which is B and C or its equivalent to B and first part of the graph spam followers and of course, targeted follower and that makes sense.

(ReferSlideTime:16:35)



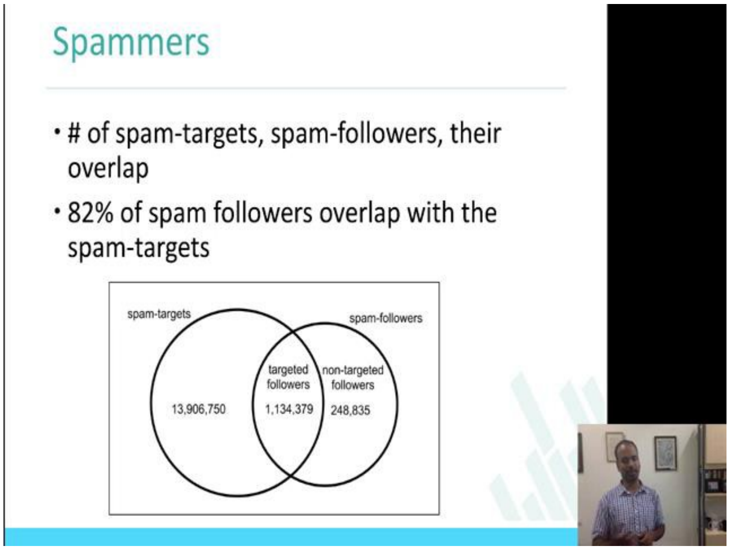
Spammers

- 379,340 accounts has been suspended in the interval, Aug 09 – Feb 11
  - –Spam-activity or long inactivity
- 41,352 suspended accounts posted at least one blacklisted URL shortened by bitly, tinyurl

So, what they formed was they formed 379,340 accounts that has been suspended in the interval of this period August 2009 to February 2011, spam activity of course, these accounts were suspended because there was a high spam activity and login activity becauseif you donotlogin toyour accountfor sometime, Twitter can actually suspend your account. 41,352 suspended accounts posted at least one blacklisted URL shortened by bitly or tinyurl. So, there is a set of URL shorteners called bitly, tinyurl.

Some of you may have used it, if not please go look at them. The idea for the URL shorteners is, if you want to share a long URL, particularly because of the social networks' influence, social networks' growth, these kinds of URL shorteners have actuallybecome very, verypopular becausewhenIwantto do apost inTwitter, which is only 140 characters, I do not want to really spend alot of a space in just posting the URL, insteadIwouldactuallysend it totheURLshortener, whichwillreducethe link, if it was like 100 characters it will just give me into bitly, bitly dot com slash some 6 or 8 unique characters which would redirect me to the actual website. So, 41 thousand suspended accounts posted at least one URL, which was actually shortened.

(ReferSlideTime:18:02)

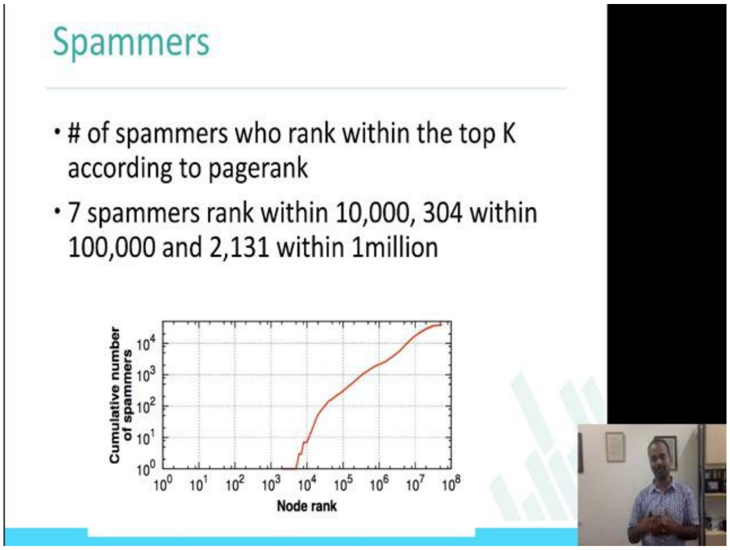


Spammers

- # of spam-targets, spam-followers, their overlap
- 82% of spam followers overlap with the spam-targets

Let us look at just the spam. So, if you remember the terminology spam targets, spam followers, entire followers. So, number of spam targets followers that the graph at the bottom, actually the Venn diagram at the bottom shows you spam targets were about 13 million, targeted followers, were about 1 million and the spam followers were about 248 thousand, 82 percent of the spam followers overlap with the spam targets alright, whichis the followers, spam followers, who are going to be following some accounts are actually part ofthe spam targets itself, 82 percent ofthe spam followers overlap with the spam targets.

(ReferSlideTime:18:46)



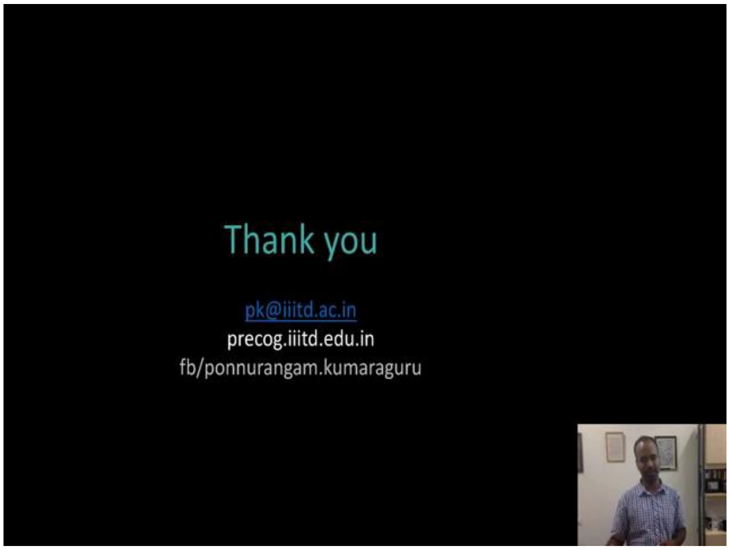
Spammers

- # of spammers who rank within the top K according to pagerank
- 7 spammers rank within 10,000, 304 within 100,000 and 2,131 within 1million

So, what isthegoodway?So,here isan interesting wayofactually looking at it thedata. So, this is cumulative, CDF, which actually shows you node rank at the x axis and cumulative number ofspammers inthe yaxis. There is some interesting conclusions thatyou can actually draw from this, which is to say that the number of spammers who rank within the top k according to the Pagerank. This is ranked according to the rank of the user. So,ifyou seewithinthe first 10,000userstherEareactually7spammers, what does this mean? This means that, if you would actually list down rank of all the users of Twitter, look at the followers, look at the node rank which is the Pagerank, in-degrees of those followers,

you can actually see 7 spammers within the top 10,000 users and 304 within the 100,000 and 2131 within the first 1 million users, which is if you take 10 lakh users, the top 10 lakh users with PageRank, with the in-degree as high, list them, you will see that two thousand users of this 10 lakh users are actually spammers alright. So, that gives you sense of you know the spammers are actually very popular alright. It essentially shows that these users have high in-degree which isthe context ofthe problemthat wetrying to study which is link farming.

(ReferSlideTime:20:30)



Thank you

pk@iiitd.ac.in

precog.iiitd.edu.in

fb/ponnurangam.kumaraguru

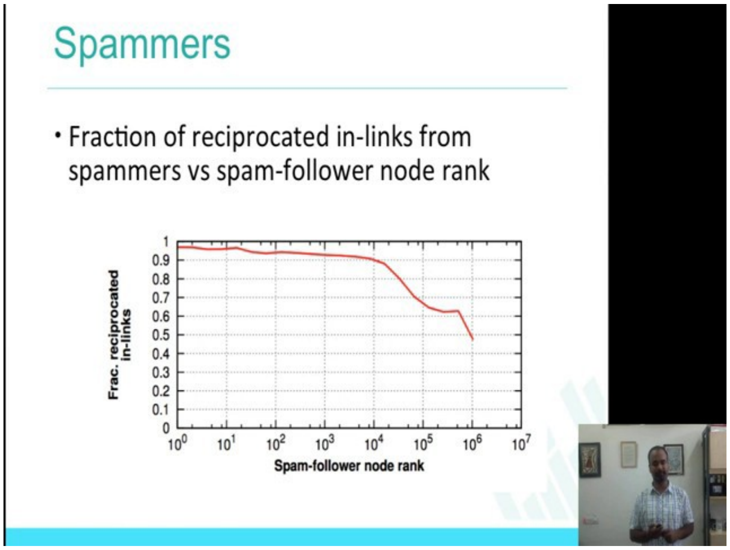
With that, I will actually stop the second part ofthe week 6. I will continue with the rest of the results and analysis soon.

Unit-4

Link Farming in Online Social Media

Welcome back to the course Privacy and Security in Online Social Media. So, this is week7, Ihopeyou gotachance tolookatthe content inweek6where welookedat link farming spam in Twitter and some kind of work which was able to find out what link farmers are what is the characteristic of link farmers. So, today we will continue, this week we will continue a little bit about the same topic, we will finish it and move onto something else.

(ReferSlideTime:00:44)



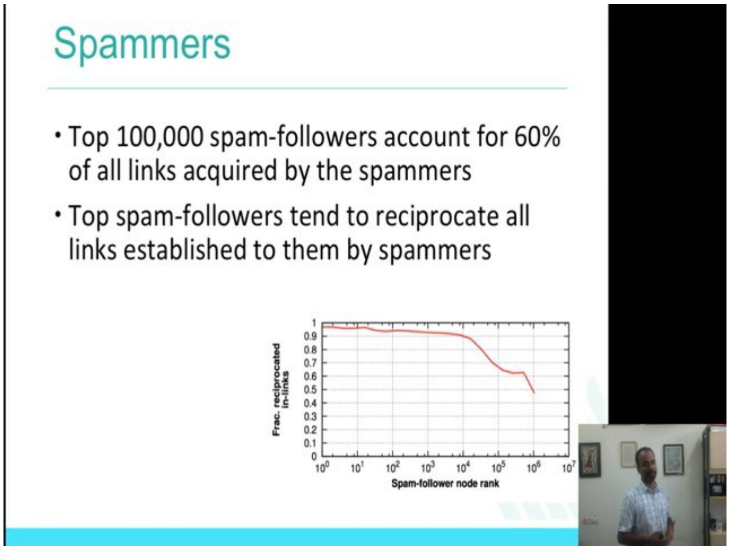
Spammers

- Fraction of reciprocated in-links from spammers vs spam-follower node rank

If you remember last week, I showed you about what is link farming and somedata about link farming. So, here is a graph which has on x axis spam follower node rank whichiswhatistheprobabilitythat spamfollowersaretheaccountswhichactually follow spam.IfyourememberthegraphthatIshowedyoua,b,c,d,e,fwheretherewas something as spam follower and then something that was spam followings.

So, x axis is spam follower node rank which is the rank of the account, which is being a spamfollower and then onthe yaxis is fraction reciprocated in links, which is I think we had briefly mentioned this last week also, the probability of you following me when I actually follow you, reciprocity so to say.If I follow you what is the probability that you will follow me back that is what is put on the y axis. So, fraction of reciprocated in links from spammers versus spam follower node ranks.

(ReferSlideTime:01:56)



Spammers

- Top 100,000 spam-followers account for 60% of all links acquired by the spammers
- Top spam-followers tend to reciprocate all links established to them by spammers

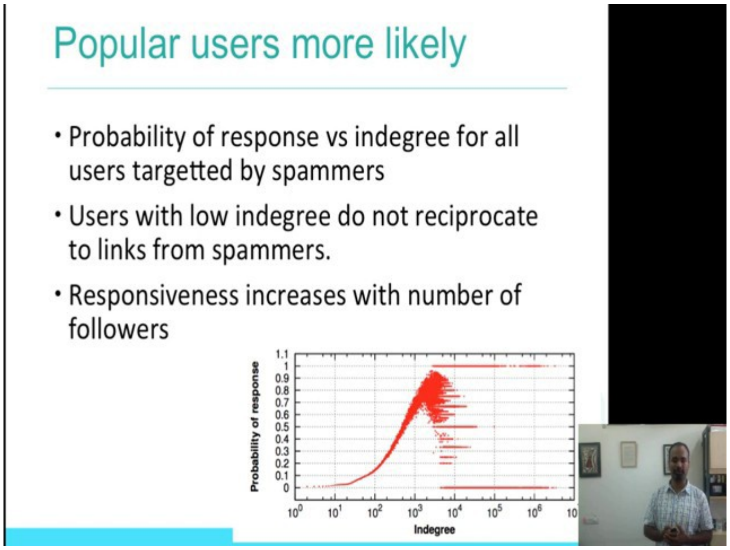
That is if you look at the top 100,000 spam follower accounts for 60 percent of all links acquired by the spammers. So, what does this mean? This means top 100,000 spam follower accounts for 60 percent of all links acquired by the spammers. So, if there were 100 links that were created for the spammers, 60 percent of them are coming from the spam followers, which is an account which actually follows back the spammers. Top spam followers tend to reciprocate all links established to them by spammers.

So, if you look at this graph the way to read this graph is on the x axis is the log scale, which is 1, 10, 100, 1000, 10000, 100000 and 1 million. So, that is how it is written on the x axis. y axis is the probability of you following me back, if I follow you. So, if you look at the first let us take a look at the first value 1 or 1 to 10. So, here the probabilities isalmostclosetoonewhichisthetoprankedspamfollowerswhichistheaccounts which are having the chances of following you back is very high. If youarrange the spam followers in the node rank which is the number of followers that they may have is actually very high.

So, what does this mean? This basically means that the probability of a spam follower following a spammer is very high, which is what spammers actually make use of, which isifthereisaprobability,ifthereisachancethatyouwillfollowmeback,thespammers will keep following people like you and there is a high chance that you will follow me back and therefore, spammers increase their in-links, which means on the topic that we discussed about link farming which means that my node rank is increasing, which is, spammers' node rank is increasing which is what they want because of that their content will show up on search, their content becomes more popular and therefore, they will probably benefit from it, I hope that connects the dots.

So, I am going to use the same data, to actually emphasize on what is the behavior of these spammer? And what the link framers do? What do the spam followers do?

(ReferSlideTime:04:38)



Popular users more likely

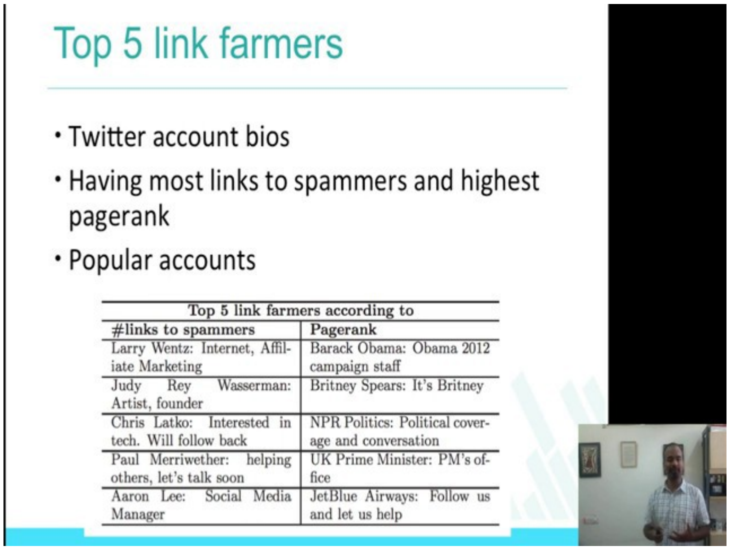
- Probability of response vs indegree for all users targetted by spammers
- Users with low indegree do not reciprocate to links from spammers.
- Responsiveness increases with number of followers

Let us look at another graph. This graph is showing the probability of response which is intermsofjustrespondingtoarequestwiththeindegree,indegreeisnumberoflinks thattheyhave.So,this is showingyou probabilityofresponsewas in-degreeforallusers targeted by spammers. If there are maybe, 10 to the power 7 users were targeted by the spammers, what is the probability that they are going to actually respond?

Users with low integrity do not reciprocate to links from spammers. If you look at the graph let us take less than 100, less than 1000 which is if I have followers which are less than 1000,, there is very less probability that I will actually follow the spammer back.Let us look at in the graph again, 10 to the power 3. Let us look at the value 10 to the power of 3, the probability of that users even then 10, 100 to 1000 if you see, the probability of somebody following back - the spammer is actually about 40 percent, 50 percent around 60 percent.

Ifyoulookatthelaterpartofthegraphonthexaxisandintheyaxis whichisabout0.7, responsivenessincreaseswithnumber offollowers,in-degreeisthenumberoffollowers. Asthefollowersincrease,thechancesofsomebodyfollowing you backwhenyou senda requestishigh.Ihopethatissinkingin,again letmereiteratethepointwhichis,onthex axis in this graph we are seeing the in-degree which is the number of followers, y axis is the probability of response when a request is sent for following or when somebody follows you, chances of you following me back. Users of less in-degree do not reciprocate to spammers; whereas, users with larger in-degree which is larger number of followers tend to follow back more.

(ReferSlideTime:06:52)



Top 5 link farmers

- Twitter account bios
- Having most links to spammers and highest pagerank
- Popular accounts

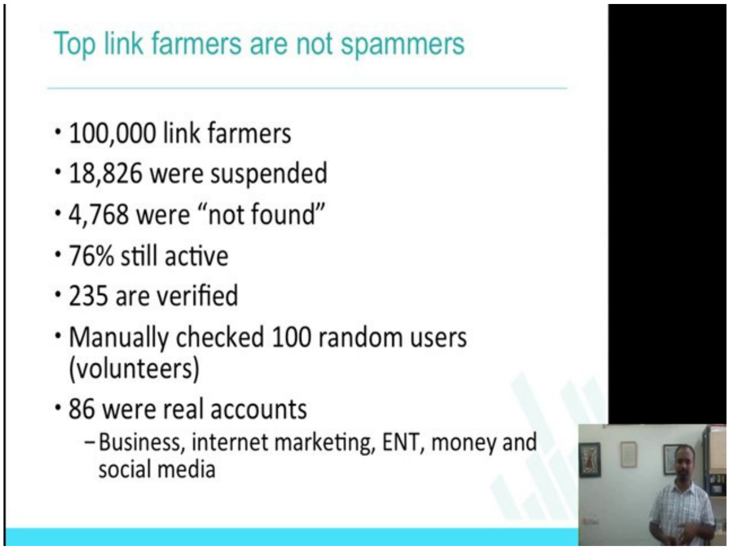
Top 5 link farmers according to	
#links to spammers	Pagerank
Larry Wentz: Internet, Affiliate Marketing	Barack Obama: Obama 2012 campaign staff
Judy Rey Wasserman: Artist, founder	Britney Spears: It's Britney
Chris Latko: Interested in tech. Will follow back	NPR Politics: Political coverage and conversation
Paul Merriwether: helping others, let's talk soon	UK Prime Minister: PM's office
Aaron Lee: Social Media Manager	JetBlue Airways: Follow us and let us help

So, what they did was after looking at these two things, which is the probability of somebody following spammer part is high, they actually looked at the top five link farmers and looked at the bios, what are the accounts and here is a sense of what these accounts are Larry Wentz, Internet affiliate, Marketing; Judy Rey Wasserman, artist, founder.

So, these are all the accounts which had the links to spammers; top five link farmers according to the links to spammers according to the Pagerank; the word Pagerank is nothing, but the links that you have to be out which is the in degree and the out degree that is what is PageRank is. Chris Latko, interested in Tech, will follow back and Paul Merriwether,helpingothers,letustalksoon;Aaronlee,socialmediamanager.So,itjust basically shows you what kind of users of the top five link farmers which is creating these links to others and then getting others to follow you back. Internet, social media manager will follow back; these are the kinds of accounts.

Ifyoulookatthepagerankwhichis,higherthelinksofhowyouarelinkedtoothersalso this Barack Obama, Obama 2012 campaign staff; Britney Spears; NPR politics; UK prime minister; JetBlue Airways. So, this is also showing that it is not just the real spammers,butmaliciousintention,theyactuallydoinglinkfarming;evenlegitimate accounts even more so popular accounts are actually part of the link farming ecosystem andtheyincreasetheirfollowers.Thatisakindofrevolutionthattheywantedtoactually get whichis tolook atthe bio,the accounts oflinkfarmers andhave some understanding of what kind of users these are.

(ReferSlideTime:08:57)



Top link farmers are not spammers

- 100,000 link farmers
- 18,826 were suspended
- 4,768 were "not found"
- 76% still active
- 235 are verified
- Manually checked 100 random users (volunteers)
- 86 were real accounts
  - Business, internet marketing, ENT, money and social media

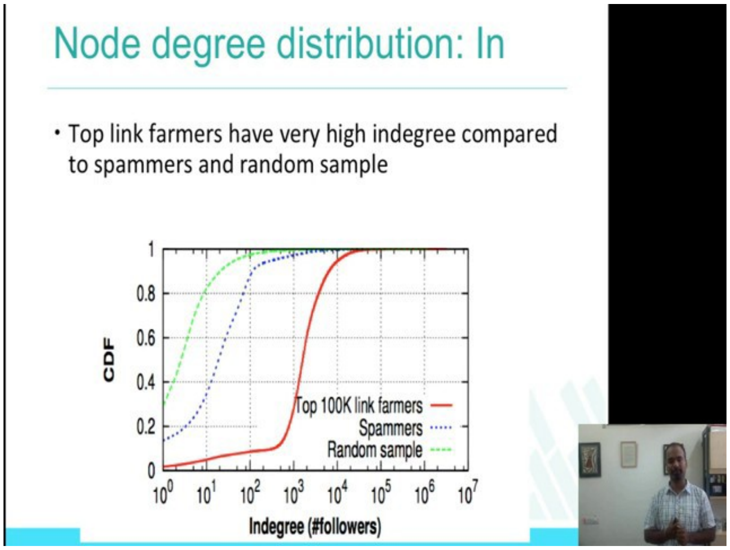
Interestingly top link farmers are not the spammers. They looked at top 100,000 link farmers at the point of analysis of which of course, 18826 were suspended. Twitter did something, figured that all these are actually spammers, these are actually malicious accounts. Therefore, they suspended it. 4768 accounts were not found which is that they probably deactivated the account the account does not exist, whereas, 76 percent of the account of 100,000 link farmers which is, how do they get this hundred thousand link farmers; they do the graph of node rank, they do the graph of what, who were the most popular link farmers and they got this hundred thousand link farmers and of this 76 percent were still alive.

Interestingly of that, 235 were verified accounts, if you remember verified accounts are the accounts which has a blue tick next to the account and these are the accounts which arelegitimatethatistheyknow,theyshowthattheyaretherealpeoplewhichisAmitabh Bachchanhas the real account,verifiedaccount; Obama has a verifiedaccount whichare the real peoplewhotheysaytheyare.

They manually checked 100 random users of 235, but volunteers of course, they got some user volunteers to verify whether to look at these 100 random users and said somethingabouttheusers. Theyfoundthat86wererealaccounts,theyactuallygotmore than one people to look at it, therefore if more than one person says that it is a real account, there's a high probability that its a real account. They actually found that of 86 real users, people were like had the account as business, internet marketing, entrepreneurship, money and social media. These are the topics that the 86 real accounts are talking about. It just gives you the sense of, it also connects very well to the Twitter account bios that we saw in the slide which is top five link farmers.

So, this shows that the top link farmers are not really the ones who are standing in real world, but they could be actually, they are actually real accounts, they are actually verified accounts

(ReferSlideTime:11:27)



Node degree distribution: In

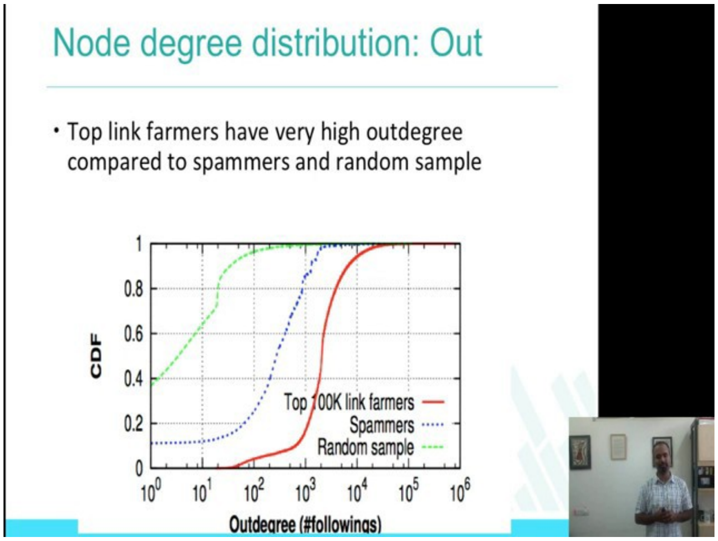
- Top link farmers have very high indegree compared to spammers and random sample

Let us delve little bit more into the node degree distribution which is how the in-degree followers are for top 100,000 link farmers, for spammers and for random sample. Well whatisthegoalhere?Thegoalhereistotryandcomparethehundredthousandlink farmers they found with spammers, which are real spammers; their accounts that are recorded and random sample of users, if they compare these three types of users, the observation can be very helpful to understand what is the property of these 100,000 link farmers..

If you look at this graph, this graph basically shows that top link farmers have very high in-degree compared to spammers and random sample. So, let us go through again the graph in detail; x axis is in-degree, which is again log scale which is 10 to the power of1 to 10 to the power of 7, that is the number of followers, cumulative distributionfrequency CDF is on the y axis. The way you look at this is that the more the value onthe red graph, red line is which is if you look at the 1000 users which has 1000followers which is in degree which is very high. If you look at that, the CDF is about 0.1.

So, top link farmers which is, if you arranged the in-degree in particular order, the top link farmers have very high in-degree compared to spammers which is the graph for the spammers and the random samples is very different and also the CDF is actually verylow within the in degree, which is 10 to the power of 2, 10 to the power of 3. Top link farmers have very high in degree compared to spammers and random sample.

(ReferSlideTime:13:22)

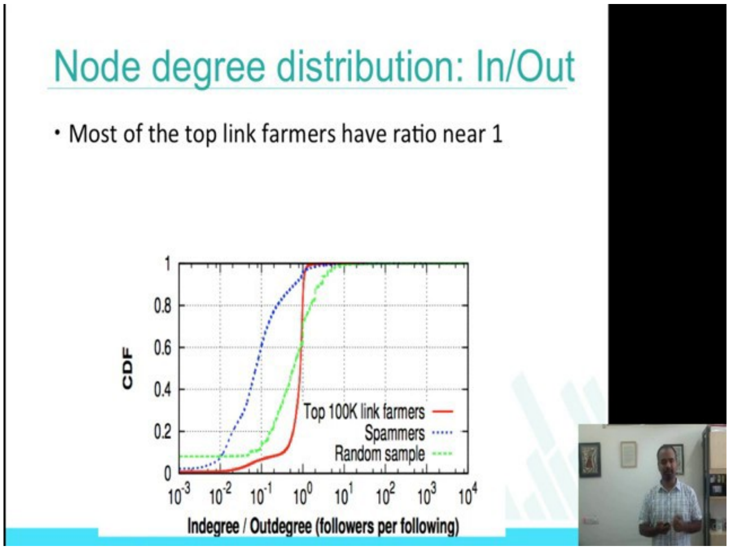


Node degree distribution: Out

- Top link farmers have very high outdegree compared to spammers and random sample

Interestingly, they also found that for the out-degree also, the graph looks very similar which is, the number of followings that I have and the top link farmers have very high out degree compared to the spammers and random sample and slightly very different from, slightly different from the in degree graph, but still again the 100,000 link farmers graph is much higher in terms of out degree compared to the spammers and the random sample.

(ReferSlideTime:13:57)



Node degree distribution: In/Out

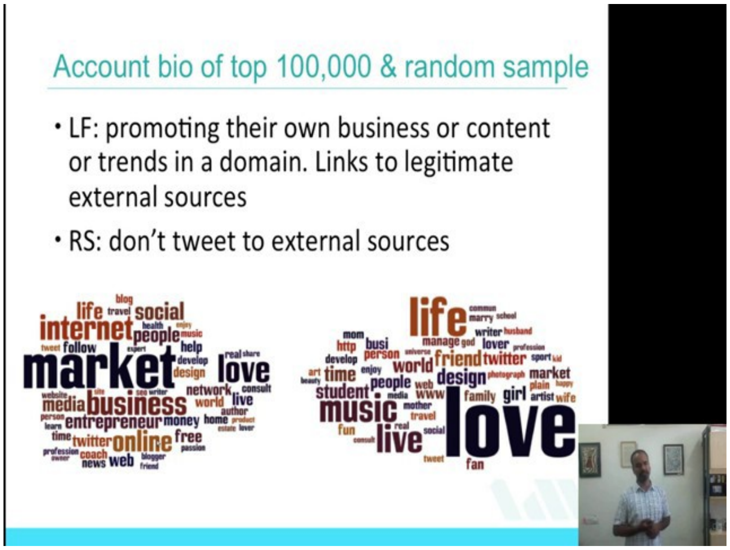
- Most of the top link farmers have ratio near 1

They also did an interesting analysis on finding the ratio between in degree and the out degree, this ratio is actually very useful because if you look at really large followers account like Amitabh Bachchan or Obama, the number of accounts that follows themwill be very high versus the number of followings that they have is actually very low. That is one parameter, one way to look at the legitimate accounts. If you look at again legitimate accounts like mine, probably the number of followers and the number of followings\ are actually very close to each other.

So, this is what they want to find out which is, if you find the ratio of in verses out, how is this;whatisthepatternoftheusers;mostofthetoplinkfarmershavearationearto1, if you look at the graph 10 to the 0 is 1 and interestingly, the value for the top link farmers,whyistheratiobetweenfollowersandthefollowingsisequalto1,whichisthe pattern that I was telling you for real users like mine. So, therefore, link farmers also have this similar behavior. So, the other point to take away from it is given all this, it is going to be hard to find out who is the link farmer that is the kind of intuition that isbuild behind all this analysis.

Again, if you see this graph x axis is so, 10 to the power of minus 1, minus 2, minus 3 is where the in-degree verses out-degree ,when the out is much larger than the n would be the one that are less than 1. So, you can clearly see that the top link farmers which is the red color has the ratio of one, whereas, if you look at spammers and the random sample they are not really one, there is some difference with the examples that I took like Amitabh Bachchan, Obama and myself. So, I hope that makes sense in terms of what distribution is in-degree, out-degree distribution? What is the ratio of in degreeversus out degree? It give you a sense, go through the slides, go through the materials and if there is any confusion or any more clarifications needed, feel free to post it on forum, I will be actually happy to help in understanding these content also.

(ReferSlideTime:16:16)



Account bio of top 100,000 & random sample

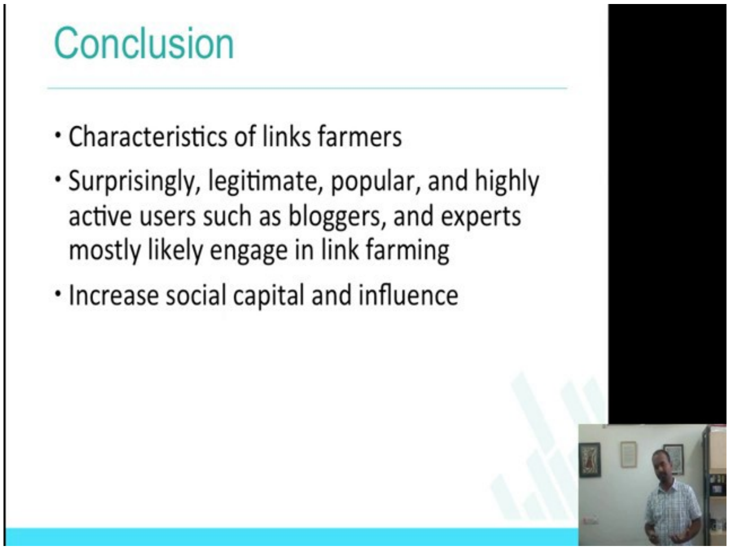
- LF: promoting their own business or content or trends in a domain. Links to legitimate external sources
- RS: don't tweet to external sources

Now,ifyou lookatthebiooftop100,000linkfarmersandrandomsample,justtogetan comparison ofwhat are the people, what are the accounts which are actuallylink farmers andwhataretheaccountswhicharerandomsampletalkingabout?Theleftoneis actually the link farmers, the right one is actually random sample. You can clearly see here that the left one is talking more about market, online, internet, social, love which probably is in random sample also - life, music, live, love, right.

So, theone some conclusion thattheydo fromthisanalysis is thaton theleftyou seeLF, you can see that promoting their own business or content or trends in a domain, links to legitimate external sources. Of course, they are basically talking about some business, talking about some links that are outside twitter, outside the network that they are promoting this content. Thatis isright, donot tweetto external sources which is,thereis not a lot of links to other sources.

Again this will be a pattern that this research formed, but the pattern if you like to study this today, it may be very different also. I am kind of looking at some of these classical worlds which looks at some of the question that we should be asking in privacy and security in online social media topic.

(ReferSlideTime:17:43)



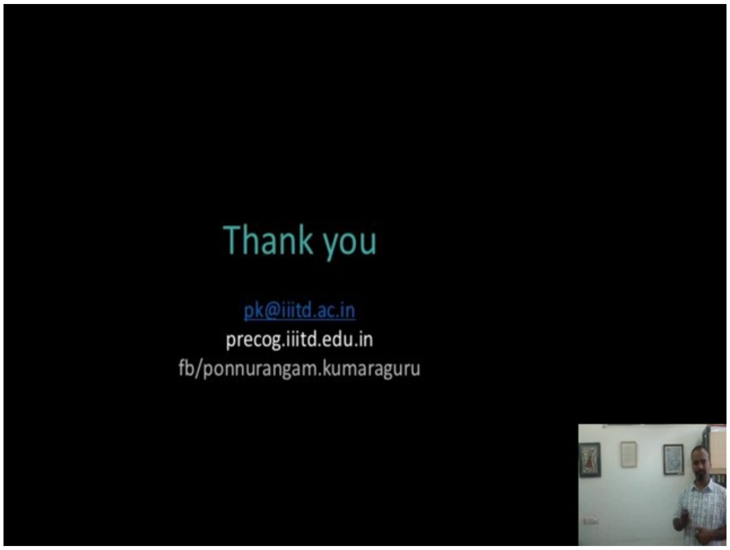
Conclusion

- Characteristics of links farmers
- Surprisingly, legitimate, popular, and highly active users such as bloggers, and experts mostly likely engage in link farming
- Increase social capital and influence

So, the final conclusion for this part of the work is, this part of the course / lecture is characteristics of link farmers we found the in degree verses out degree ratio is actually prettyhigh.Thein-degreeisveryhighcomparedtospammersandrandomsample.The out-degree is also very high; the probability of spam followers is also very high in termsof requests sent to them or if Ifollow you, there is a high probabilitythat you will follow me if you are a spam follower. Surprisingly, legitimate popular and highly active users such as bloggers and experts, most likely engage in link farming, these are accounts like Britney Spears, Obama all of these accounts actually have link farming behavior.

So, the problem is that this increases the social capital and the influence because if the link farming engage, if the concept, I will go back to the first slide again on this topic. If there is high probability of, if there are chances that you let your social reputation, the links between the users are higher, their social recognition are I mean today social reputation, influence is all measured by number of followers you have and the kind propagation of of the content that you have and therefore, link farming can be pretty effective in terms of increasing your social capital and influence.

(ReferSlideTime:19:10)



Thank you

pk@iiitd.ac.in

precog.iiitd.edu.in

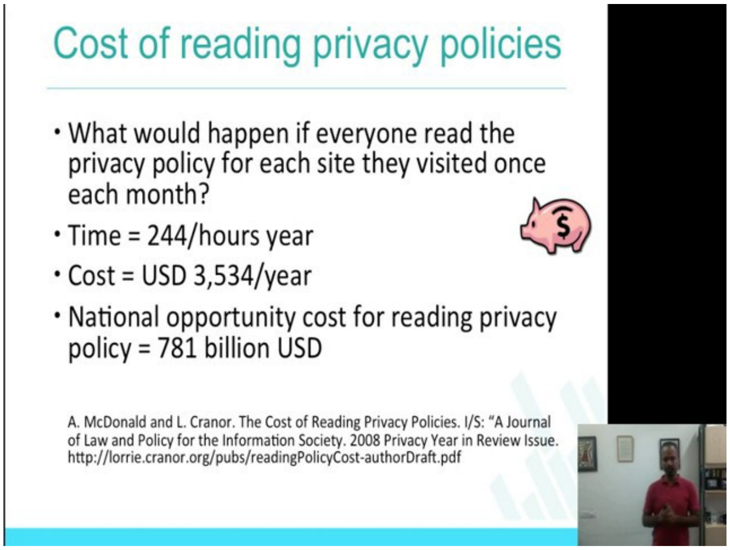
fb/ponnurangam.kumaraguru

WiththatIwillactuallystopthisparticularpartofthelecture,whichisweek7.1.

Nudges

Welcome back to the course Privacy and Security in Online Social Media, week number 7 - second section of the week number seven. So I hope you got a chance to look at the content that we made available for the week 7.1.

(ReferSideTime:00:27)



Cost of reading privacy policies

- What would happen if everyone read the privacy policy for each site they visited once each month?
- Time = 244/hours year
- Cost = USD 3,534/year
- National opportunity cost for reading privacy policy = 781 billion USD

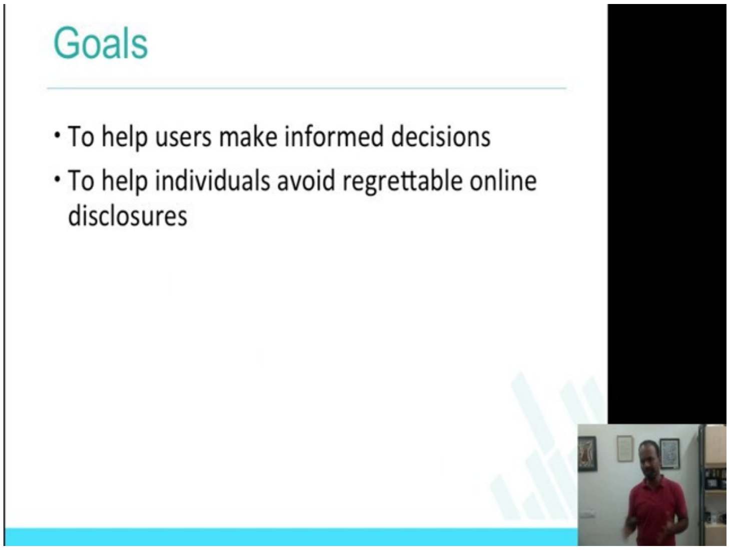
A. McDonald and L. Cranor. The Cost of Reading Privacy Policies. I/S: “A Journal of Law and Policy for the Information Society. 2008 Privacy Year in Review Issue. http://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf

I do not know,how many of you actually read privacy policies. Researchers have shown that people do not read privacy policies. Just let us do some exercise now. So, some of you may have actually done online transactions in the last one week or almost all of you would have done some kind of online transactions, logging to your Facebook, logging to your Twitter, transactions of buying things online or between banks. How many of you haveeverreadanyprivacypolicyofthewebsitesthatyouhaveinteractedwithinthelast week or a month, must you really low. And it is not just what I am saying, people have actually studied this exhaustively.

So, here is one piece of work which says, if we were to make people read privacy policies,whatitwoulditcost.Thestudy wasdoneintheUS(ReferTime:01:19) anditis actually part of PhD thesis work, where the question was, what would happen ifeveryone read the privacy policy for each website they visited once each month. Time taken would be about 244 hours per year, which is basically, that questions more on the lines of economics of what would it cost. It is 244 hours,convert it into money,

, and in total number, if we just look at the US population, national opportunity cost for reading privacypolicywould be 781 billion US dollars, which is, if I get everycitizen of the US to read the privacy policy for least one month for the websites they have seen, it would cost in some 781 billion dollars. That is a lot of money, and that is a lot of opportunity that is being lost because they are spending time on reading this privacy policy. While they could make the pair, which they would (Refer Time: 02:14) use to make the decision.

(ReferSideTime:02:18)



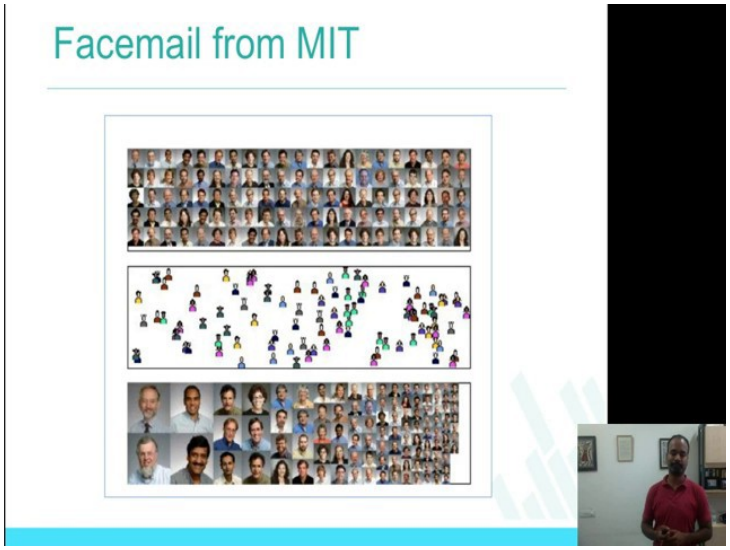
Goals

- To help users make informed decisions
- To help individuals avoid regrettable online disclosures

So, keeping that context in mind, which is reading the privacy policies, researchers started asking some questions, and broadly also, there is this whole area whereresearches are working on, and technologies is being built to help users make informed decisions,whichishowcanIhelpuserstomakeinformeddecisionswiththe information that is presented to them, with the information that they can actually use from these services.

Thespecificgoalistohelpindividuals avoidregrettableonlinedisclosures,whichis,can we actually build technology, can we actually build something for the users to use, so thattheycanactuallybehappyaboutthecontentthatthattheyarepostingoravoidbeing regrettable for the information that they are disclosing online.

(ReferSideTime:03:10)



Facemail from MIT

Sohere is onetechnologythat was builtat (ReferTime:03:14) MIT,meaning evennow when we are using this discussion forum for this class, when you type in the discussion forum, add the course name, you will actually find out in this plot, when if you use this technology,you do not really know who are the people who are getting this email, right, (Refer Time: 03:36), because the groups are set up differently, people have been signed up into the groups, you really don't get to know who you are interacting with (Refer Time: 03:44). For example, in a mailing list in your company and mailing list in your college or mailing list that you maintain for yourself with your friends, so all of this you do not really get to see who is getting these emails (Refer Time: 03:57).

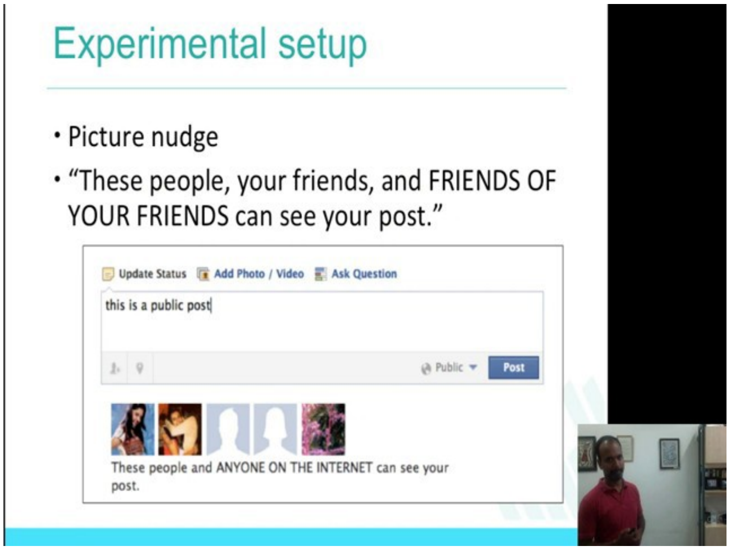
So,whattheseresearchersatMITdidwas,theyareactuallysaid,okay,wheneverthe email is going to go, because this would, this could, be a problem also, right,because you do not know who is getting the email. So, it could be a problem the information that you are sharing could actually go to people whom you do not want them to see this content.

So what they said was, okay, if there is a email address the state called (Refer Time: 04:19) psosmnptel2016 at abc dot com, they would actually show you the profilepictures of the people who are going to get this email, so that is a way by which to show that how many people are actually going to get the email, that is the information on the slide also, their profile pictures.And it would also show you who are the people who are gettingthe,becauseoftheprofileyou gettoknowmoreaboutthepeoplewhoaregetting to see this email.

So, this just helps users to make a decision on sending this email, because let us take if you were to send an email, and if you wrongly typed the mailing list address, it could actually end up going to a wrong mailing list. You wanted to send it only to 10 people, whereas the email list that you send is actually going to 100 people, so this can actually help avoid (Refer Time: 05:10).

In another version of it that they built also where they were actually showing you thebottom one, when they were actually showing you the profile picture of the person who you have been interacted more slightly bigger than others ones. This is also helping you to make a judgement on who is getting the email, who were the people who are getting these e mails and who you have interacted with more frequently than others. So this is justananotherexampleofhelpinguserstomakeinformeddecision,whetheryouwantto send the email or not.

(ReferSideTime:05:48)



Experimental setup

- Picture nudge
- "These people, your friends, and FRIENDS OF YOUR FRIENDS can see your post."

this is a public post

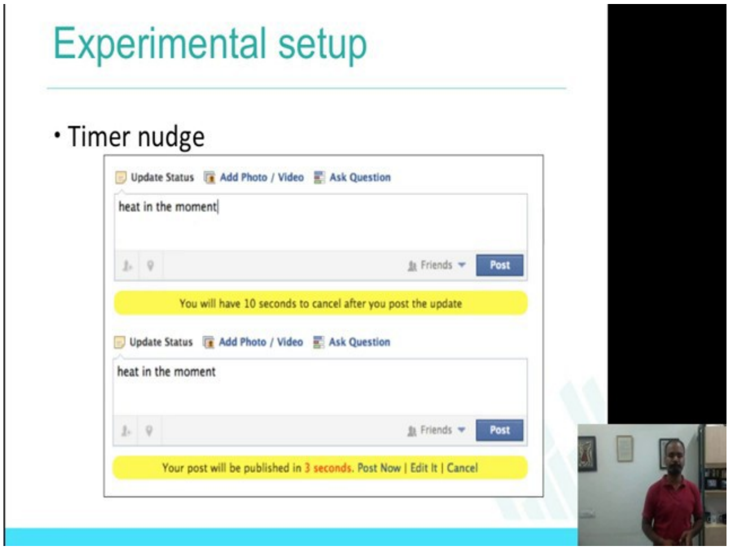
These people and ANYONE ON THE INTERNET can see your post.

So, keeping that in mind, which is to help users make this decision - informed decision - here is a piece of work that was done to look at the context of just Facebook. So in this experiments that they did, what they did was they built Chrome browser extension,which would actually show you, you go to facebook.com, when you are going to do a post it is actually going to nudge you with some information. They have different set of sub nudges, I will walk through what they did.

And we will also see how effective it was, which was effective, which was not effective. Because today, you could actually, meaning, I am sure some of you have experienced that you did some post and which you did not want somebody to, somebody in your friends networktosee,andtheygottoseethepost,andtherehavebeenincidences inthe past also. If you remembered even in the week 1 or week 2, I showed you someexamples about MI6 chiefs (Refer Time: 06:44) while posting some content that wentpublic, where it was not intended to go public.

So, these kind of incidents have actually made people to think about building technologies that will help them, help users make that decisions. So here is the first idea called picture nudge. What it does is if you wanted to do a post, when you are doing a post it is actually going to stop you and saythat these people, which is the profile picture connected to the face mail, it is the same thing as in the email mailing list. These people yourfriendsandfriendsofyourfriendscanseeyourpost,itisgoingtostopyouandtell you this information. Again in the bottom screenshot which showing you these people and anyone on the internet can see your post. So, this actually helps you to make a decision on whether you want to actually do this post or not.

(ReferSideTime:07:38)



Experimental setup

- Timer nudge

heat in the moment

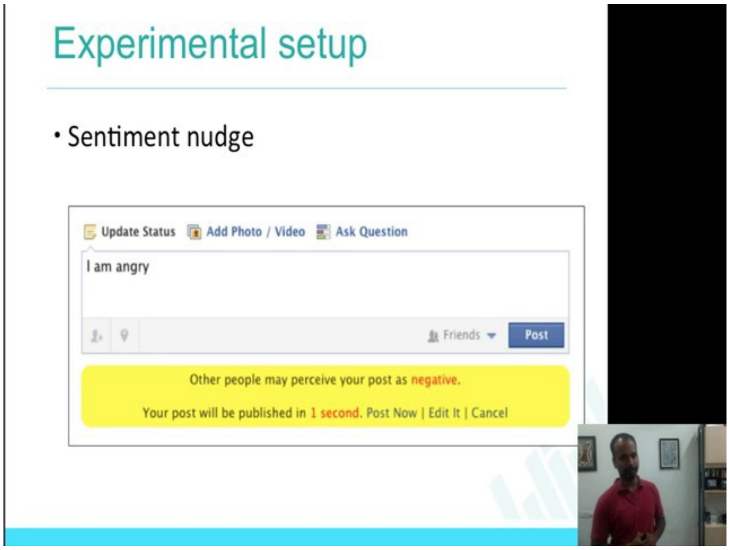
You will have 10 seconds to cancel after you post the update

heat in the moment

Your post will be published in 3 seconds. Post Now | Edit It | Cancel

Here is a second experimental set up which is timer nudge. Here it is not showing youthe profile pictures. But it showing you that you have ten seconds to cancel your post. This information just lets you to say that essentially, you are doing this post, you really want to do this post, wait for ten seconds, if you want to change your mind, do it now, and then do the update. That is the timer update.

(ReferSideTime:08:10)


Experimental setup

- Sentiment nudge

I am angry

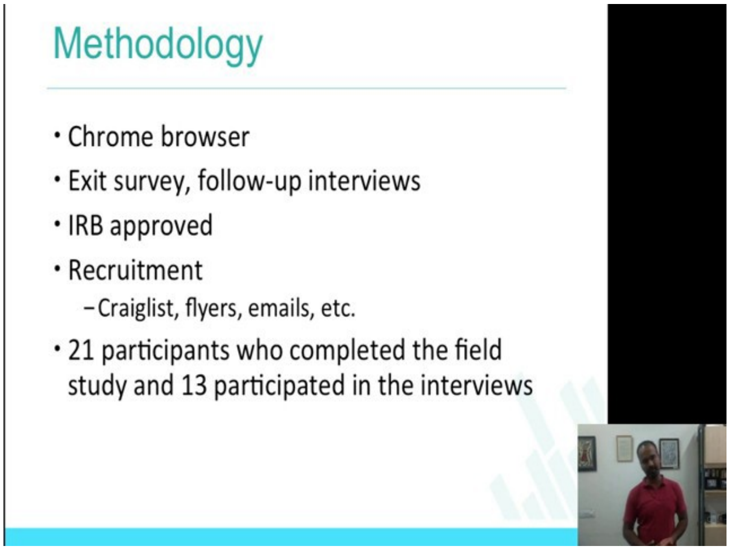
Other people may perceive your post as negative.

Your post will be published in 1 second. Post Now | Edit It | Cancel

Here is next one which is sentiment nudge. Again there have been incidences all around the world where people have actually posted the things on Facebook and it is actually backfired on them. Backfired in terms of actually very negative effects also for the content that they have posted.

So, to avoid such things here is a setup where it shows you and in this case I am angry,so it shows you that it is negative; other people can perceive your post just negative. So, negativesentimentisattachedtothepost,sopleasebecareful,doyoureallywantpostit, and things like that, so that is the sentiment nudge. Let us look at all the three - picture nudge with pictures, timer nudge with actually ten seconds time, and sentiment nudge which gives you the sentiment of the post that you are making.

(ReferSideTime:09:04)



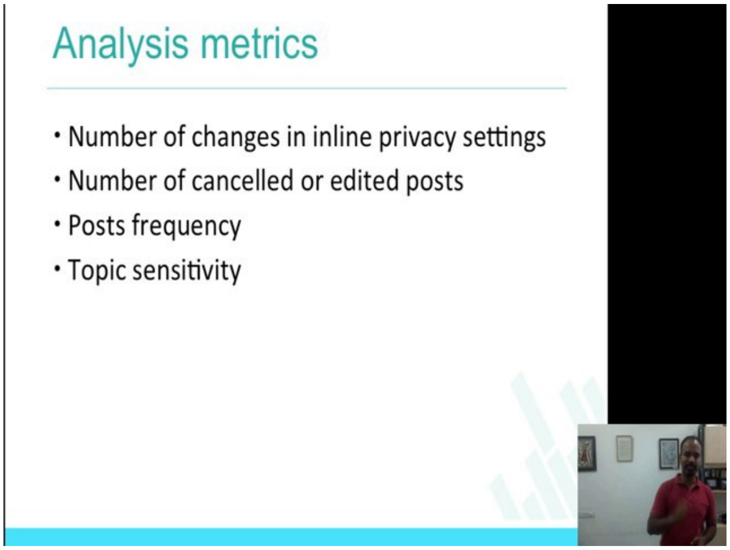
Methodology

- Chrome browser
- Exit survey, follow-up interviews
- IRB approved
- Recruitment
  - –Craiglist, flyers, emails, etc.
- 21 participants who completed the field study and 13 participated in the interviews

So methodology of the study that they did was Chrome browser as they said, they did exitsurveys, which iswhenpeoplecompleted thestudy,theyasked themsomequestions and also asked them questions in terms of both quantitative and also interviews with the participants. So, IRB is institutional review board, which is basically to say that if youareinteracting orcollecting datafromusers,itwillbehuman subjects,you reallywantto make sure that things do not go wrong when they are actually doing the study, and they should not feel offended and things like that. And that is why IRB approval is necessary when you are interacting with human subjects.

And of course, users were recruited in multiple ways. (Refer Time: 09:52) Of course,puttingflyersallaroundtheplacesendingoutemails,Craigslist,allof that.Theygot21 participants who completed the field study, because it is the Chrome browser plug-in, they could actually use it at home or wherever, and 13 people participated in the interviews, which is the exit survey.

(ReferSideTime:10:12)

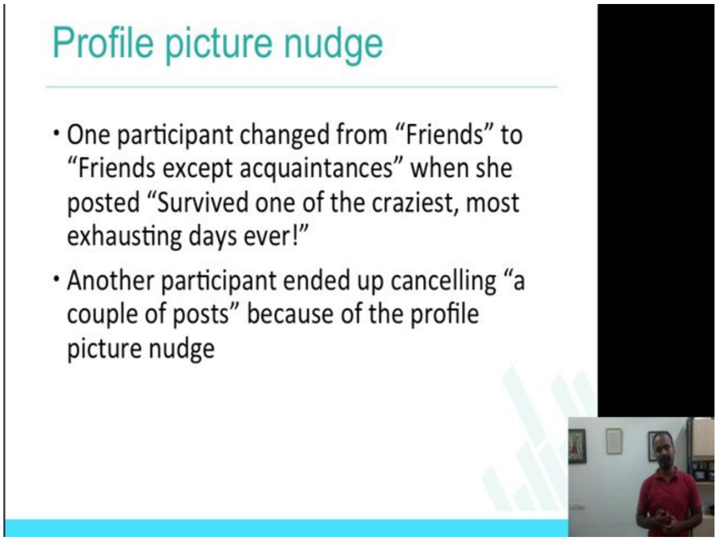


Analysis metrics

- Number of changes in inline privacy settings
- Number of cancelled or edited posts
- Posts frequency
- Topic sensitivity

So, in the metrics that they are actually used to analyze what is going on within this context of providing nudges of these things. Number of changes in inline privacy settings, number of canceled or edited posts, post frequency and topic sensitivity, essentially they were trying to understand by giving these nudges are people changingthe behavior. And if they change the behavior what are they changing, so that is the conceptthattheywereactuallytryingto study,that is themetrics that theyweretrying to study.

(ReferSideTime:10:43)

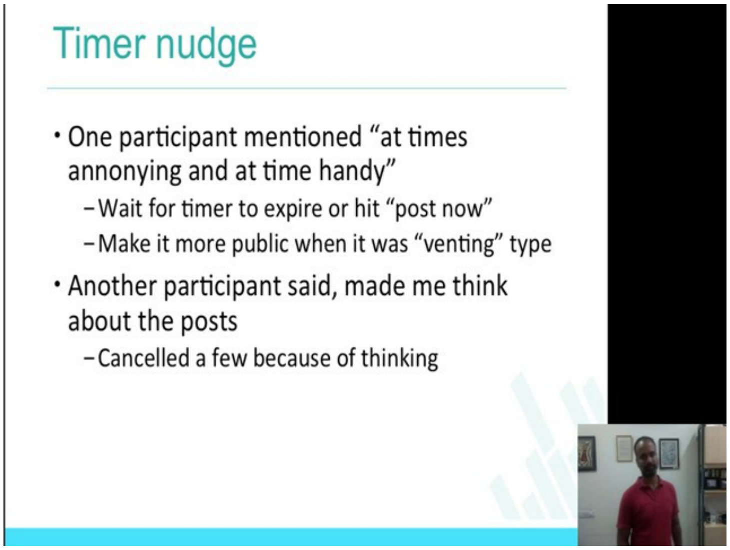


Profile picture nudge

- One participant changed from “Friends” to “Friends except acquaintances” when she posted “Survived one of the craziest, most exhausting days ever!”
- Another participant ended up cancelling “a couple of posts” because of the profile picture nudge

Profile picture nudge. So the first one, I think I have one slide for a nudge to tell you what happened in this study. One participant changed from friends to which is probably all of them, friends except acquaintances, when she posted survived one of the craziest, most exhausting days ever. So essentially people are changing the group in which they are sharing the content depending on theinformation that thenudge is actuallyproviding them. Another participant ended up canceling a couple of posts, because of the profile picture nudge basically saying that oh it is going to actually lot more people than what I thought, what I think, so let me not do (Refer Time: 11:22) the post.

(ReferSideTime:11:24)

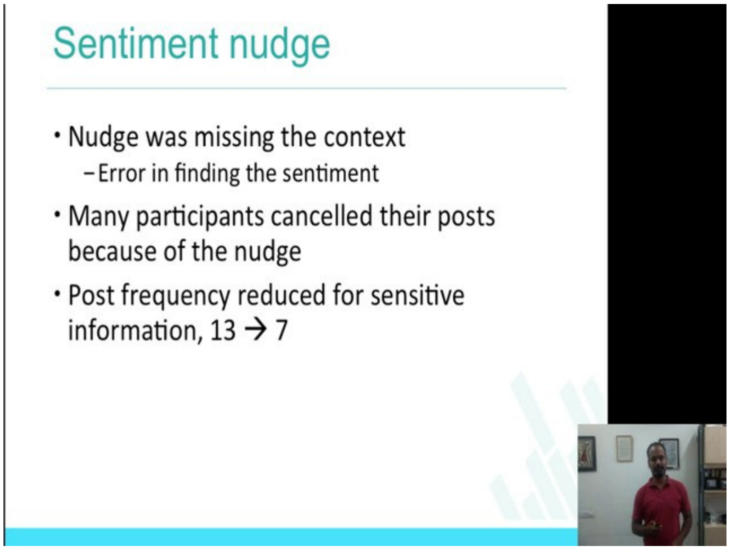


Timer nudge

- One participant mentioned "at times annonying and at time handy"
  - Wait for timer to expire or hit "post now"
  - Make it more public when it was "venting" type
- Another participant said, made me think about the posts
  - Cancelled a few because of thinking

Timer nudge, one participant actually said at times annoying, and at times handy. Because I am pretty sure, right, because if you are doing, let's take, 10 posts a day or, like, 5 post a day also, it is going to stop you for every post ten seconds and then only togo. Waiting for a timer to expire or hit the 'post now'. So it is essentially a feature that could be provided. Make it more public, when it was venting type, right, make it more, so, it can be more suggestive. Another participant said, made be think about the posts, which is the same, which is the behavior that is actually important for these nudges to actually make within the users(Refer Time: 12:05). Changing the user behavior, it is abig need while using this technology.

(ReferSideTime:12:15)



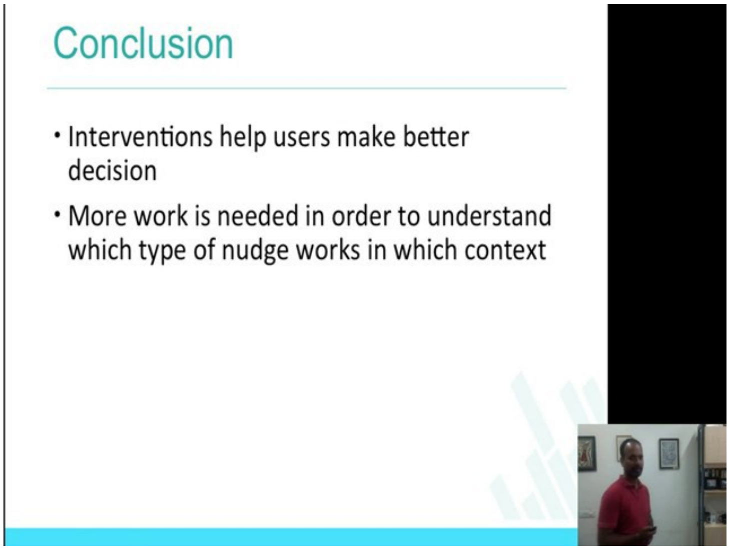
Sentiment nudge

- Nudge was missing the context
  - Error in finding the sentiment
- Many participants cancelled their posts because of the nudge
- Post frequency reduced for sensitive information, 13 → 7

Sentiment nudge, nudge was missing the context, of course, right, because I think the whole context of, in what context I am doing the post is actually very important to find out sentiment. And even human beings, it is actually hard to get the sentiments, so, the nudge was, the tool was actually making, browser plug-in was actually making errors while the calculating or finding out what these sentiments are.

Many participants canceled the posts, because, I think it is because if the posts are negative, and many people are going to be actually offended by the post within your network or in public, it is actually going to be bad for you. So, people are actually canceled the post. And the post frequency also reduces; they were actually doing 13 posts, it went down to 7.

(ReferSideTime:13:07)



Conclusion

- Interventions help users make better decision
- More work is needed in order to understand which type of nudge works in which context

So, essentially all of this is helpful in making the, helping the users to make better decisions. And particularly when it comes to posting something that others are going to be offended, posting something that people are going to, people whomyou do not intend to actually share the information, all of this is actually helpful, these technologies are helpful, for the users to make a better decision.

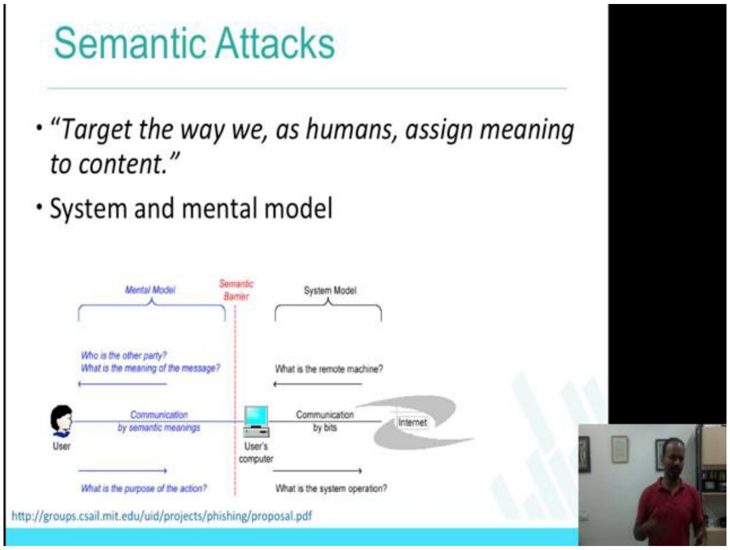
And of course, more work is needed to understand which type of nudges work in which context, because the contexts could be very different - I am just doing you a quick update, saying, I am, for example, I was actually doing a lot of updates, yesterday,about the convocation at IIIT-Delhi with the hash tag IIIT Convo 5. And if I wait, if it was (Refer Time: 13:55) going to stop me, and probably I did like thirty, forty posts on Twitter. But every time if it is going to stop me for ten seconds, that is not going to be good, so, and also, in terms of the, in terms of the sentiment that it is showing,everything has to be done slightly better.

So, I think more work is needed, I am sure there are people in the class who areinterested in taking some of this, it may be interesting projects to work on. With that I will stop the 7.2 part of the week. And I will continue on something which is in the context of phishing; and phishing in the context of social media in the next part of this week also.

Semanticattacks:Spearphishing

Welcomebacktoweek7andthisisthethirdpartoftheweek7.Inthisclass,inthis section what will see is, we will see about PhishingAttacks in online social networks.

(ReferSlideTime:00:18)



Semantic Attacks

- *"Target the way we, as humans, assign meaning to content."*
- System and mental model

Mental Model
Semantic Barrier
System Model
Who is the other party?
What is the meaning of the message?
What is the remote machine?
Communication by semantic meanings
Communication by bits
Internet
User
User's computer
What is the purpose of the action?
What is the system operation?

http://groups.csail.mit.edu/uid/projects/phishing/proposal.pdf

So, this is a slide from an MIT PhD thesis which actually looked at what a semanticattack is? Semantic attacks are attacks that happens where humans are targeted. So, for example, Bruce Schneierwho is supposed to be a expert in security classified the different types of attacks that could happen as physical, syntactic and semantic, but physical attacks are the were happening like 15-20 years before where the attackers would actually get physical access to the machine.

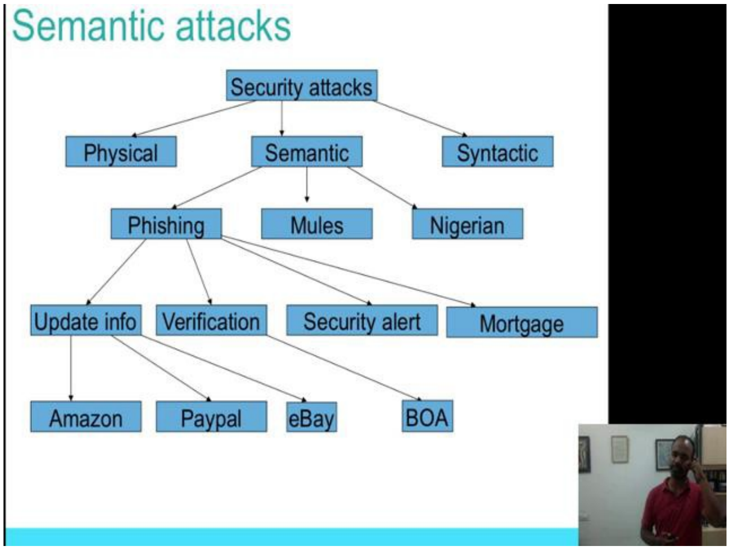
Whereas in syntactic attacks are the attacks that were happening around the programs, around the systems that are built, which is more like the denial of service attacks, buffer overflowattacksandattackslikethat,butthesemanticattacksareattackswhichtarget the way we as humans assign meaning to the content which is that what do we because the specific attack that we will be talking about is phishing.

For example, ifyou get anemail frompk at iiitd dot ac dot in now,talking about NPTEL course and whichhas a link saying please give your user name and a passwordto seethe content heremost likelythat you'regonnaactuallyclickthe linkandgivethe information which may be a phishing link also. So, that is started the way you actually think you are seeing an e mail that is coming from legitimately pk at iiitd, and the system thinks that you are actually going to this free website forum psosm on NPTELdot come slash logindot html, but actually it is a phishing website. it is targeted phishing website.

So, system and mental model, what this is in this PhD thesis they actually nicelyput itthat semantic barrier, which is the difference between what system thinks we are doing and what you think that system is doing will actually be called semantic barrier in the larger the barrier is it is because actually difficult to not fall for such types of products. So, ifyou lookat the mentalmodelit says, who is the other partywhat is the meaning of the message. So, the example that I said also what is the meaning of the email we got what is the meaning of the who is sending the message and information is all mental model ofthe user, but a system modelwhois the remote machine where booked website I am going to access and information like that.

So, user model or the mental model which what users think that is happening, system modelwhich is what, the systemthinks that the user are doing, the barrier between them the difference between them is actually called semantic barrier. and the larger the barrier is, its actually hard to actually fix the problem. So, this is what we will use this is what we will actually talk about mostly in the section called phishing.

(ReferSlideTime:03:23)



Semantic attacks

Security attacks

Physical

Semantic

Syntactic

Phishing

Mules

Nigerian

Update info

Verification

Security alert

Mortgage

Amazon

Paypal

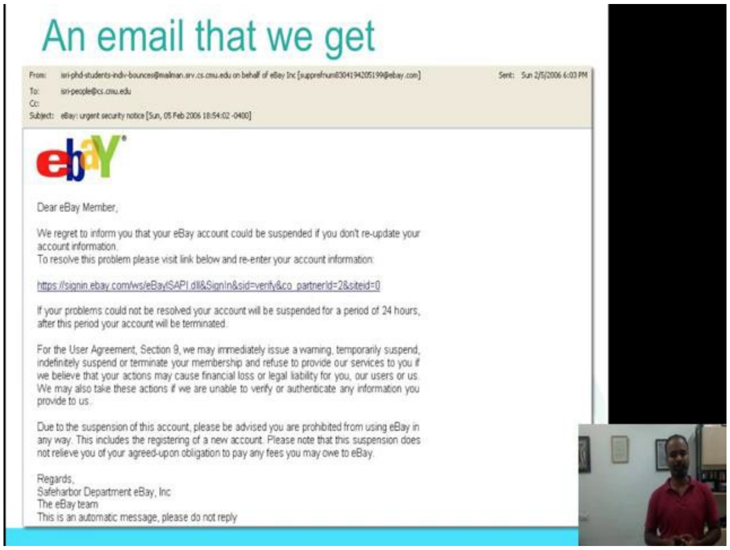
eBay

BOA

Here is the broad category of semantic attacks: Security attacks - physical, semantic and syntactic which is what Bruce Schneier did and if you look at Semantic attacks, you can actually go through multipe categories Phishing, Mules, Nigerian, 4 1 scams and attacks like that, and in phishing also there are multiple categories – update your information, banks and in your ICICI banks sending you a message saying that, please update your information within next 24 hours or your account would be closed. Verification, saying that, we want to verify whether it is really you, please click and verify. Security alert, Microsoft is updating the latest version of MacOS, thereis an update.

Here is the link, please go and update. Mortgage information, meaning your mortgage, the due is coming closer, please click this link and do something. All of these kinds of categories ofattacksare called phishing attacks and almost allcompanies todayprobably are undergoing, are part of, or being victims of this attack of phishing. Even academicinstitutes probably are victims of phishing attacks.

(ReferSlideTime:04:36)



An email that we get

From: isri-phd-students-indv-bounces@mailman.srv.cs.cmu.edu on behalf of eBay Inc [supprefnum83041942051 99@ebay.com] Sent: Sun 2/5/2006 6:03 PM
To: isri-people@cs.cmu.edu
Cc:
Subject: eBay: urgent security notice [Sun, 05 Feb 2006 18:54:02 -0400]

Dear eBay Member,

We regret to inform you that your eBay account could be suspended if you don't re-update your account information.
To resolve this problem please visit link below and re-enter your account information:

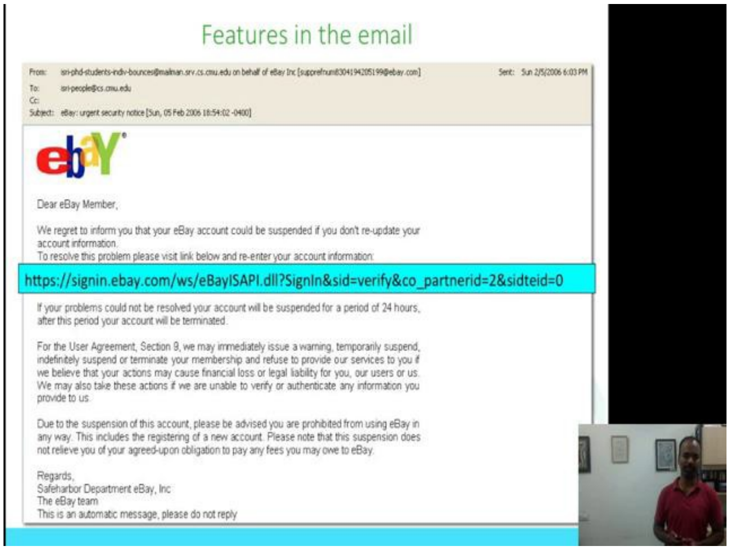
https://signin.ebay.com/ws/eBayISAPI.dll&SignIn&sid=verify&co_partnerId=2&siteid=0

If your problems could not be resolved your account will be suspended for a period of 24 hours, after this period your account will be terminated.

For the User Agreement, Section 9, we may immediately issue a warning, temporarily suspend, indefinitely suspend or terminate your membership and refuse to provide our services to you if we believe that your actions may cause financial loss or legal liability for you, our users or us. We may also take these actions if we are unable to verify or authenticate any information you provide to us.

Due to the suspension of this account, please be advised you are prohibited from using eBay in any way. This includes the registering of a new account. Please note that this suspension does not relieve you of your agreed-upon obligation to pay any fees you may owe to eBay.

Regards,
Safeharbor Department eBay, Inc
The eBay team
This is an automatic message, please do not reply



Features in the email

From: isri-phd-students-indv-bounces@mailman.srv.cs.cmu.edu on behalf of eBay Inc [supprefnum83041942051 99@ebay.com] Sent: Sun 2/5/2006 6:03 PM
To: isri-people@cs.cmu.edu
Cc:
Subject: eBay: urgent security notice [Sun, 05 Feb 2006 18:54:02 -0400]

Dear eBay Member,

We regret to inform you that your eBay account could be suspended if you don't re-update your account information.
To resolve this problem please visit link below and re-enter your account information:

https://signin.ebay.com/ws/eBayISAPI.dll?SignIn&sid=verify&co_partnerid=2&sidteid=0

If your problems could not be resolved your account will be suspended for a period of 24 hours, after this period your account will be terminated.

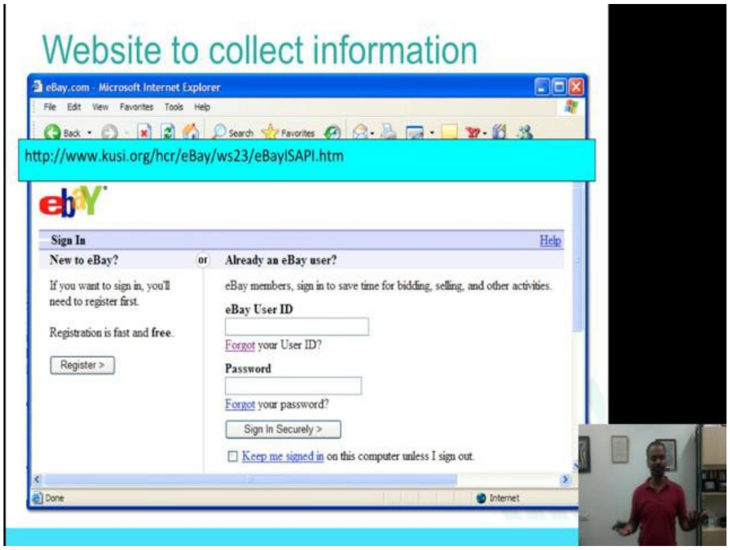
For the User Agreement, Section 9, we may immediately issue a warning, temporarily suspend, indefinitely suspend or terminate your membership and refuse to provide our services to you if we believe that your actions may cause financial loss or legal liability for you, our users or us. We may also take these actions if we are unable to verify or authenticate any information you provide to us.

Due to the suspension of this account, please be advised you are prohibited from using eBay in any way. This includes the registering of a new account. Please note that this suspension does not relieve you of your agreed-upon obligation to pay any fees you may owe to eBay.

Regards,
Safeharbor Department eBay, Inc
The eBay team
This is an automatic message, please do not reply

Here is a simple example, which is an email that the 3 parts of the email whichisactually makes the legitimate email and the difference between the legitimate email and the phishing email, which is the subject line, subject line and urgency in the message on there are there is the line. These are the three things that happen that is a part of the phishing email which at least one wants to keep attention on. Subject eBay urgent notification from billing department. We regret to inform you that your eBay account could be suspended if you do not update your account information and then there is alinkthereandthenwhenweclickonthelinkittakes youtoawebsitecalledkusidotorg.

(ReferSlideTime:05:15)



Website to collect information

http://www.kusi.org/hcr/eBay/ws23/eBayISAPI.htm

eBay.com - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Done
Internet

Sign In Help

New to eBay? or Already an eBay user?

If you want to sign in, you'll need to register first.

Registration is fast and free.

Register >

eBay members, sign in to save time for bidding, selling, and other activities.

eBay User ID

Forgot your User ID?

Password

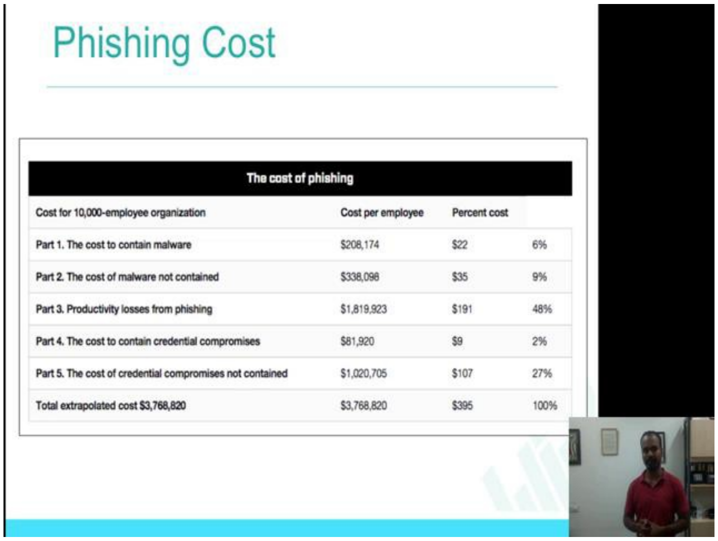
Forgot your password?

Sign In Securely >

□ Keep me signed in on this computer unless I sign out.

Which supposedly should be takingyou toeBay sign in page, so thatis the sharing that is a very classical phishing attack when there are multiple ways ofa changing these kind of phishing attacks.

(ReferSlideTime:05:34)



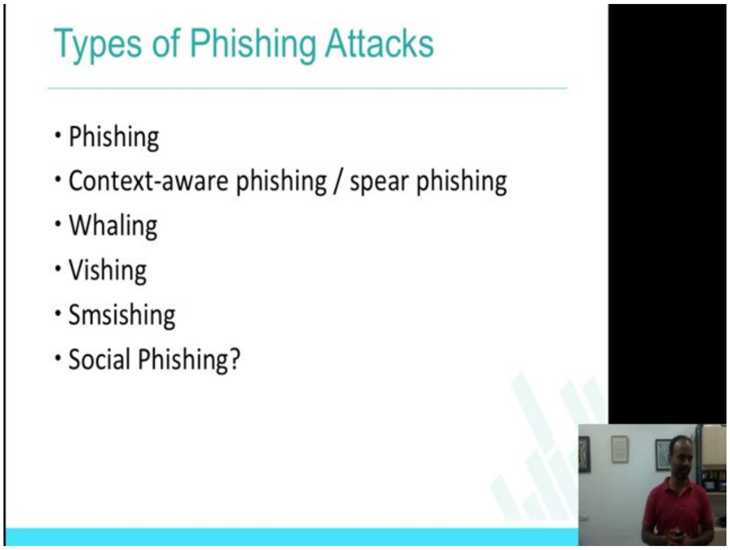
Phishing Cost

The cost of phishing

Cost for 10,000-employee organization	Cost per employee	Percent cost	
Part 1. The cost to contain malware	$208,174	$22	6%
Part 2. The cost of malware not contained	$338,098	$35	9%
Part 3. Productivity losses from phishing	$1,819,923	$191	48%
Part 4. The cost to contain credential compromises	$81,920	$9	2%
Part 5. The cost of credential compromises not contained	$1,020,705	$107	27%
Total extrapolated cost $3,768,820	$3,768,820	$395	100%

Here is some cost again, economics about phishing, some costs that is relevant to the topic ofphishing. Costsofhundred thousand employees organization, which is the, ifthe phishing attack happens what would be the cost to contain the malware, the cost to contain a malware, the cost of malware not contained. So, if you look at the cost its actually pretty high in terms of actually even the phishing attack.Total extrapolated cost is 3 million 76; 3 million plus dollars, all right. So, it needs a lot of money that is spent every year, FTC and many other organizations in US actually try to course aboutphishing and there is an organization called anti-phishing working group which actually specifically works on the problem of phishing and how to actually reduce it.

(ReferSlideTime:06:28)



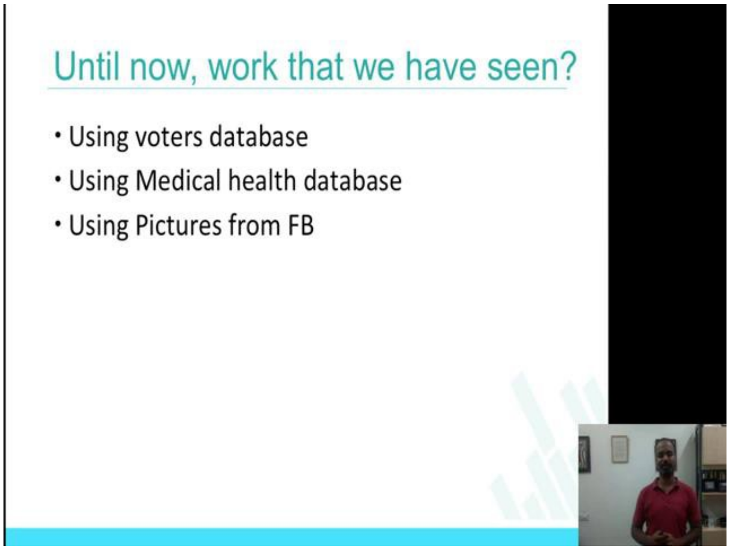
Types of Phishing Attacks

- Phishing
- Context-aware phishing / spear phishing
- Whaling
- Vishing
- Smsishing
- Social Phishing?

So, here are some kinds of Phishing Attacks I think we probably briefly mention this in thepast also.So,I'llgooverquickly,phishingwhichisaclassicalonethat Ishowedyou, Context-aware phishing the email that I talked about sending in to the students takingthis course, Whaling is an attack which is sent to the chief executive officers of the company, Vishing is over the phone, Smsishing is over the SMS. So, what is social phishing?That is what the topic that we have been discussing for the rest ofthis week.

SocialPhishing;doesanybodyknowwhatsocialphishingis?

(ReferSlideTime:07:10)



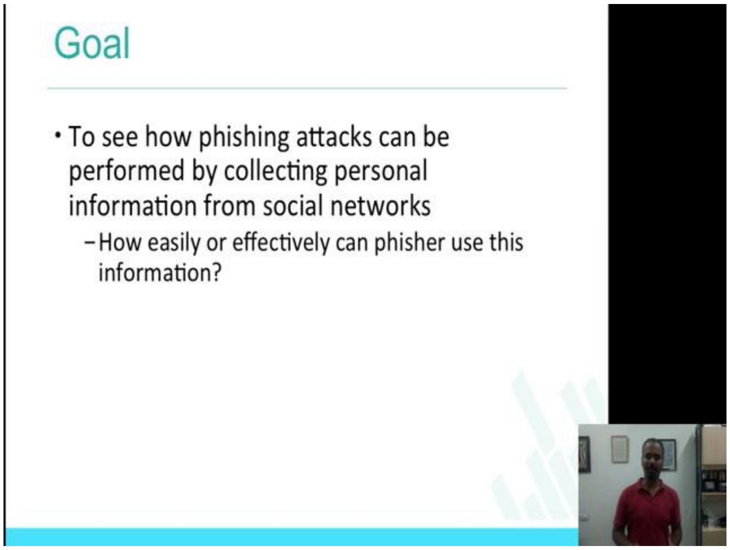
Until now, work that we have seen?

- Using voters database
- Using Medical health database
- Using Pictures from FB

Social phishing is nothing but looking at the information from the social context then using that to actually phish, it is not about finding whether you are taking a course that could be many other information that I defined on from a Facebook page, from the facebook account, things thatyou've done and things like that. So, the topics thatwehave seen until now are using older data we saw Latanya Sweeney's work using medical health data, again Latanya Sweeney's work, using pictures from FB voter data.

We also saw the work that was down in collecting pictures from the university campus collecting this information and making some judgments about the user. Finding the people who they see in the campus whether they will get the right profile from the Facebook. So, those are the topics that we saw, but we never saw about what social phishing is.

(ReferSlideTime:08:02)

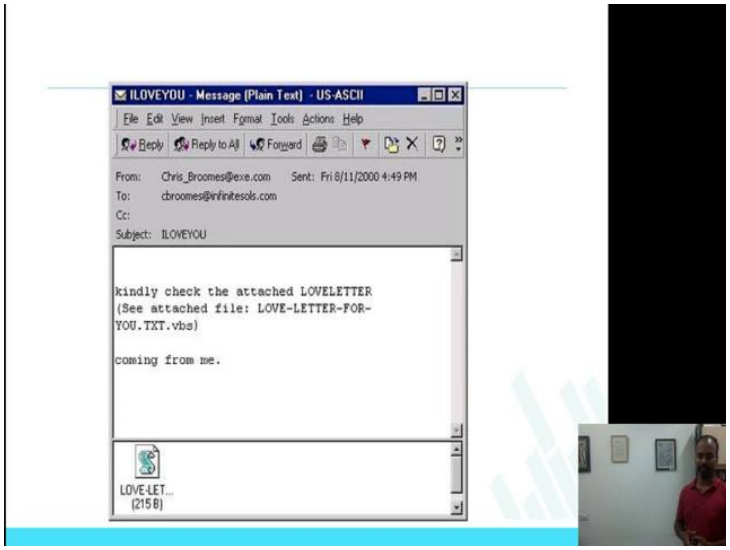


Goal

- To see how phishing attacks can be performed by collecting personal information from social networks
  - How easily or effectively can phisher use this information?

So, here is a goalthe goal is to see how phishing attacks can be performed by collecting personal information from social networks right. So, it is not about sending into the CEO’s, it is not about sending to the students of this class. It is about actually can I collect some informationabout you fromyour social network behavior and use it against you, how easily or effectively can phisher use this information. Again there is a very classicalworkthat was done some years back. So,it'llbe nice to actuallyknow howthey did it some years back. I am pretty sure these studies can be done again to seehow itgoes and this study was done in the US.

(ReferSlideTime:08:41)


ILOVEYOU - Message (Plain Text) - US-ASCII
File Edit View Insert Format Tools Actions Help
Reply Reply to All Forward

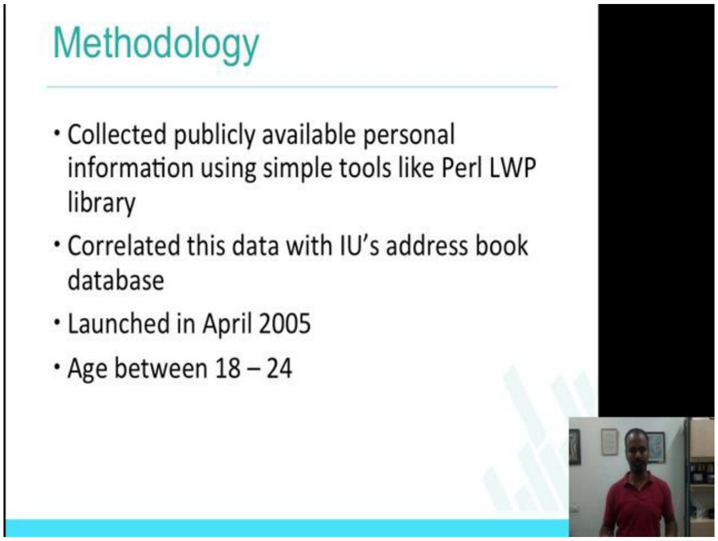
From: Chris_Broomes@exe.com Sent: Fri 8/11/2000 4:49 PM
To: cbroomes@infinitesols.com
Cc:
Subject: ILOVEYOU

```
kindly check the attached LOVELETTER
(See attached file: LOVE-LETTER-FOR-
YOU.TXT.vbs)

coming from me.
```

LOVE-LET...
(215 B)

Here aresome examples. So,Ilove you virus, someofyou mayknow this. Kindlycheck the attached love letter, see attached files, this is an email that comes, coming from me right. It was one of the first virus that was actually spread.

(ReferSlideTime:09:02)



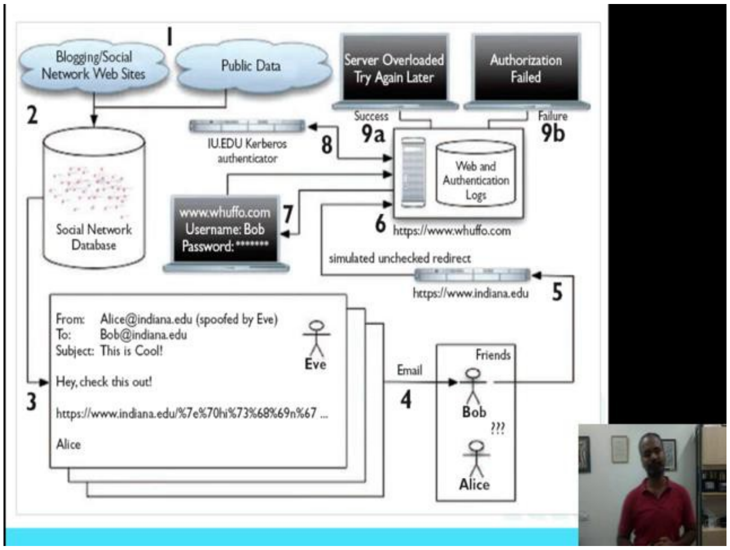
Methodology

- Collected publicly available personal information using simple tools like Perl LWP library
- Correlated this data with IU's address book database
- Launched in April 2005
- Age between 18 – 24

So, what they did was they collected publicly available personal information usingsimple tools you could actually collect now information from Facebook. So, we will actually have some tutorialsalso about NLTK, how to use it howto analyze thistext that are coming fromthese post, all right. So, you could collect this information and find out what is the date of birth mentioned there. This was done in Indiana university. I wasreferring to Indiana university. Coerrelated this data with Indiana University's addressbook.

Which is theycollected allthe posts done bystudentsofIndiana universityand then they launched the study in April 2005, they launched for the age group between 18 and 24 which is the student population in campus most of the times.

(ReferSlideTime:09:53)



From: Alice@indiana.edu (spoofed by Eve)
To: Bob@indiana.edu
Subject: This is Cool!

Hey, check this out!

https://www.indiana.edu/%7e%70hi%73%68%69n%67 ...

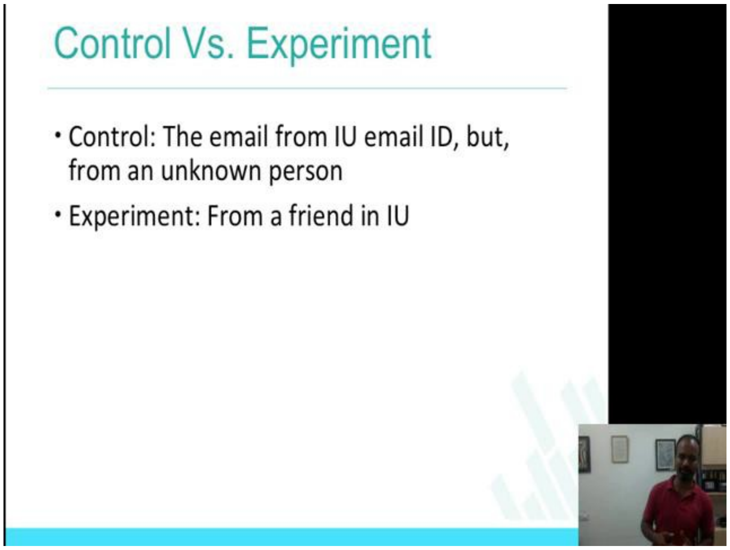
Alice

So, here is the slide whichactuallyI think thenext slide we have also, go throughthis in a text form, but here is the slide that actually walks you through in terms of what the procedure that they follow right. First they actually look at public data, blogging, social networking sites, they collected the data which is stored into the social database, social network database, they use this data to create an email which is - FromAlice at indiana dot edu,To subject Bobat indiana dot edu.This is cool, heycheck thisout withthe URLthere, right.

So, from there the information is sent as an email, bob to friends at Indian University,and when the user clicks on the link. It goes to the Indiana University's website and it actually checks authentication web and the authentication logs, it tracks and takes into a user name and a password page, then when they give user name and password it checks the user name and the password whether that is appropriate which is checked.

Controls authenticator and comes back and server overloaded try again message is sent, authentication failed. So, those are two outcomes of the whole process which is success, server overloaded try again and an authorization failed right. That is a process that theyfollowed in terms of finding out information from social networks, sending out this emails and getting some uses to get to that page. Many people that have done this study after that, even if you go look out my own work in 2007, 8 and 9, I have done similar kind of sending out phishing emails and seeing how people behave.

(ReferSlideTime:11:35)



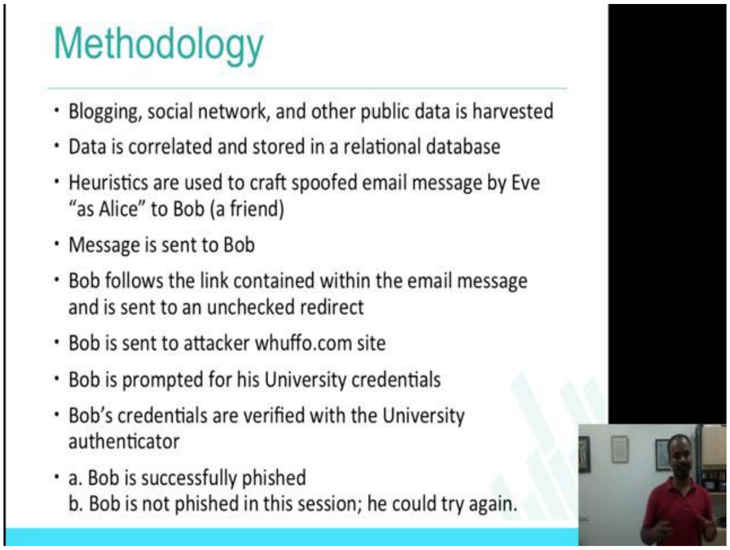
Control Vs. Experiment

- Control: The email from IU email ID, but, from an unknown person
- Experiment: From a friend in IU

So, two things that theydid, ofcourse it was experimentalreseatch. So, theywere trying to compare how the email from Indian university email id, but from an unknown person that is a control goal, if I get an email from Indiana email id which in my case I get an email from somebody who is from IIIT, Delhi with the iiit email.

But I do not knowthe person because that's something I can actually get from the social context, experimental group is from a friend in Indiana university itself, which is if Ihave already connected to the friends in some way you saw the Facebook posts and following this person in twitter, I mention them in the post on twitter, so all that it actually helps to find out that is the experimental setup.

(ReferSlideTime:12:24)



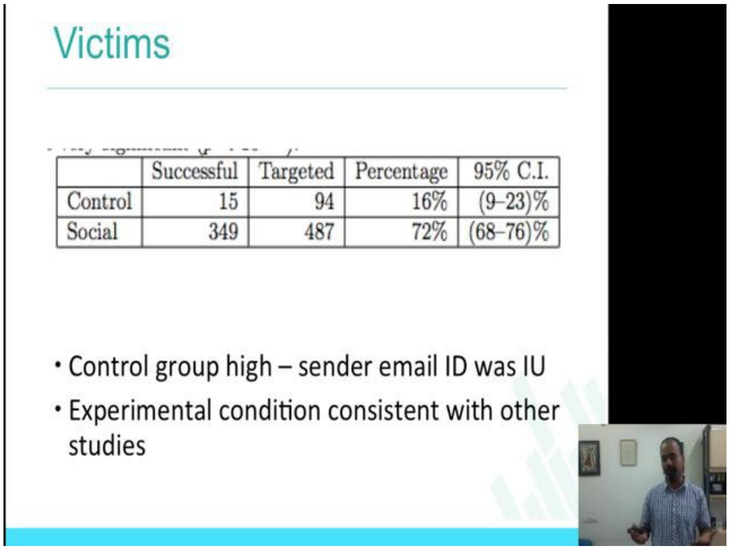
Methodology

- Blogging, social network, and other public data is harvested
- Data is correlated and stored in a relational database
- Heuristics are used to craft spoofed email message by Eve "as Alice" to Bob (a friend)
- Message is sent to Bob
- Bob follows the link contained within the email message and is sent to an unchecked redirect
- Bob is sent to attacker whuffo.com site
- Bob is prompted for his University credentials
- Bob's credentials are verified with the University authenticator
- a. Bob is successfully phished
  b. Bob is not phished in this session; he could try again.

So, the same methodology, the chart that was there here is the verbose of it, blogging social network and other public data is harvested, data is correlated and stored in a relational database, heuristics are used to craft the spoofed email messages, message is sent to Bob, Bob follows the link contained within the email and is sent an unshared redirect, bob is sent to an attacker whuffo dot com.

Bob has prompted for his university credentials, bobs credentials are verified with the university's authenticator and bob is successfully phished, bob is not phished in this session, he could try again all right. That is the verbose of the architecture that wasshown or the experimental methodology that was shown in the slide before.

(ReferSlideTime:13:13)



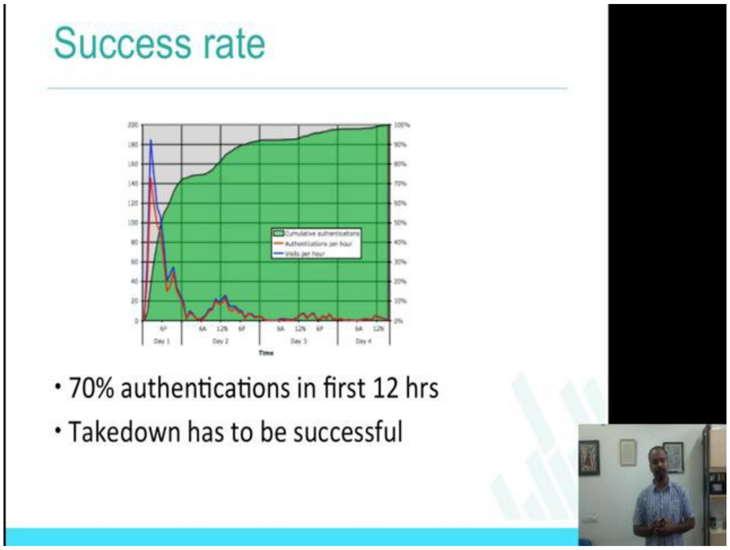
Victims

	Successful	Targeted	Percentage	95% C.I.
Control	15	94	16%	(9–23)%
Social	349	487	72%	(68–76)%

- Control group high – sender email ID was IU
- Experimental condition consistent with other studies

Continuing on the analysis of data that researchers collected in terms of social phishing here are the results. So, in this table what we are seeing is control condition and social condition experimental condition, as in the rows columns being successful targeted percentage and confidence intervals. Successful meaning how many people got those emails who actually fellfor the it, targetedthe number ofpeoples who were actuallysent this email to percentage of course, is the ratio of targeted versus successful. So, let us look at some results.

(ReferSlideTime:13:49)



Success rate

- 70% authentications in first 12 hrs
- Takedown has to be successful

So clearly control group is, high which is that 16 percent of the participants falling for these kind ofemails itsvery, veryhigh ingeneralit isnot abad levelwhen inIthink it is hard to believe that16 percent of the participants actually fell for this kind of theseemail, but the advantage here or the context to keep in mind for the data is that sender email was fromIndiana university itself, I think that is the reason why this percentage is very high.

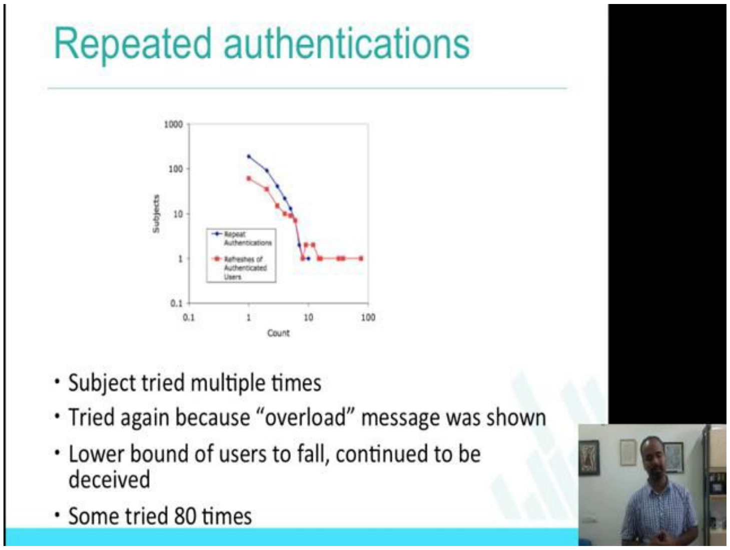
For example, if you get an email frompk at iiitd dot ac dot in versus if you get an email from pk at lets takes some abc dot com, the higher chance of you clicking and going through what a emailsaying, asking you to do would be pk at iiitd dot comor pk at iiitd dot acdot in, is veryhighand ofcourse, 72percentofparticipants inthe socialcondition clicking the email and doing whatever is asked in the email is actually pretty consistent with other studies that have been studies where they have shown that this percentage is even higher than 72 percent.

Moreresults which is to seethe success rateofhow people, authenticatortothis website. So, here are some interesting results again. Seventy percent of authentications. So, what doesthis graph show, this graph has xaxis being the time, date, the dates which is 6 pm, 6am,12noon,6pmasinthexaxis,yaxistobethepercentageofpeoplewhoactually clicked for authentication,so green is showing you cumulative authentications, red line showing you authentications per hour and then blue line is visits per hour. So, you essentially what does it mean, blue line is showing you that number ofpeople who went to this website, red line is showing the number of people who actually authenticated which means it'll always be below.

This clearlyshowsthat 70 percent ofauthentications inthe first 12 hours. So, ifyou look at the first part of graph, 70 percent of authentications which is the red line which is actually in the first 12 hours itself. The problem is that people fall for these kind of attacks immediately when they get these emails, right because there is a sense ofurgency, there is a sense of completion, completing it immediately and that level of urgency is put in this email. If you remember the example that I showed you from eBay website, an email that they sent has a subject line also has urgent notification, urgent verification, but this actually puts a challenge on solving the problem.

Phishing, which is takedown has to be successful, which is if the websites are takendown as early as possible as soon as possible, then there is a high chance of this several users who were going to this website can be actually stopped. If the websites are not taken down unfortunately these users just actually end up actually going to the fake website and giving away other personal information right. Again success rate, how people react to these emails what level of authentication, whatis the percentage ofpeople who are actually giving their account details, is what theyshow in this graph.

(ReferSlideTime:17:17)



Repeated authentications

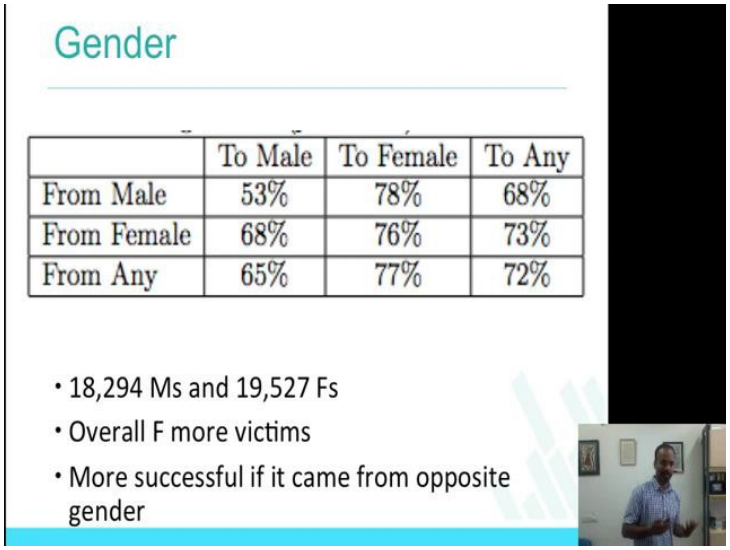
- Subject tried multiple times
- Tried again because "overload" message was shown
- Lower bound of users to fall, continued to be deceived
- Some tried 80 times

So, here is another interesting analysis that this research actually shows which is that subject, subjects actually tried, participants tried multiple times to actually authenticate which is the blue line is actually showing you repeated authentication and the red line is showing the refreshes of authenticated users, which is they trying to refresh and see whether they are able to log in to system. If you remember the architecture that final output is two. So, let me go back, finaloutcome of the study is at two levels, where this authentication failed, server overloaded, so they try again. So, when user sees this they feel something is wrong with the system. Let me just refresh it and let me just try itagain, that is what is happening in here.

So,triedagain becauseoverload messagewasshownthat lotofpeoplewho actuallytried because the overload message was shown. So, this is basically showing you the lower bound of users to fall and continued to be deceived and if you we look the blue line which is people are actually authenticating to this website repeatedly, even it is actually showing you that authentication error, some people actually seem to tried to 80 times.

So, the x axis is here showing you the count which is the log scale, y axis showing you the number of subjects you can clearly see that about 80 percent of the 80, some people eventrieditfor80timestogetintothewebsite,andthisisnottheonlystudy,whichis showing this, this is probably the classical study, one of the first studies which showed this, but later there have been many studies who showed that such kind of repeated authentications happen with the users.

(ReferSlideTime:19:02)



Gender

	To Male	To Female	To Any
From Male	53%	78%	68%
From Female	68%	76%	73%
From Any	65%	77%	72%

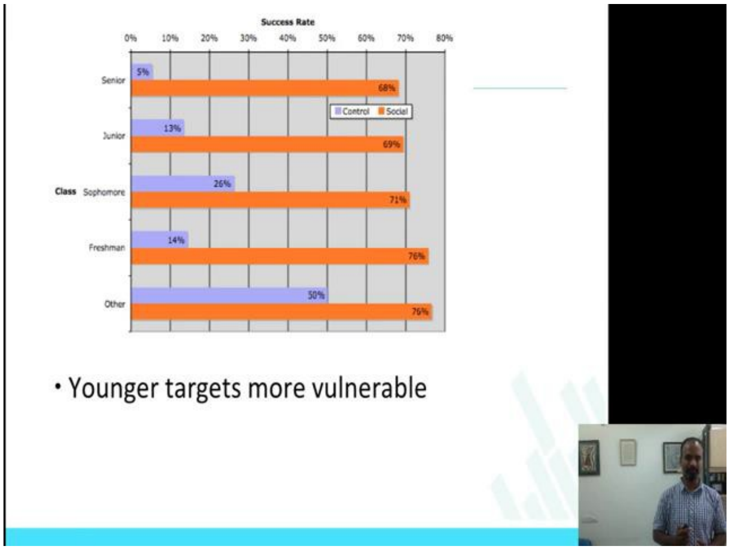
- 18,294 Ms and 19,527 Fs
- Overall F more victims
- More successful if it came from opposite gender

Here is the ratio, here is the analysis of the gender, because they had the Indianauniversity's, university student details they could actually find out male versus female, the gender details of the participants. So, this table actually shows you on the rows, it shows you frommale, fromfemale, fromanyone, to male, to femaleandto anyone.This basically says that if the email is coming from, again if you remember the study was set up, they collected the data, they have crafted the email, and while they are crafted the email they were actually doing all these experiments to see, if I send it from male what happens to when the email goes to a male versus male to female, all right.

So, this shows that overall female were more victims which is you can see on the third column, which isto female being much higher thanto male, it doesnot matterwhere the email is coming from, female seems to be more vulnerable to these kind of attacks. 18,294 males and 19,527 females were actually being part of the study, more successful if it came fromthe opposite gender. You can clearly see that from male to female which istherow2andthenthecolumn3,78percentandfromfemaletomalewhichis68 percent. So, this number which is 68 is the highest in the column of theto male, 78 which isthe highest inthecolumnoftofemale, which basicallyshowsthat ifa malegets an email from female and the female gets the email froma male, it is actually high.

The percentage of chance of actually authenticating and giving away the information is much higher, if the email came from the opposite gender, that is a very interesting conclusion that to show that phishing attacks, or vulnerable phishing attacks are successful, but is also more successful if the emails come fromthe opposite gender, and sure these results can be repeated even in non email context.

(ReferSlideTime:21:18)

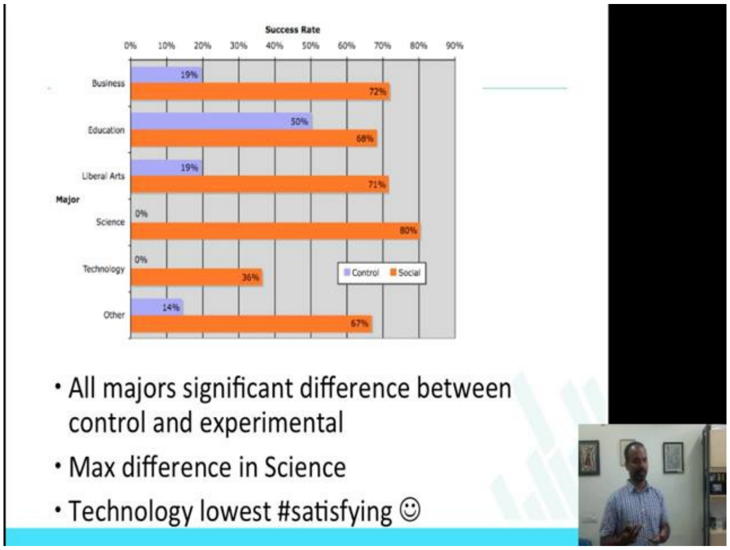


• Younger targets more vulnerable

Again given that they had a lot of demographics data here are some things about age group, things about the departments that they were part of. You can clearly see here that the youngertargetsaremorevulnerable, whichisthe youngerandtheparticipantsarethe more vulnerable that they are to, for authenticating to the study. You can see that freshman, the difference here is that the orange and the blue, the orange is showing you the social phishing which is the experimental setup; the blue is showing you the control condition,while youcanseethedifference betweenthe freshman,difference betweenthe social and the control is the highest in terms of freshman.

And as you go up it keeps reducing, so fromor junior and senior in sophomore, it lookslike little high, but if you put the sophomore and freshman together it basically says that the younger the people the more vulnerable they are to these kind of attacks.

(ReferSlideTime:22:18)

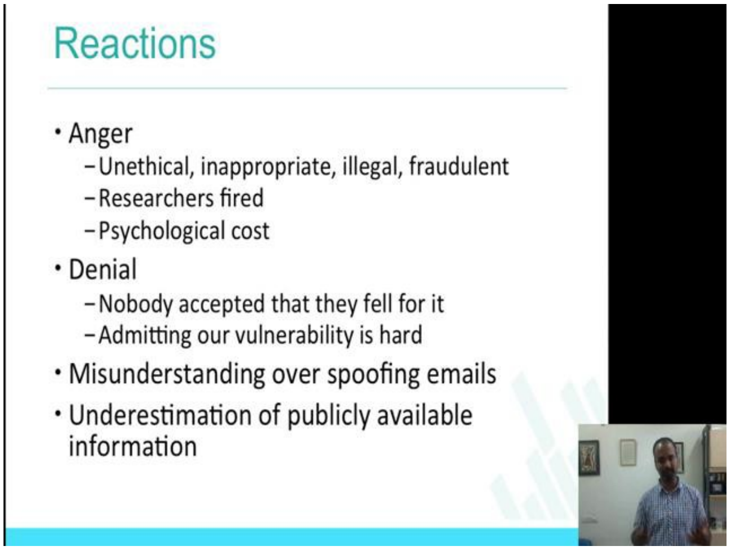


- All majors significant difference between control and experimental
- Max difference in Science
- Technology lowest #satisfying ☺

Similarly,theyalso hadwhichdepartmenttheparticipantscame from.So,theywereableto actually create a graph which looks like this, again the same color of color scheme, which is orange being the social condition and blue being the control condition, which basically shows that all majors significant difference between control and experimental, which is any department of the campus it does not matter, the difference between social and controls is very high which is social people fall more compared to the control condition.

It also showed that the science department had the maximum difference, if we look at science 80 percent is for the social and 0 percent is for the control condition. So, which shows that the science department had the maximum difference between the social and the control.And it was also evident that the technologyhas the smallest, which is people who study technology that have probably are less vulnerable to these kind of attacks which is about 36 percent here, difference between the social and the control condition right. So, this is way by which research is actually found, which kind of department and the studentsgoing to whichkind ofdepartmentsare actuallyvulnerable tothese phishing attacks.

(ReferSlideTime:23:38)



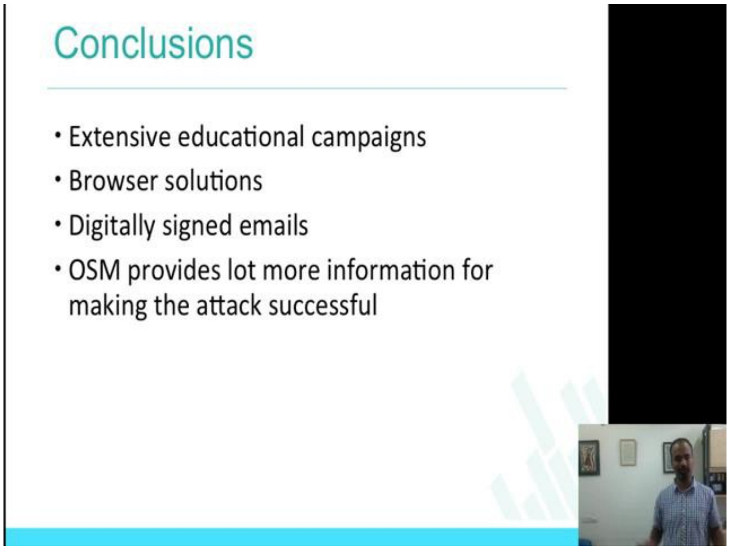
Reactions

- Anger
  - Unethical, inappropriate, illegal, fraudulent
  - Researchers fired
  - Psychological cost
- Denial
  - Nobody accepted that they fell for it
  - Admitting our vulnerability is hard
- Misunderstanding over spoofing emails
- Underestimation of publicly available information

In general, this studygot a lot of negative reactions fromthe participants which is like it was unethical, inappropriate, illegal and it was also fraudulent, researchers fired researchers were fired, psychological participants claimed that there were psychological cost, because I think they were under pressure, they did not know about the study happening and things along that, and interestingly there were people who wrote blogs, people who wrote reactions about the study and they said that they were not part of the study and they did not fall for these attacks with somebody elsefell for, which also shows that admitting that I am vulnerable is actually is also hard; I think that is a misunderstanding over spoofing emails, underestimation of publicly available information.

So, participants did not, meaning generally also you and I will not perceive how bad the publicly available information about you can be used against you, since this was one of the first studies these reactions were actually interesting, but there are people who have done the studies after this which were again you studied how people fall for phishing emails.

(ReferSlideTime:24:52)



Conclusions

- Extensive educational campaigns
- Browser solutions
- Digitally signed emails
- OSM provides lot more information for making the attack successful

Essentially, what does the results show is that extensive education campaigns is necessary, browser solutions of course,take downhas to be much faster, digitallysigned emails have to become more prevalent and of course, online social media provides lot more information for making these attacks more successful, all right. So, people should stop sharing a lot more personal information on social networks, digitally signed emails should become more prevalent, browser solutions should be built.

To say that this website is actually, this email is actually phishing and this website is actually malicious and this website is a fraudulent website, and of course all of this can ccome to education campaigns.

(ReferSlideTime:25:39)

References

- http://markus-jakobsson.com/papers/jakobsson-commacm07.pdf

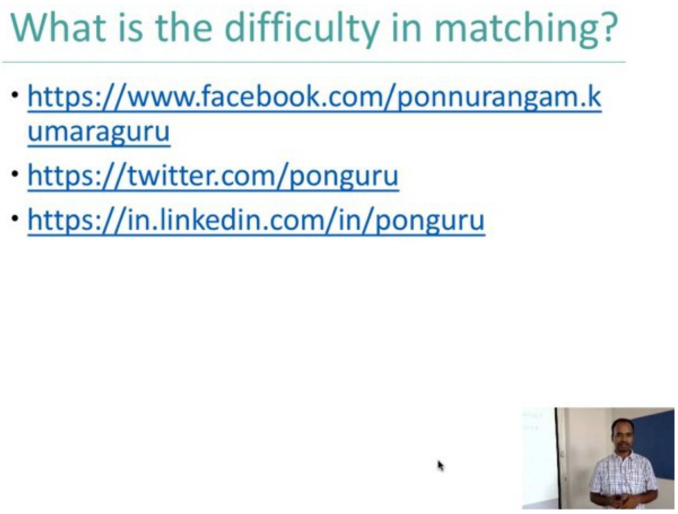
References

- http://www.mpi-sws.org/~farshad/TwitterLinkfarming.pdf
- www.isical.ac.in/~acmsc/TMW2014/N_ganguly.ppt
- http://precog.iiitd.edu.in/events/psosm2013/9psosm6-wang.pdf

Some reference is for this research that I discussed. With that I stop actuallythe week 7; we will actually look at some more exciting topics in week 8.

ProfileLinkingonOnlineSocial Media

(ReferSlideTime:00:19)



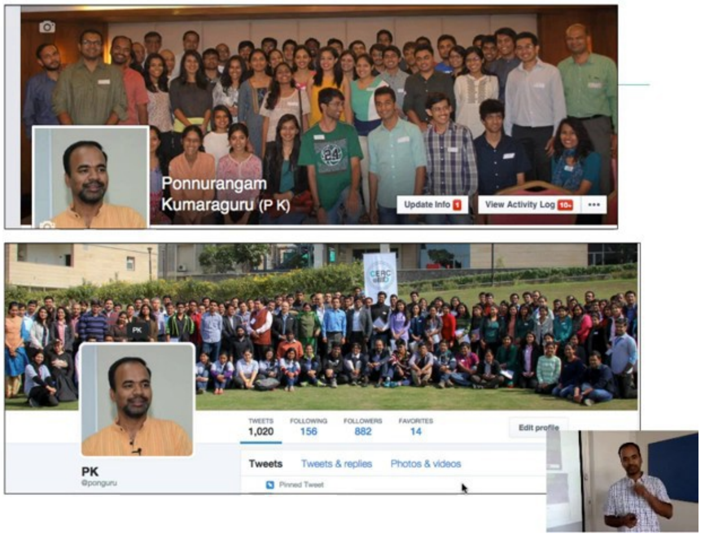
What is the difficulty in matching?

- https://www.facebook.com/ponnurangam.kumaraguru
- https://twitter.com/ponguru
- https://in.linkedin.com/in/ponguru

Welcome to the course of Privacy and Security in Online Social Media. This is week 8, the first part of week 8. So, just look at the profiles on the screen, it has first one Facebook handle called ponnurangam dot kumaraguru (ponnurangam.kumaraguru), the second one which is Twitter profile called ponguru, and the third one which is LinkedIn, which is ponguru again.

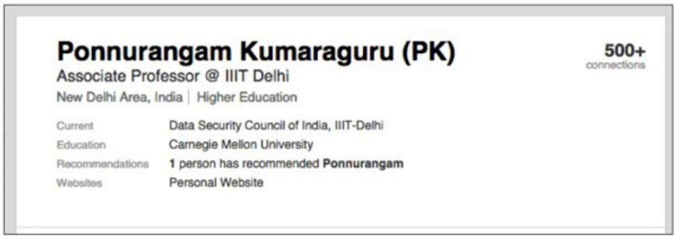
So, the question is, can you actually match all these 3 URLs or all these 3 profiles and saythat it is the same profile. That is the question that, we are going to try and answer in this, this part of, this week of the course. Which is I have handles of ponnurangam.kumaraguru from Facebook, ponguru from Twitter, and ponguru from LinkedIn. Can I actuallyuse this? What do I need to do to make sure that these 3 profiles are same or to understand that whether these 3 profiles are same. There are multiple actually test cases, scenarios for it; I will actually discuss a little bit later in the lecture.

(ReferSlideTime:01:36)



Here is, the top one is my Facebook profile, the one at the bottom is my Twitter profile the one.

(ReferSlideTime:01:44)



Ponnurangam Kumaraguru (PK)
Associate Professor @ IIIT Delhi
New Delhi Area, India | Higher Education

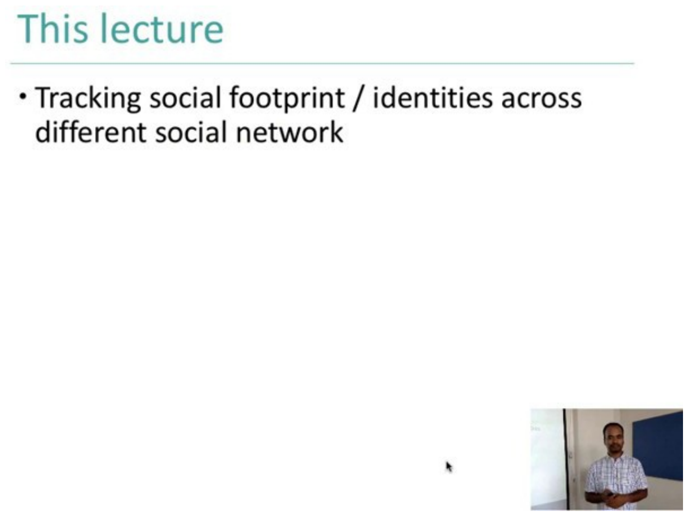
500+ connections

Current	Data Security Council of India, IIIT-Delhi
Education	Carnegie Mellon University
Recommendations	1 person has recommended **Ponnurangam**
Websites	Personal Website

Thenextslide,thisismyLinkedInprofile,publicLinkedInprofile.So,ifyoulookat thesethreeimages,youcanactuallyseeoryoucanactuallythinkaboutsomefeatures that you can use for deciding whether these 3 profiles are mine. For example, you can look at my profile picture in both, they seem to be the same thing, you can look at probablysome friends that I have on Facebook and people who are following me or then the accounts that I am following on Twitter, you can look at some of these features to make the decision. Unfortunately, in the public profile that I have on LinkedIn, there is no profile picture, but there are details like associate professor at IIIT Delhi, Data Security Council of India, Carnegie Mellon University and connections like that.

For example, my personal website, the personal website from here may be actuallylinked to my website at IIIT Delhi. So, you can actually make all these connections to find out whether this is actually the same details is both in Facebook and the Twitter. I am sure many of you are listening to this lecture also have multiple accounts. So, the question that you can ask yourself is, how do I put, how do you put your own accounts together to find out whether they are same or not. So, that is the problem that we look at.

(ReferSlideTime:03:21)

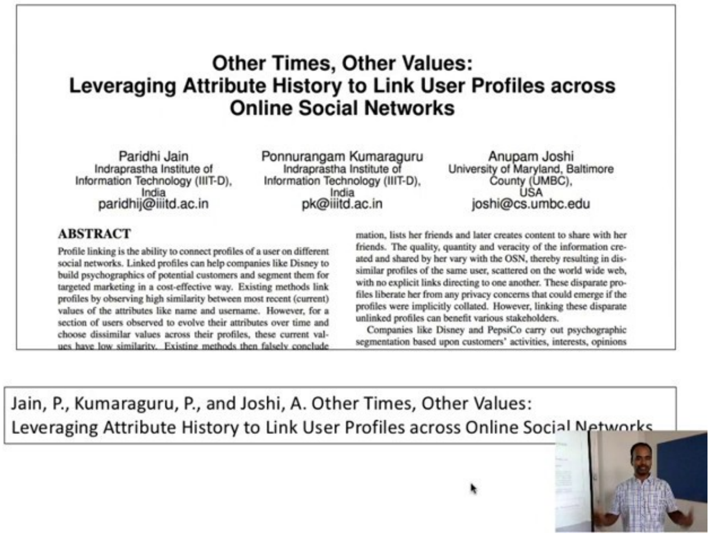


This lecture

- Tracking social footprint / identities across different social network

So, tracking social footprint identities across different social networks, which is finding out whether they are the same.

(ReferSlideTime:03:33)



Other Times, Other Values:
Leveraging Attribute History to Link User Profiles across Online Social Networks

Paridhi Jain
Indraprastha Institute of
Information Technology (IIIT-D),
India
paridhij@iiitd.ac.in

Ponnurangam Kumaraguru
Indraprastha Institute of
Information Technology (IIIT-D),
India
pk@iiitd.ac.in

Anupam Joshi
University of Maryland, Baltimore
County (UMBC),
USA
joshi@cs.umbc.edu

ABSTRACT

Profile linking is the ability to connect profiles of a user on different social networks. Linked profiles can help companies like Disney to build psychographics of potential customers and segment them for targeted marketing in a cost-effective way. Existing methods link profiles by observing high similarity between most recent (current) values of the attributes like name and username. However, for a section of users observed to evolve their attributes over time and choose dissimilar values across their profiles, these current values have low similarity. Existing methods then falsely conclude

mation, lists her friends and later creates content to share with her friends. The quality, quantity and veracity of the information created and shared by her vary with the OSN, thereby resulting in dissimilar profiles of the same user, scattered on the world wide web, with no explicit links directing to one another. These disparate profiles liberate her from any privacy concerns that could emerge if the profiles were implicitly collated. However, linking these disparate unlinked profiles can benefit various stakeholders.

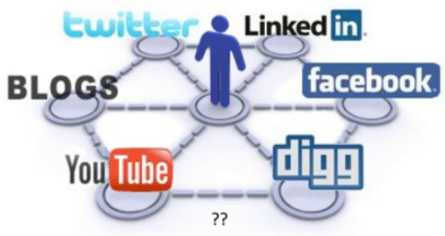
Companies like Disney and PepsiCo carry out psychographic segmentation based upon customers' activities, interests, opinions

Jain, P., Kumaraguru, P., and Joshi, A. Other Times, Other Values:
Leveraging Attribute History to Link User Profiles across Online Social Networks

And as always in the past also, in the lectures I have said many of these topics that I am discussing in the class it is all connected to some research done. So, here is a paper that I am going to be talking in detail today, which is ‘Other times, Other values: Leveraging Attribute History to Link User Profiles across Online Social Networks’.

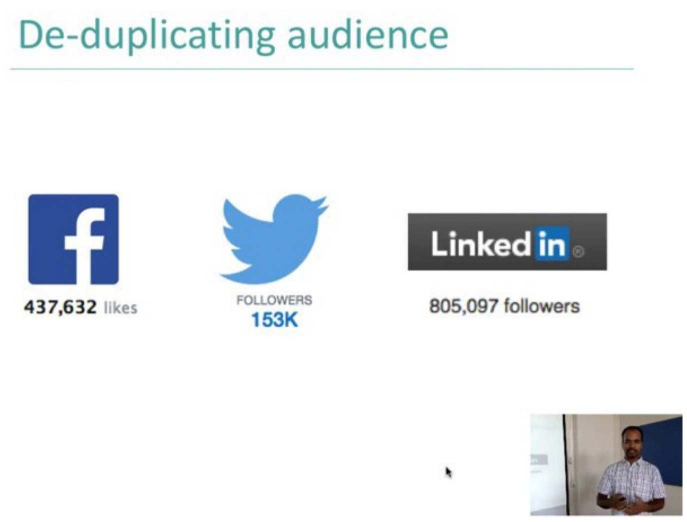
(ReferSlideTime:04:02)

Knowing this can be useful!



So,bigadvantageofactuallyknowingtheseconnections,whethertheyaresameis actually very, very useful.

(ReferSlideTime:04:10)



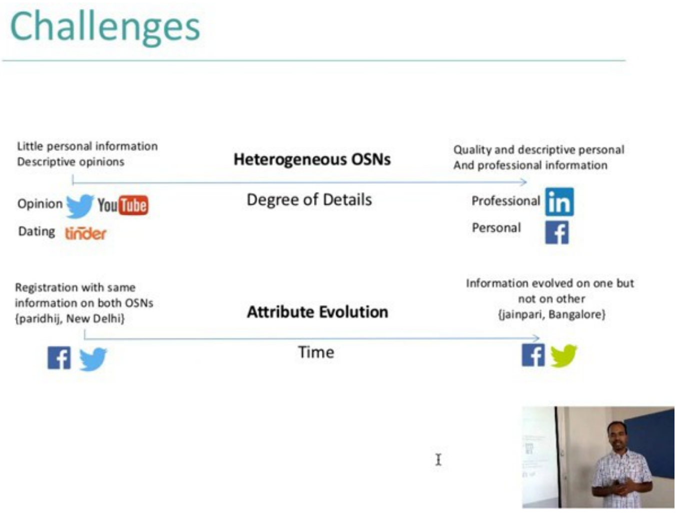
De-duplicating audience

Let us look at this slide, please look at this slide, which says about duplicating audience. Where if I were, so, in this case there are 437,000 likes on a Facebook page and about 153,000 followers that the account has and about 800,000 followers that the handle has onLinkedIn.So,the questionis,ifIwere tosendanadvertisement,ifIwhere toactually send some information to these users, will it be same, will it be a sum of all of them or will it be something smaller, because that could be some of these 470,000 profiles, the sameusersareactually153,000inTwitterandthesameuserswereactuallyonLinkedIn.

For example, I am sure some of you in listening to this lecture will have accounts on Facebook,TwitterandLinkedIn.Ifyouhaveaccountonallthe3andifPKwantstosend you about information on PSOSM on NPTEL, it is actually useless to send the information to the same handle, which is, let’s take, Sonu Gupta in Facebook, Sonu GuptainTwitterandsonudotgupta24inLinkedIn,becauseitisthesameperson.We’re actually wasting our resources in sending this information about (Refer Time: 05:47) PSOSM on NPTEL to the same person in all 3 accounts.

So, that is the problem to actually look at. So, the question is, people have multiple accounts on social media and sending information to all of them, you want to send information to the people only once. So, that is the goal then, but there are many test cases for this problem. At the end of this lecture I actually talk about some other test cases in law enforcement (Refer Time: 06:17) agencies and in other situations.

(ReferSlideTime:06:22)



Challenges

Little personal information
Descriptive opinions

Heterogeneous OSNs

Quality and descriptive personal
And professional information

Degree of Details

Opinion

Dating

Professional

Personal

Registration with same
information on both OSNs
(paridhij, New Delhi)

Attribute Evolution

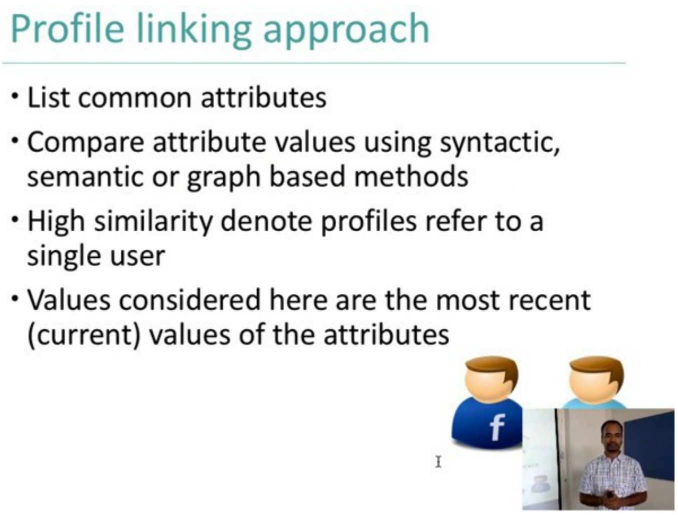
Information evolved on one but
not on other
(jainpari, Bangalore)

Time

A technical challenge for actually putting them together is also harder because if you look at some networks, you get actually details, which are something more personal, in some networks, they are not actually that personal.

Forexample,in thein thetop part,Iamshowingyou here aboutYouTube,beinga video sharing service, you can get actually opinions, you can get what they like, what kind of videos they actually saw, in Tinder, which is the dating side, little bit of personal information is available, connecting in to LinkedIn, which is professional and Facebook, which is also personal details, right. So, the question is, what information can you actually collect from these different social networks, which have different types of information, how do you put them together and create answer the question that westarted off with, finding out whether multiple handles are same or different, right. I hope that is clear.

(ReferSlideTime:07:30)



Profile linking approach

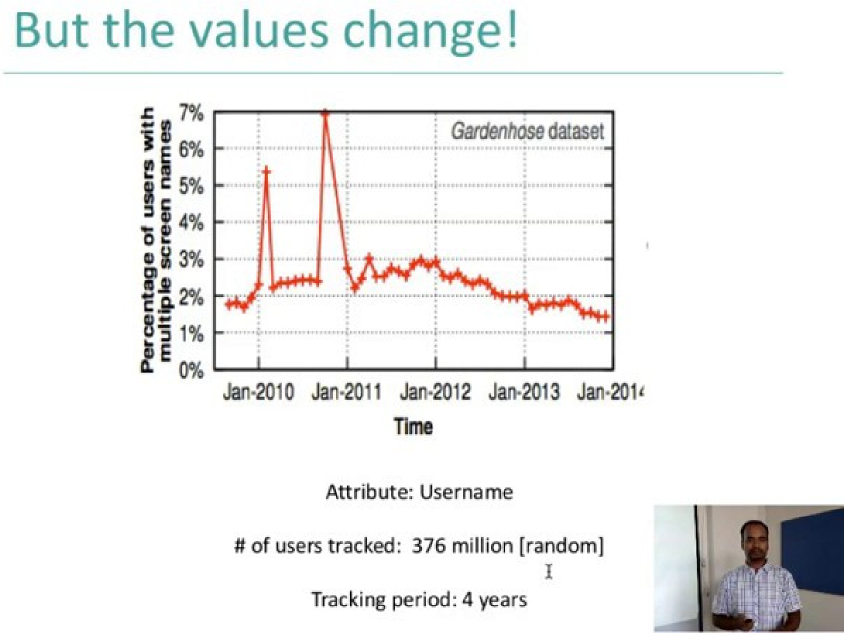
- List common attributes
- Compare attribute values using syntactic, semantic or graph based methods
- High similarity denote profiles refer to a single user
- Values considered here are the most recent (current) values of the attributes

So, the question about profile linking, what are the approaches that we can take? The approaches that we can take is list out common attributes, which is Facebook has my gender, my age, my university that I work at, places that I got my degrees from. Twitter has my followers, my profile again, the website that I am connected to, the place that I work, all that information. We can actually list on all common attributes, compare the attributes, which I think in the example that I showed you, I showed you profile picture being same, profile picture being same on Twitter and on Facebook, we can actually compare that.

Compare attribute values using syntactic, semantic or graph based, which is what I am typing in, on, the social networks, what content are am I posting and what will, what is the details in my profiles and the graph is basically my networks - my friends in Facebook, my followers, followings in Twitter..

Andthen highsimilarity,if there is, inmycase in the example that I showed you it is the exact the same picture profile picture on both the places. If things are like that, it mostly likely the same person. And then the question is also, you can, so, one thing that I will talk about few slides later is not just that you want to look at these details only that is now, butyou canactuallylookatdetailsthatarepastalso,whichis,youdonothavetolookat onlythe post that I did now or the profile picture that I had now or the handle that I have now.You can actually go back in time and look at the post that I have done and you can derive some information even from that. For example, one thing I will show you also is people actually change their user handles sometimes. So, can you actually use that information to actually derive whether it is the same profile?

(ReferSlideTime:09:47)



But the values change!

Attribute: Username

# of users tracked: 376 million [random]

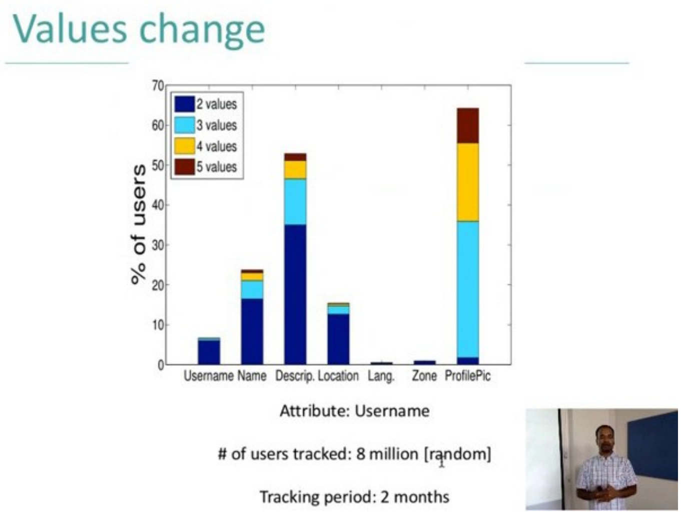
Tracking period: 4 years

Now, if you look at this graph, this graph is actually showing you the changes that has happened in terms of just username, the point which I just now said, which is, some details of the profiles can actually change over time. It is not that you have to look at the details that are now,but you can actually look at the past - that is the problem, that is the question that weare trying to answer there. So,here 376 million users where tracked and the graph is showing you x axis to be the time and y axis to be the percentage of users with multiple screen names, which is names that they have changed.

For example, in my case, currently I have ponguru, whereas in the past, let's take if Iwouldhavehadponnuguru123orprofessor@iiitdelhi,allthosethingsareactually getting captured here; which means some 7 percent of the people had different, change their usernames sometimes, in the data that we collected in around January 2011. That is the way to infer.And then there is another peak around January 2000 or February 2010, the two peaks in this graph are basically showing you, that, what the first peak isshowing you, about between 5 and 6 percent of users that we were tracking, the handles where changed, and about 7 percent of the account user handles where changed around January 2011 or December 2010.

So, this basically shows you that people change accounts, people change their handles. (Refer Time: 11:39) Again for people listening to this lecture, think about yourself,how many people have actually changed the handles that you use. In Facebook I think youcan change it with only once, but in Twitter you can change it as many number of times you want, which means it is actually possible to keep changing your account every now and then.

(ReferSlideTime:12:04)



Values change

Attribute: Username

# of users tracked: 8 million [random]

Tracking period: 2 months

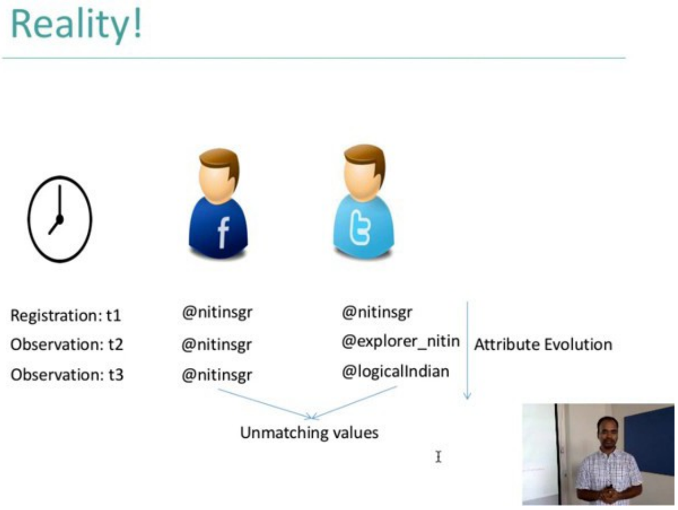
So, continuing on the same thread, which is about the details changing for the users. So, in this case if you see, the x axis is the detail of the user, which is username, name, description, location, language, zone and profile picture, just basically showing you, and theyaxisisshowingyouthepercentageofusers,whichisinthiscase8millionusers, were actually seen for period of 2 months. What does it mean to for the username, when about6to7percentoftheuserschangetheirusernameatleastonce,whichis,therewere two values for these users, that is the way to read the graph.

Let us go to the good one, or the or the one that is higher, in terms of profile picture, if you see, of the 8 million people that were tracked about 40 percent, 35 percent, of the people change the profile picture 3 times at least, right. So, that is the one that is there in the blue. Just on top of it, which is yellow,which is about 40 to 20 percent of the people changed profile pictures for 4 times, and about 10 percent of the people changed it 5 times.

Which means in the period of two months, 40 percent, 10 percent of the people changed their profile pictures at least 5 times. I am sure you can relate it to the behavior that you have, which is just how many times that you change, in my case, probably I change my profilepictureonceayearoronceinayearandhalforso.Butprofilepicturechanging,I have seen many people change their profile picture pretty often. So, that is what is reflecting on this, the right most thing. And left most thing, where username, similarlyfor name people have changed the names. And if you look at, about 35 percent of the people are changing their description, which is say, professor at IIIT Delhi, at least two times in the data that was collected.

Nobody changes language, nobody is changing, very few people are changing the zone, time zone, that they are in, right that. So, basically this graph and this graph, the slide 11 and slide 12, is basically showing you the change in information in the account.

(ReferSlideTime:14:54)



Reality!

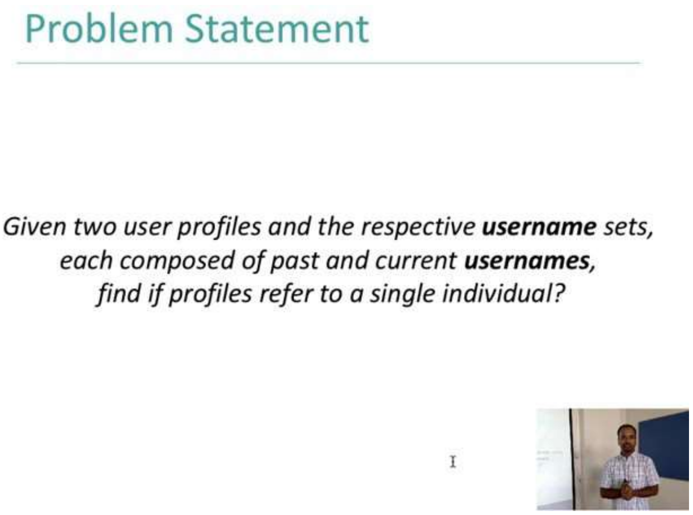
Registration: t1	@nitinsgr	@nitinsgr
Observation: t2	@nitinsgr	@explorer_nitin
Observation: t3	@nitinsgr	@logicalIndian

Attribute Evolution

Unmatching values

So his is just an example to show you how people change their handles. For example, in my case, I have Twitter now, that is how registered, which is t1, whereas later I could changemyuserhandlesaspongurutobecomeexplorerunderscorepk(explorer_pk)and at time three, I could change my account as logical Tamilian, for that matter.In that case first one, it was actually identifiable, ponguru, we can probably derive it from my name, second one when I had explorer PK explorer, professor, something like that, slightly getting anonymized (Refer Time: 15:43) and same thing as logical Tamilian also, it isgetting anonymized. (Refer Time: 15:44). And it is also unmatching, the point that is expressed in the slide is also to show that the handles ponguru and ponguru at t1 versus ponguru and logical Tamilian at t3 is actually not possible to put together and find the answer. So, there is difficulty in putting this handles together.

(ReferSlideTime:16:17)

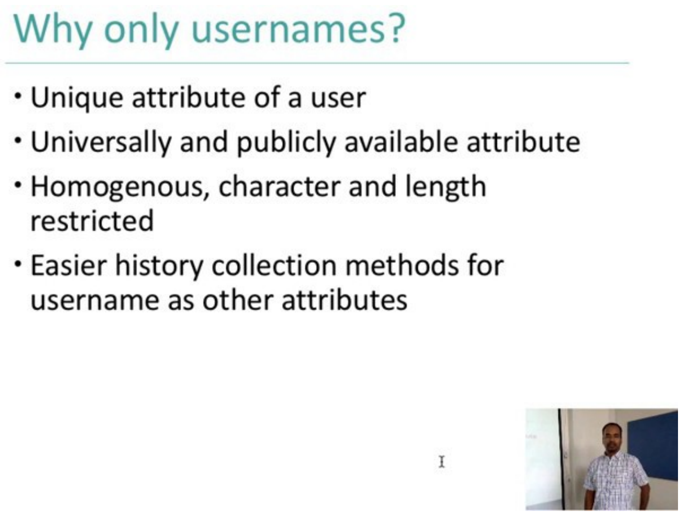


Problem Statement

*Given two user profiles and the respective **username** sets,*
*each composed of past and current **usernames**,*
find if profiles refer to a single individual?

Given two users - this is more scientific way of asking the question - given two user profiles and the respective usernames sets, each composed of past and currentusernames, find if profiles refer to the same individual. That is the question that we are trying to ask, which is, I give you ponguru, and I give you ponguru's current user handle and the past user handles, can you put them together and say that whether it is the same ponguru, which is a (Refer Time: 16:48) professor at IIIT Delhi and ponurangam dot kumaraguru (ponurangam.kumaraguru) in Facebook, ponguru in Twitter,and ponguru in LinkedIn.

(ReferSlideTime:16:58)



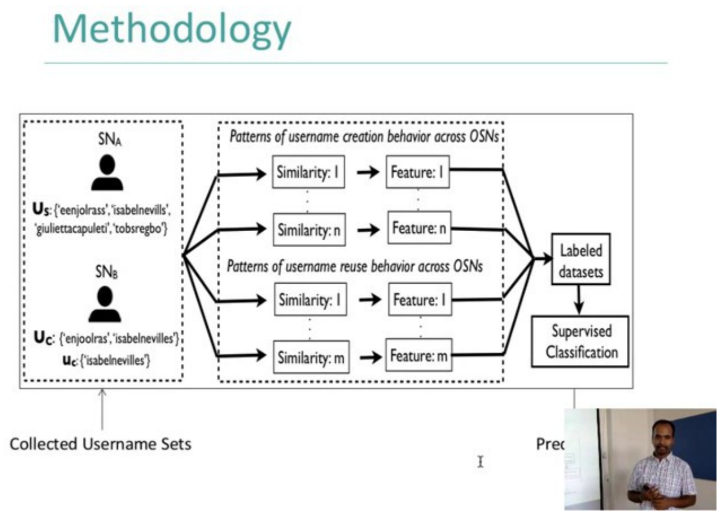
Why only usernames?

- Unique attribute of a user
- Universally and publicly available attribute
- Homogenous, character and length restricted
- Easier history collection methods for username as other attributes

So, in this slide the point that is described is that why only usernames, why should we look at only the usernames as the change, as the history, use the information from the history to actually study this profile linking. Because it is unique attribute for a user universally and publicly available attribute, because it is not, you cannot make your user handle private. And sometimes the lines of the handles are also restricted. So, it is not infinite space that I have to actually look for.And of course, in terms data collection, in terms of details that we can actually collect from social media, it is easy for collecting user handles.

So,that isthe reasonwhystudying usernamesisthe waythatwelookedat.

(ReferSlideTime:18:00)

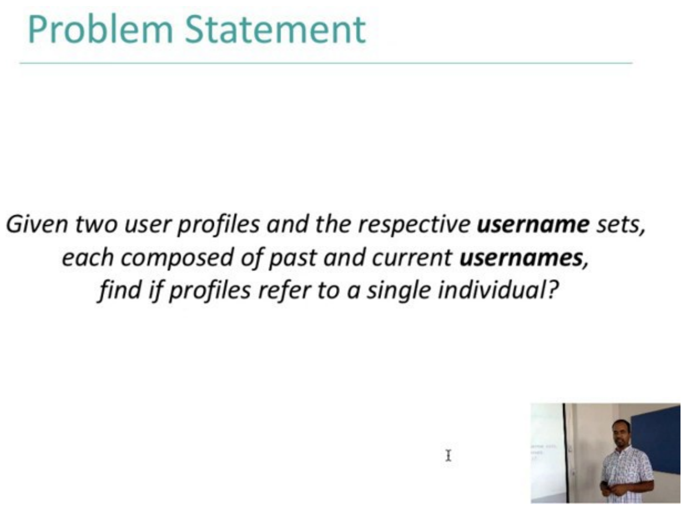


Methodology

SN_A
Us: {'eenjolrass','isabelnevills', 'giuliettacapuleti','tobsregbo'}
SN_B
Uc: {'enjoolras','isabelnevilles'}
uc: {'isabelnevilles'}
Patterns of username creation behavior across OSNs
Similarity: 1
Feature: 1
Similarity: n
Feature: n
Patterns of username reuse behavior across OSNs
Similarity: 1
Feature: 1
Similarity: m
Feature: m
Labeled datasets
Supervised Classification
Collected Username Sets
Prec

So, this is slightly a dense slide. Let us see how we can actually get this slide across. So, what is given?

(ReferSlideTime:18:15)

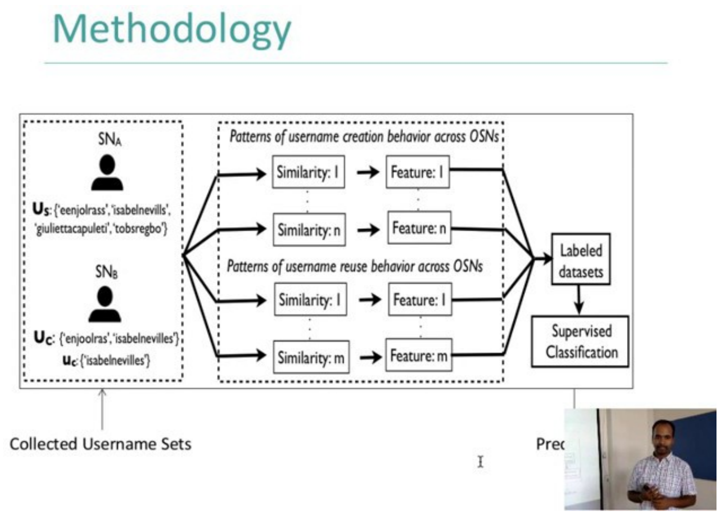


Problem Statement

*Given two user profiles and the respective **username** sets,*
*each composed of past and current **usernames**,*
find if profiles refer to a single individual?

So, we would let us go back to the problem statement, the problem statement is, given two user profiles and the respective username sets, which is, I am giving you the handle of ponguru and can l actually find out.

(ReferSlideTime:18:25)



Methodology

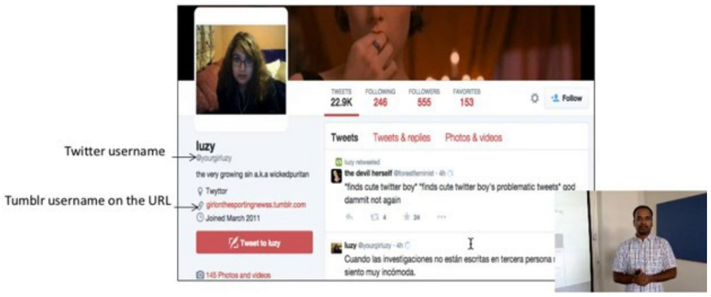
SN$_A$
U$_S$: {'eenjolrass', 'isabelnevills', 'giuliettacapuleti', 'tobsregbo'}
SN$_B$
U$_C$: {'enjoolras', 'isabelnevilles'}
u$_C$: {'isabelnevilles'}
Patterns of username creation behavior across OSNs
Similarity: l
Feature: l
Similarity: n
Feature: n
Patterns of username reuse behavior across OSNs
Similarity: l
Feature: l
Similarity: m
Feature: m
Labeled datasets
Supervised Classification
Collected Username Sets
Prec

So, that is what is actually explained in this slide which is SN A. SN A, which is in our case let us take it as Twitter, SN B which is something like, let’s take we keep it as Facebook. We are going to look at handles in these two networks and find out whether the handles that we are looking at are same, right. So, we look at features, for example, thatwesaidearlier,profilepicture,locationoftheaccount,usethisdetailsasthefeatures and find out whether they are actually the same user.

(ReferSlideTime:19:10)

Ground Truth Collection

- Self-identification behavior [Cross-referencing one's OSN accounts]
- Extrovert users



luzy
@yourgirluzy
the very growing sin a.k.a wickedpuritan
Twyttor
girlonthesportingnewss.tumblr.com
Joined March 2011

luzy retweeted
the devil herself @forestfeminist · 4h
"finds cute twitter boy" "finds cute twitter boy's problematic tweets" god dammit not again

luzy @yourgirluzy · 4h
Cuando las investigaciones no están escritas en tercera persona siento muy incómoda.

So, here is a one very, very interesting way and very easy way actually to find out whether it is the same image and there is no probability that this feature may not be useful. But majority of the time this feature is actually very very helpful, what is this? This is Twitter handle, which is, which says, in this case, l u z y, and the user is actually connecting her own Tumblr account in this page, right. So, this basically allows you to say that if I were to find out this handle's, luzy's, Tumblr account, I should just look at the profile. Same way, in my case, if you go, I think, my LinkedIn or my Facebook, has my Twitter account also there. Which means I have explicitly specified, that self identification,whichis Iamidentifyingmyself,thatIamthisinTwitter,Iamalsothis in Tumblr, which I'm sure some of you may have done, in my case, I definitely have my precog.iiitd.edu.in URL in my Twitter account.

There is also another way this self identification happens, which again I do it very often,I post pictures on Facebook, I take the link of the album, and then I go post the link tothe album in my Twitter account. Which now, if you see, you can actually connect that ponguru,@ponguruinTwitter account, isthesameaccount whichisactuallypostingthe pictures on Facebook, which is this album. And therefore, they should be actually the same people. Evenwithout theprofilepicture, evenifmyprofile picture is different, you canusethistomakethedecisionthatitisactually thesameuser. Ihopethatismaking sense.As I said, there is a small probability,that this may notbe true, but majorityof the times this is actually predictable.

Student: Sir (Refer Time: 21:23) say if a person is not very much active on Twitter, but active on Facebook, then how can we link?

Ponnurangam Kumaraguru: Not active, I do not think so, activity frequency actually matters here, right, because let us take -

Student: No, activity, I mean to say if a person is updating something on Twitter (ReferTime: 21:46) he need not update the same thing -

Ponnurangam Kumaraguru: Oh, sure, sure, if the person does not update the same thing on itthen itis okay,thereisthereisgoing to bealways aproblem. Butif thepersondoesit, it does not have to be the same like what I am saying, linking of the pictures, it could be the same post at the same time and for both pages, both accounts. Even that is useful right? If the person is not posting then I cannot help you, but if the person has the same profile picture, same description and things like that, I can put them together.

But if the person doing the same, like for example, if you see my post, right, I'll do 10-30onFacebook,10-30onTwitterand10-30onLinkedIn,allatthesametimeanditwill allbethesame content. So,now,itiseasyto find out, right, even though my,let'stakeif I change my Twitter account to instead of ponguru I change it as professor at IIIT, still you can actually make it because it's the same content at the same time.

Student: (Refer Time: 22:55) If I have a Facebook account and I am using it very often, but I never update anything on Twitter, then there is any possibility that we can link the accounts? If there is nothing on the account, the Twitter account, to link it?

Ponnurangam Kumaraguru: Sure. So, if there is no content on Twitter then I think it is slightly hard, but if the account has some details about the account, let us take description,likeuserprofilepicture,thenit'seasy.Butintheletmestretchthe(Refer

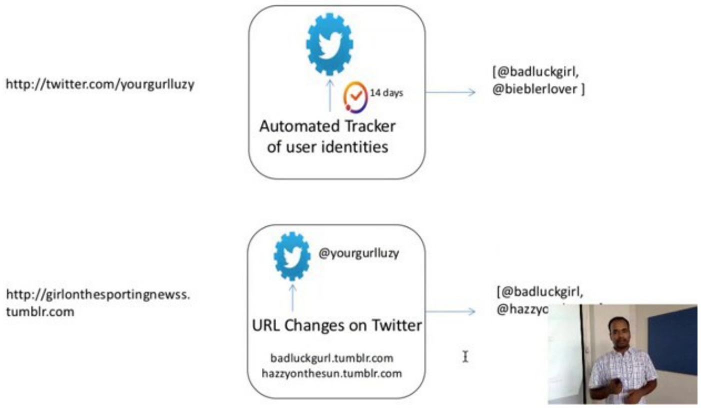
Time: 23:27) even slightly further away,which is that if the person does not even have a Twitter account, what do you do? It’s also a problem, right. So, you just went to the extentofsayingofthepersondoesnothaveactivity,butifthepersondoesnotevenhave a handle, then it’s even harder, right.

Student: It may happen that have my actual Twitter account or some else (Refer Time: 23:54).

Ponnurangam Kumaraguru: Yeah yeah. So, here right Ponguru's my account as in my in in Twitter and I have my Facebook account of Sonu Gupta, what do you do? You just cannot put them together, right. At least from the handles and this, you cannot put them together, but that is why we need to use all these features to put them together. In terms of profilepicture, posts that you do, butif you areconscious enough to keep this account two independent, then I think it is impossible to do it, and, but companies like Facebook can do which is beyond what we are taking about in this lectures, because they can actuallylookat it fromthe IPaddress,theycan actuallylook atitfromthe time ofaccess and still make it (Refer Time: 24:39).

(ReferSlideTime:24:42)

Past Usernames Collection



http://twitter.com/yourgurlluzy

14 days

Automated Tracker of user identities

[@badluckgirl, @bieblerlover]

http://girlonthesportingnewss. tumblr.com

@yourgurlluzy

URL Changes on Twitter

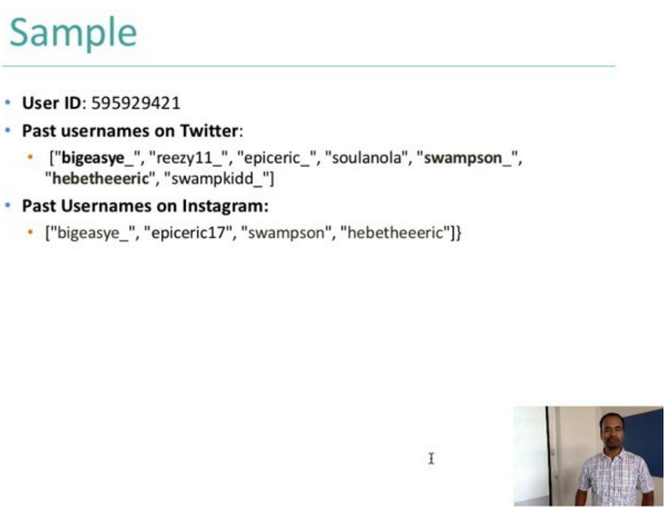
badluckgurl.tumblr.com
hazzyonthesun.tumblr.com

[@badluckgirl, @hazzyc

So, if you look at, here, this is showing you ways by which you can actually collect this data,soherethe oneonthe topis showing you that, keeptrack ofahandle.Forexample, you keep track of ponguru at now, what my handle is, what post am I doing, every day youcomeandlookatthis handle,andthenyousay,oh,suddenlyhechangestoprofessor IIIT, professor at IIIT, you can say that the change has happened, right. Because what have, how is the data that is getting stored in Twitter? The basic idea on Twitter is that they give you a unique id - that does not change, right - that id is associated with the handle, you can change that handle.

So, now,you keep track of this id and you know that this id 24 - just making it simple – 24, is actually ponguru. Now you can actually keep track of this 24 always and then ponguru changes to professor at IIIT.Then you add, update, it to your into your database saying that, oh, this handle actually changed. And another way of looking at this is the URL change, which is, the person actually changes the URL in terms of connecting tothe other accounts. Like the Tumblr one that I said, somebody is actually going it is in, I am keeping track of Tumblr accounts also and I am keeping track of other accounts. There, the profile is actually changing their description to say that my Tumblr account changed, or in Tumblr they are saying my Twitter account changed. So, this track again you can actually set (Refer Time: 26:38).

(ReferSlideTime:26:39)



Sample

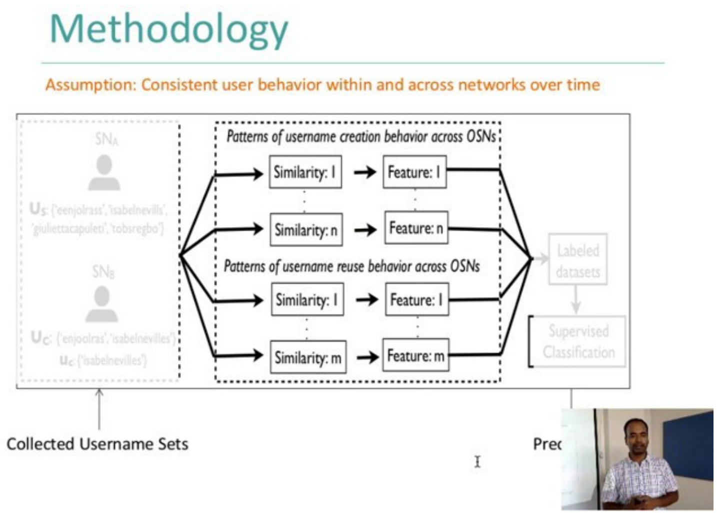
- **User ID**: 595929421
- **Past usernames on Twitter**:
  - ["**bigeasye_**", "reezy11_", "epiceric_", "soulanola", "**swampson_**", "**hebetheeeric**", "swampkidd_"]
- **Past Usernames on Instagram:**
  - ["bigeasye_", "**epiceric17**", "swampson", "hebetheeeric"]}

So, here is an example of users whose accounts have changed, whose handles have changed. User ids, as I said before, 24, in this case is 595929421, that is the handle. That isthe user id. Wekeep like a track of that, and if you look at the names that this handle has changed it, when it from bigeasye underscore, to reezy11, to epiceric underscore, to something else, to swampson x y, swamkidd, right. It changed 1 2 3 4 5 6 7 times, it has actually changed the handles.

And if you look at the same user account in Instagram, this seems to also changed four times, but there is a connection between the users handles that the person had in Twitter and then Instagram also. Right, you can actually look at this to also make sense,oh, that thereis this swampsoninInstagram,then swampsonunderscore inTwitter,is itthesame person? Like, for example, I am sure many of you were, who have common names, for example, Ponnuragan Kumaraguru is not so common, so, if you want to create anaccount probably you are the only one, you can get the handle, but let's take Shristi Gupta it is so common.

That if you want to create an account now any of the social networks, you are not goingto get (Refer Time: 28:21) Shristi Gupta. So, you are probably going to get Shristi Gupta 123,SonuGupta246,19o7,things likethat.This was,therefore;youcanactuallyuse this information also, that some parts of the handle is very similar, so, are they the same people?You could usethis Jacqard'sdistance, and there aremany other - Editdistance– there are many other measures by which you can find out whether the, how far is the handle from each other can be also used to say whether it is the same person.

(ReferSlideTime:28:59)



Methodology

Assumption: Consistent user behavior within and across networks over time

SNA

Us: {'eenjolrass', 'isabelnevills', 'giuliettacapuleti', 'tobsregbo'}

SNB

Uc: {'enjoolras', 'isabelnevilles'}

uc: {'isabelnevilles'}

Patterns of username creation behavior across OSNs

Similarity: l → Feature: l

Similarity: n → Feature: n

Patterns of username reuse behavior across OSNs

Similarity: l → Feature: l

Similarity: m → Feature: m

Labeled datasets

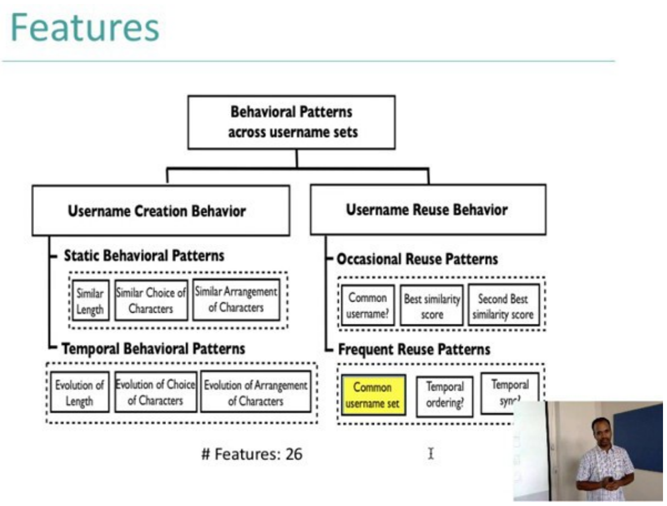
Supervised Classification

Collected Username Sets

Prec

So, here is one version of the same slide that I showed you earlier, which is, usernames are collected, which is what we discussed now. Now we look at some features that you can actually use to put them together and then we'll find out what the predictions are. This is the same slide that I had about 5 or 8 slides before on the whole process of actually identifying whether the handles are same.

(ReferSlideTime:29:33)



Features

Behavioral Patterns across username sets

Username Creation Behavior

- Static Behavioral Patterns
  - Similar Length
  - Similar Choice of Characters
  - Similar Arrangement of Characters
- Temporal Behavioral Patterns
  - Evolution of Length
  - Evolution of Choice of Characters
  - Evolution of Arrangement of Characters

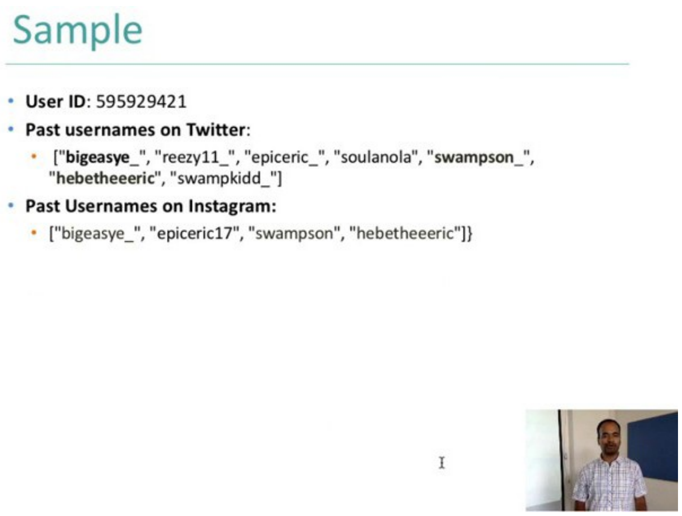
Username Reuse Behavior

- Occasional Reuse Patterns
  - Common username?
  - Best similarity score
  - Second Best similarity score
- Frequent Reuse Patterns
  - Common username set
  - Temporal ordering?
  - Temporal sync?

Features: 26

So, here these sets of features that probably actually used in our work. Usernamecreationbehavior,she,shejust(ReferTime:29:46)categorizedthefeaturesintodifferent buckets. Similar length in terms of username, similar choice of characters, similar arrangement of characters ponguru in Twitter and tongorea 24 in LinkedIn and temporal behavioral feature also, evolution of length, I started with 6, now at 7, now it is 8, what kind of characters are changing, evolution of choice of characters.

(ReferSlideTime:30:22)

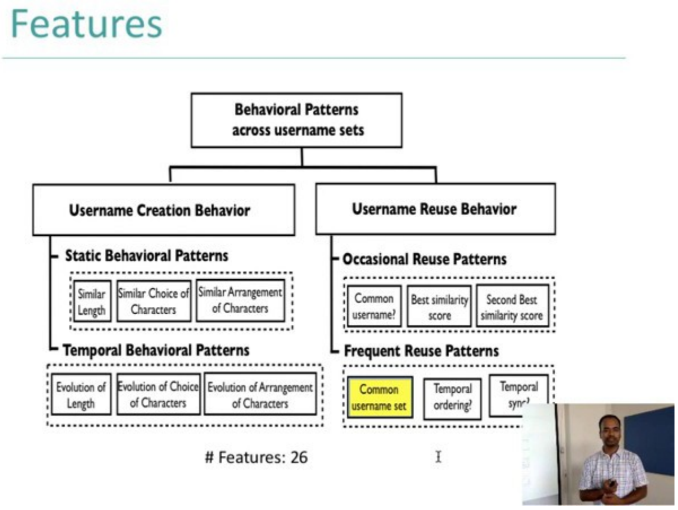


Sample

- **User ID**: 595929421
- **Past usernames on Twitter**:
  - ["**bigeasye_**", "reezy11_", "epiceric_", "soulanola", "**swampson_**", "**hebetheeeric**", "swampkidd_"]
- **Past Usernames on Instagram:**
  - ["bigeasye_", "epiceric17", "swampson", "hebetheeeric"]}

If you see here, all these features, some of these features can be discussed here. If you look at the account details then probably this user started, both Twitter and Instagram is just the same, and then after some point in time the person had epiceric (Refer Time: 30:39) the third in Twitter is the same as the second one in Instagram. Fifth in Twitter is very similar to the third in Instagram. So, you can find this evolution and make something out of this also. Occasional reuse patterns.

(ReferSlideTime:31:01)



Features

Behavioral Patterns across username sets

Username Creation Behavior

Static Behavioral Patterns

Similar Length

Similar Choice of Characters

Similar Arrangement of Characters

Temporal Behavioral Patterns

Evolution of Length

Evolution of Choice of Characters

Evolution of Arrangement of Characters

Username Reuse Behavior

Occasional Reuse Patterns

Common username?

Best similarity score

Second Best similarity score

Frequent Reuse Patterns

Common username set

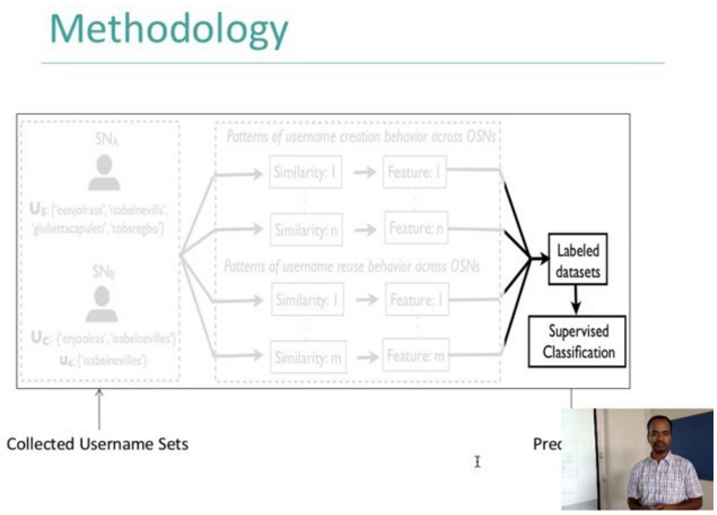
Temporal ordering?

Temporal syn

Features: 26

So, common username, the same username being used and features, temporal ordering again,characters,howtheyareplaced;youcanuseallofthesefeatures,whichis, Paridhiis calling it more as the behavioral patterns across usernames.You can use these features to say whether the handles are same and the number of features sets that we had was about 56. And if you remember the account, if you remember the trust and credibility section, which is I think week 1 or week 2 that we saw,then we actually saw 45 features in TweetCred and in trust content in Twitter, we saw about 45 features thatAditi used in terms of actually finding out whether this particular content that is posted on Twitter is credible or not (Refer Time: 31:59).

(ReferSlideTime:32:01)

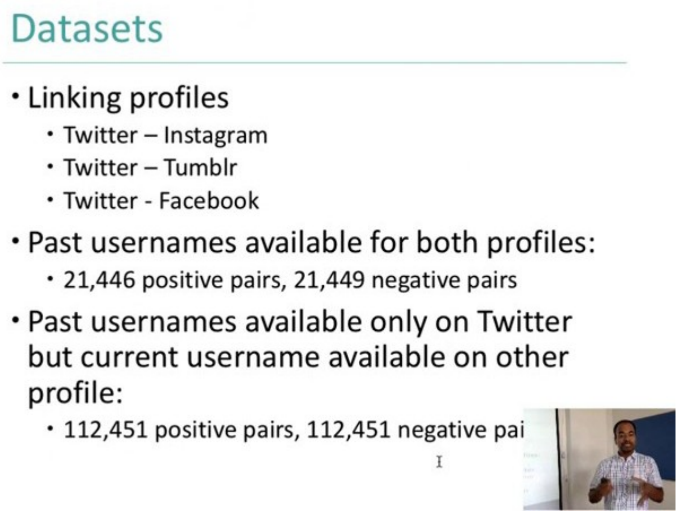


Methodology

SNA
Us: ('eenjolrass','isabelnevills',
'giuliettacapuleti','tobsregbo')
SNB
Uc: ('enjoolras','isabelnevilles')
uc: ('isabelnevilles')
Patterns of username creation behavior across OSNs
Similarity: l
Feature: l
Similarity: n
Feature: n
Patterns of username reuse behavior across OSNs
Similarity: l
Feature: l
Similarity: m
Feature: m
Labeled datasets
Supervised Classification
Collected Username Sets
Prec

So, now, we have the details from the users, details from the handles, what are the changes in the features, and we can all, we can put them all together to create a set of users, candidates sets, so to call, and then actually make a judgment, give the, give the output as, here is a probability of Sonu Gupta and Sonu Gupta 1 2 3 being the same is 0.9,versus Sonu Gupta and Sonu Gupta 0917 being the probabilityis about 0.4. So, you can actually make that output, that is the last part.

(ReferSlideTime:32:44)

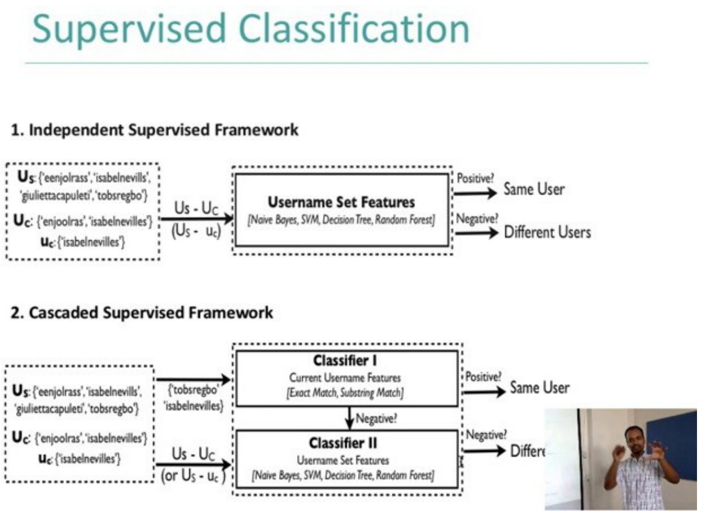


Datasets

- Linking profiles
  - Twitter – Instagram
  - Twitter – Tumblr
  - Twitter - Facebook
- Past usernames available for both profiles:
  - 21,446 positive pairs, 21,449 negative pairs
- Past usernames available only on Twitter but current username available on other profile:
  - 112,451 positive pairs, 112,451 negative pai

Now, details. So until now it is more theoretical about how this could be done. Now let just look at specific things what Parishi did in terms of actually getting the data from multiple social networks and finding out how much we can actually do well. So, let us look at some specific examples. In this case we, are looking at data collected between Twitter and Instagram, Twitter and Tumblr,Twitter and Facebook. Past usernames were collected. 21,000 positive pairs, which is, details that collected from these social networks, and about past usernames available only on Twitter, but current usernames available on other profiles is about 140,000. So, essentially the idea is that the data was collected between multiples social networks of the current and the past user handles and how this was put together, and what kind of mechanisms was used to find out, whether these handles are same.

(ReferSlideTime:33:54)



Supervised Classification

1. Independent Supervised Framework

U_S: {'eenjolrass', 'isabelnevills', 'giuliettacapuleti', 'tobsregbo'}
U_C: {'enjoolras', 'isabelnevilles'}
u_c: {'isabelnevilles'}

U_S - U_C
(U_S - u_c)

Username Set Features
[Naive Bayes, SVM, Decision Tree, Random Forest]

Positive? Same User
Negative? Different Users

2. Cascaded Supervised Framework

U_S: {'eenjolrass', 'isabelnevills', 'giuliettacapuleti', 'tobsregbo'}
U_C: {'enjoolras', 'isabelnevilles'}
u_c: {'isabelnevilles'}

{'tobsregbo' 'isabelnevilles}

U_S - U_C
(or U_S - u_c)

Classifier I
Current Username Features
[Exact Match, Substring Match]

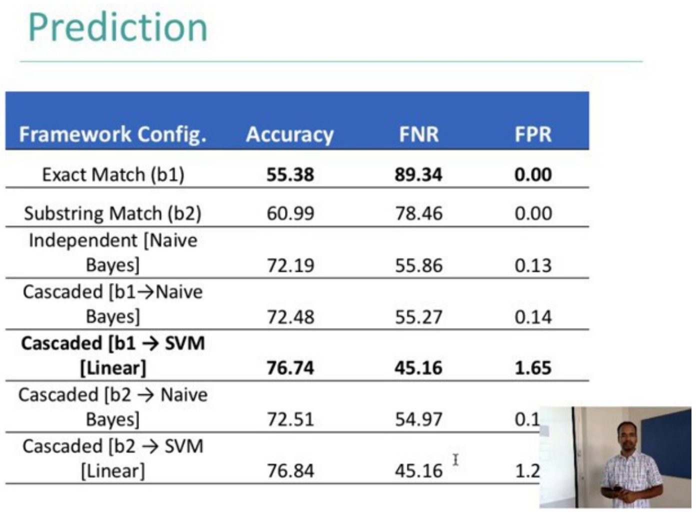
Negative?

Classifier II
Username Set Features
[Naive Bayes, SVM, Decision Tree, Random Forest]

Positive? Same User
Negative? Differe

And it is the same diagram that I showed you before, take the handles, understand some features, put the features together, and create the score. And that is what is done here - two methods are done, oneis you just do only the features, and the other method that we did was, do a classifier and then apply it and to find out whether it is the same user.

(ReferSlideTime:34:22)



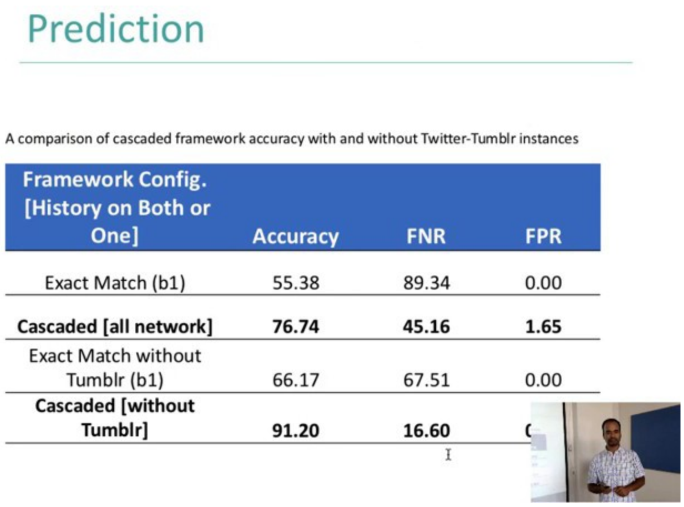
Prediction

Framework Config.	Accuracy	FNR	FPR
Exact Match (b1)	**55.38**	**89.34**	**0.00**
Substring Match (b2)	60.99	78.46	0.00
Independent [Naive Bayes]	72.19	55.86	0.13
Cascaded [b1→Naive Bayes]	72.48	55.27	0.14
Cascaded [b1 → SVM [Linear]	**76.74**	**45.16**	**1.65**
Cascaded [b2 → Naive Bayes]	72.51	54.97	0.1
Cascaded [b2 → SVM [Linear]	76.84	45.16	1.2

Herearethedifferentmethodsthatwasused,whichis,exactmatch,right,exactmatchof thehandles,substringmatch,andthenclassifiers,differentclassifierswereappliedandif youlookatthefirstoneandthefifthone,aretheoneswhichhadthemaximumaccuracy, whichbasicallysays thatif youlookatthehandles,thewaythatthehandleslook similar,

,and iftheyare exactmatch,is veryhigh probabilitythattheyarethe sameusers,that is, veryless probability,that theywould actually be different users.And if we use the SVM classifier and then apply it on to find out whether there is same users using all the, using all the 26 features that we talked about, there is high probability that will be able to, about 76 percent is the accuracy to find out whether they are actually the same handles.

(ReferSlideTime:35:29)



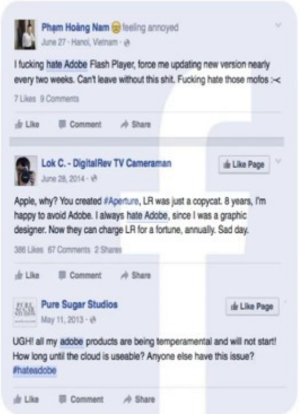
Prediction

A comparison of cascaded framework accuracy with and without Twitter-Tumblr instances

Framework Config. [History on Both or One]	Accuracy	FNR	FPR
Exact Match (b1)	55.38	89.34	0.00
Cascaded [all network]	**76.74**	**45.16**	**1.65**
Exact Match without Tumblr (b1)	66.17	67.51	0.00
Cascaded [without Tumblr]	**91.20**	**16.60**	**(**

(ReferSlideTime:35:31)

Measuring Volume of Sentiments


Phạm Hoàng Nam feeling annoyed
June 27 · Hanoi, Vietnam ·

I fucking hate Adobe Flash Player, force me updating new version nearly every two weeks. Can't leave without this shit. Fucking hate those mofos :-<

7 Likes 9 Comments

Lok C. - DigitalRev TV Cameraman
June 28, 2014 ·

Apple, why? You created #Aperture, LR was just a copycat. 8 years, I'm happy to avoid Adobe. I always hate Adobe, since I was a graphic designer. Now they can charge LR for a fortune, annually. Sad day.

386 Likes 67 Comments 2 Shares

Pure Sugar Studios
May 11, 2013 ·

UGH! all my adobe products are being temperamental and will not start! How long until the cloud is useable? Anyone else have this issue? #hateadobe

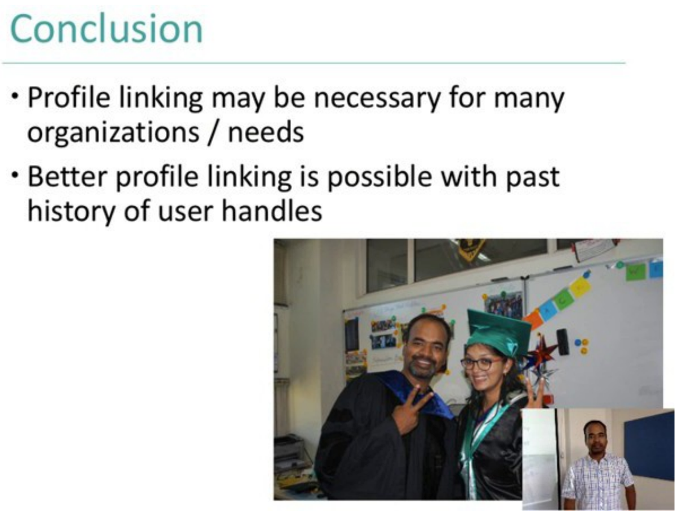
So, let me show you some, why, so, if you rememberthe motivation that I started off with is this sending this advertisement to people, I do not want to waste my money. Just let us go back and connect to the motivation also. There are, if PK or NPTEL wants to send out advertisement to all the students who are on Facebook and Twitter, we have to send this information to onlyone user onlyonce, you do not want to duplicate and waste money, that is the motivation I started, but the motivation can be many other things. There are some examples also.

So, in this case if you see, the sentiments of the user for a, let us take any topic that you take, the one on the left is on Facebook, the one on the right is on Twitter, you really want to understand whether sentiment, for example, you just look at this, you want to understand whether the sentiments of these people are expressed on Facebook and Twitter, are they same, and if they are same or if they are different, are they are same user.

So that, you can actuallymeasure that the negative sentiment of anytopic is nota sum of all negative sentiments in all social networks, but only the unique people that you wantto take a note of, (Refer Time: 36:52) right. Because if I say something positive in Facebook,andIamsamepersonwhoissayingpositiveinTwitter,itisnot,ifyou,you cannot measure the positivity as twice, but it is only once because it is only one persons sentiment, right. So, that is another motivation. The other motivation also that, the other reason why this identity resolution is an interesting problem is because you can actually look at, even law enforcement can actually use this. Which is, somebody uploaded a malicious video on YouTube and in there is another handle, which uploaded the same video on Twitter.

Now I want to find out whether it is the same person who is posting it. Somebody is actually speaking against some people or some organization on Twitter. And there is another handle which is speaking against some persons or some organization in a different network. I want to know whether it is the same person, because they do not want to be wasting the time in assuming that it'stwo people and wasting time in finding two people, but it is only one person that they have to chase and catch, right. So, that is the motivation, that is also another motivation to actually find out whether these two handles are same or not, right. So, there is very interesting motivation for doing this work, and there is a lot of interesting things one could actually try out.

(ReferSlideTime:38:29)



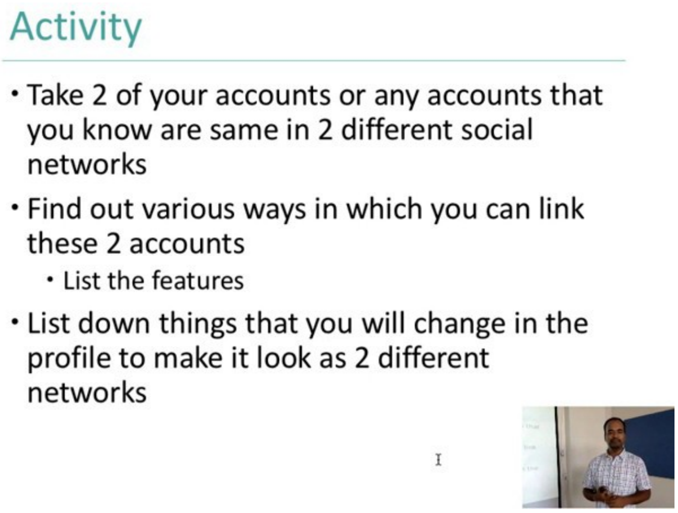
Conclusion

- Profile linking may be necessary for many organizations / needs
- Better profile linking is possible with past history of user handles

So, here the last take away from this part, which is profile linking may be necessary for manyorganizationsasthequestionsthatwesaid,Idonotwanttowastemymoney,I want to actually understand, whether,how many people have posted, what is the volume of actually positivity or negativity, or I want to find out who is actually speaking online and to link users.

And the conclusions from this work are that essentially you do not have to only bank on the current handles that people have, current information that people have; even usingtheonesthatfromthepastcanactuallyprovetheefficiency,accuracy,ofthe–andthatis Paridhi, who is in the picture, who just graduated with a PhD from the work that she did on this topic.

(ReferSlideTime:39:29)



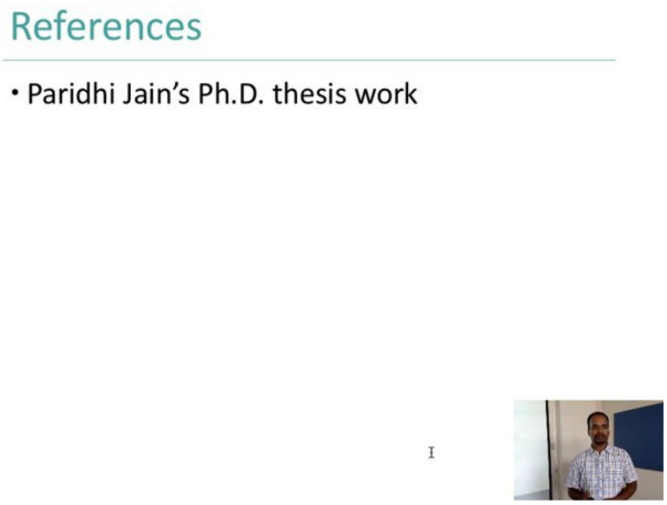
Activity

- Take 2 of your accounts or any accounts that you know are same in 2 different social networks
- Find out various ways in which you can link these 2 accounts
  - List the features
- List down things that you will change in the profile to make it look as 2 different networks

I am going leave you for this week just to try this out. We can probably connect this to the quiz that we have or homework that we have for the course also, but here is what I want to try,you to try.Take two of your accounts on your two different social networks, which is Facebook and Twitter. Let’s just stick to only Facebook and Twitter. Just take these twohandles,list downall thethings thatyou can actuallydo,findoutvarious ways in which you can actually link these two accounts. List the features, features just we talked about, right? These ones or there could be many others also. List down these, and list down things that you will change in the profile to make them look two different account networks also. You could do both ways.

You could do list down things that you will change to make it two different account, or list down things that you will do to make it the same account. Right? Share it in the forum, let’s see what you people actually come up with. I hope the activity is clear. It is that take your Facebook account, take your, take your Twitter account, list down the features that are available, that you think you can actually connect with the accounts.And list - that is the first output - second output is list down all the things that you willdo to make it look same - that is the second account, second output. The third output is, list down all the things that you will do which will make that it is two different accounts.

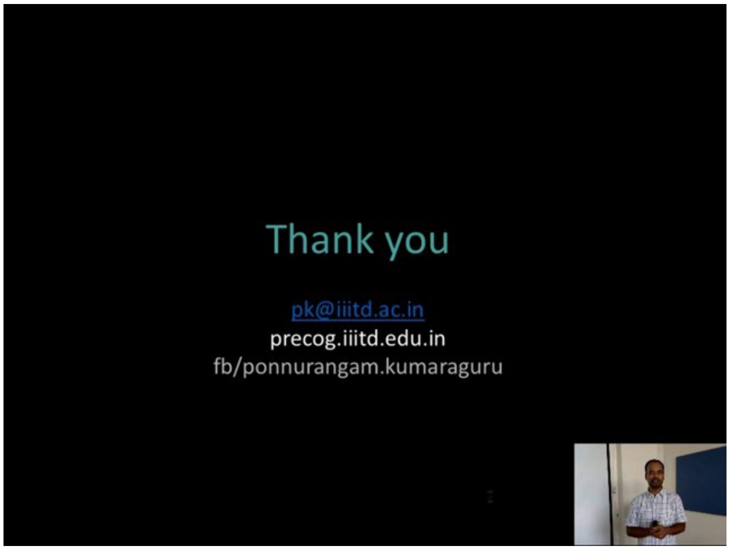
(ReferSlideTime:41:16)



References

- Paridhi Jain’s Ph.D. thesis work

(ReferSlideTime:41:19)



Thank you

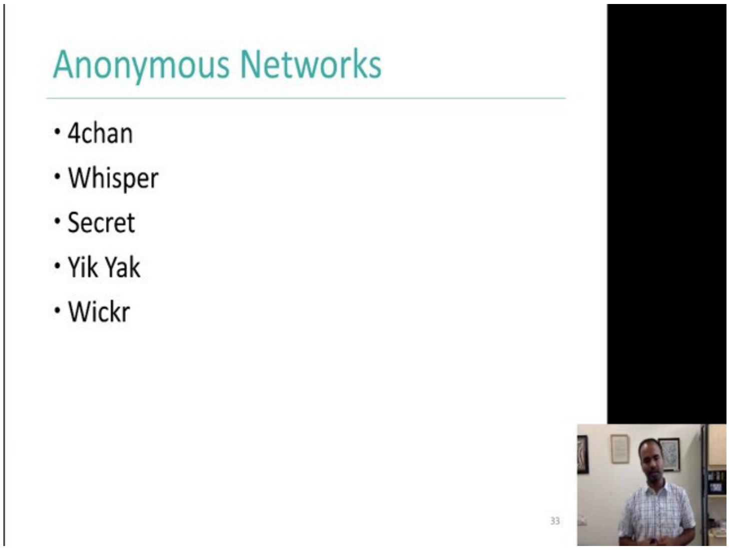
pk@iiitd.ac.in
precog.iiitd.edu.in
fb/ponnurangam.kumaraguru

So, that is all I had for this week that this 8.1. I will actually continue in a different topic when I start off with 8.2.

AnonymousNetworks

Welcome back to the Privacy and Security in Online Social Media course on NPTEL. This is week 8, and this is the second part of the week.

(ReferSlideTime:00:25)



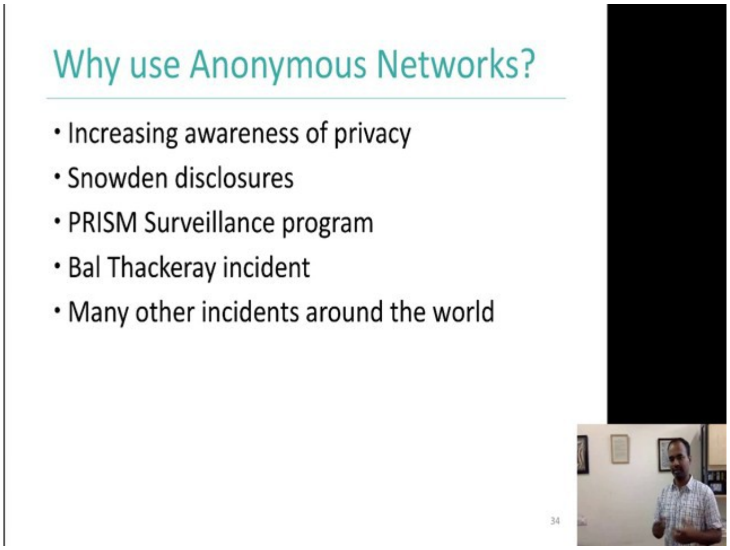
Anonymous Networks

- 4chan
- Whisper
- Secret
- Yik Yak
- Wickr

So, until now, the social networks that we are seeing is generally popular networks like Facebook, Twitter and these are called online social networks.And particularly we have also looked at Foursquare, which is a location based social network. Then I think briefly we have also talked about ephemeral social networks, which are networks where the contents that is getting generated can be actually removed after some period of time, where the contents are ephemeral, which it is like a snapshot network; where you post some content and after sometimes that content get's deleted right.

What we are going to look at this part of the lecture is something called anonymous network. Anonymous networks are networks, where it is not clearly visible or it is not possible to find out who is actually posting the content. So, we will go in detail about whatanonymousnetworksarewithsomeexamplesandIwillalsoshowyousome research done, some work done, on finding out how anonymous network behaves,compared to normal networks like facebook or twitter. Some examples of anonymous networks are 4chan, Whisper, Secret, Yik Yak, Wickr, these are the different types of anonymous social network, there are many, there are many such networks that are available there here is only a small list.

(ReferSlideTime:02:10)



Why use Anonymous Networks?

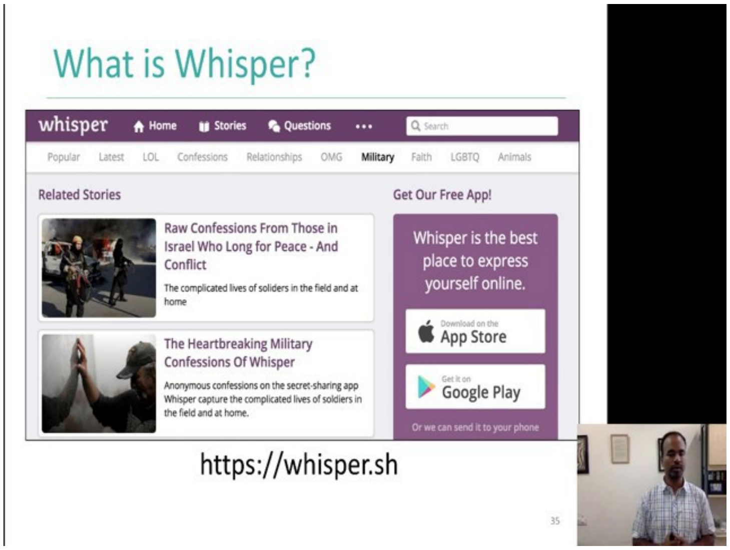
- Increasing awareness of privacy
- Snowden disclosures
- PRISM Surveillance program
- Bal Thackeray incident
- Many other incidents around the world

34

Whydoyouneedanonymoussocialnetwork?So,wealreadyhaveFacebook,wealready have twitter, why you might need a network or network of the category of anonymous network or network that gives the preference or gives the facility for having anonymity. Increasing awareness of privacy, so people are getting to know more and more about privacy,people are getting to or people want to have more privacy on online social networks. So, therefore people are looking for networks, that will give more anonymity. And there werealso incidences like Snowden; projects like PRISM were the information that ispubliclyavailable ortheinformation that isavailable totheseorganizations canbe used for other reasons also.

And of course, there is an incident in India, where the some post was done and that post called actually viral and there were consequences of the post also. So, therefore many many incidences around the world, which are happening, which is expecting, which is making users who use social networks expect more privacy, expect anonymity in the networks.Becausefor example,IfI doapostonfacebook,ifI doa poston twitteritis actually very clear that it is pk ponnurangam dot kumaraguru dot or ponguru in twitter is actually doing the post. In fact, if I wanted to say something on social networks, but I do not want to be attributed to the post then I would actually use these anonymous social network.

(ReferSlideTime:04:13)



What is Whisper?

Raw Confessions From Those in Israel Who Long for Peace - And Conflict

The complicated lives of soliders in the field and at home

The Heartbreaking Military Confessions Of Whisper

Anonymous confessions on the secret-sharing app Whisper capture the complicated lives of soldiers in the field and at home.

Whisper is the best place to express yourself online.

Or we can send it to your phone

https://whisper.sh

35

So, here is thatslide screen shotof the website whisper dot sh in thecontent is organized asinthetoporderinthisimage popular,latest,Lol,confessions,relationship,Ohmygod and they create these categories so that the content that is uploaded on whisper gets into1 of these categories and the URLis whisper dot sh. I'll let you to actually play around a little on of the website, create an account and see how the accounts work.

(ReferSlideTime:04:55)

What is Whisper?

https://www.youtube.com/watch?v=pX9I9kR2tTc

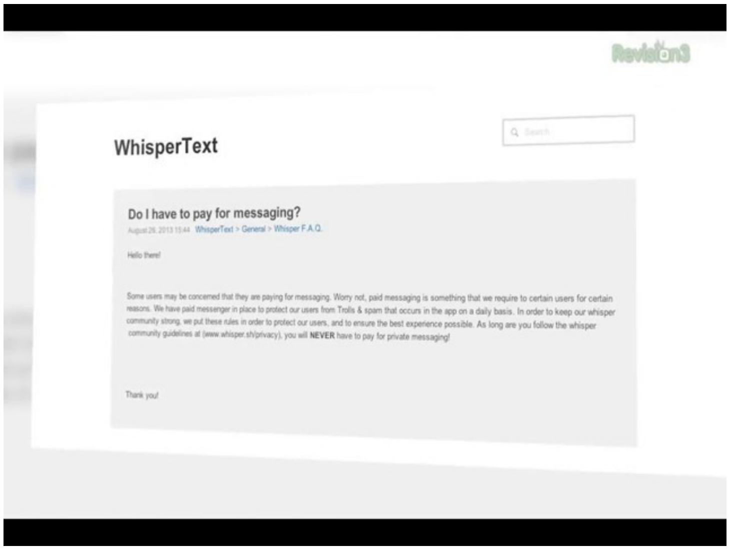
Here is a URL, here is a video, which describes some features of YouTube. We take a lookatvideonowandthenIwilldescribesomedetailswhichisfromthevideoandother features of whisper.

(ReferSlideTime:05:22)

Secret,s lies and plenty of spam on this super popular mobile app - whisper.Whisper is a confessional app that encourages you to post your secrets behind a screen of anonymity. You know kinda like that other website - Post Secret, where in users can submit those deepdarksecretstheywouldnoteventelltheirbestfriend,likeIsecretlytooknude pictures of my best friend. But whisper, which been around for a couple years and it is being steadily gaining popularity is just as much way to people to connect around us on flattering,embarrassing,tabooorsometimesdisturbingconfessionorsometimesnoneof those things pretty often.

Whisper does not harvest your email or contacts and screen names are less prominent within the app you can also change it whenever you want along with the pin that takes place of the password. So, there's definitely increased premium on anonymity than most social networks. As per your deep dark secrets, those you can post by hitting the plus. Type your whisper and it will auto generate a stock photo to go along with it. So, yes is not just a confessional, but it is meme generator which you can share via email, SMS or social.

(ReferSlideTime:06:18)



Do I have to pay for messaging?

August 26, 2013 15:44 WhisperText > General > Whisper F.A.Q.

Hello there!

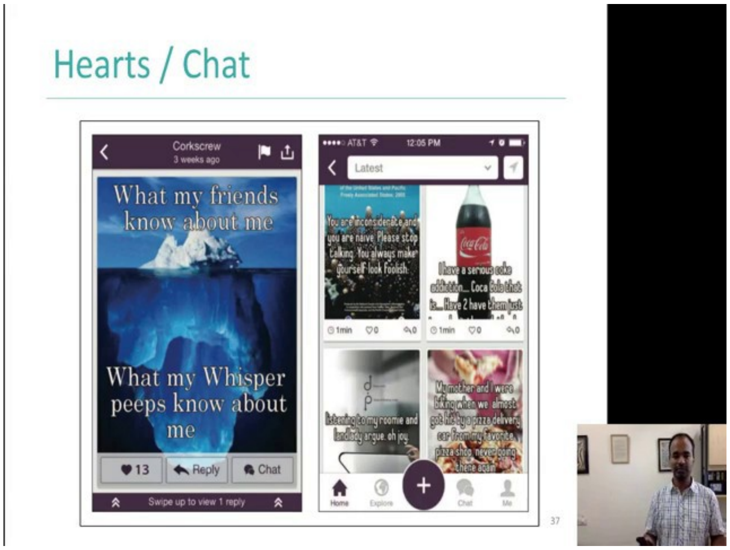
Some users may be concerned that they are paying for messaging. Worry not, paid messaging is something that we require to certain users for certain reasons. We have paid messenger in place to protect our users from Trolls & spam that occurs in the app on a daily basis. In order to keep our whisper community strong, we put these rules in order to protect our users, and to ensure the best experience possible. As long are you follow the whisper community guidelines at (www.whisper.sh/privacy), you will **NEVER** have to pay for private messaging!

Thank you!

And you can private message, there is premium messaging for certain users, trolls and spammers according to the F.A.Q, but for everyone else it is free as long as you play by the rules.Forthat reasonyou have all seen posts like these, alot ofthem peopletrying to hookup or these - message your favorite singer and if you are dumb enough to fall forthat,you areprobablyundertheageof10andgodhelpyou,butwhereverthereissecrets there are also bad apples and sometimes lies.

In September for example, someone posted a supposed murder confession on PostSecret, promptingafrenziedreddit searchfortheselfprofessedcriminal.Andtilldateno crimehas ever been found associated with that post. And this week in Arizona, a cop was arrested for having sex with a minor he met on whisper app after she posted that she wanted to get pregnant. But considering the allure of posting secrets and knowing other people's, it's unlike to slow down whisper for now.As always you can let me know what you think I am onTwitter,Facebook, Google orVK on anniegaus and you can get a free netflix trial with a signup on Netflix dot com slash wtbd, thanks for watching.

(ReferSlideTime:07:16)

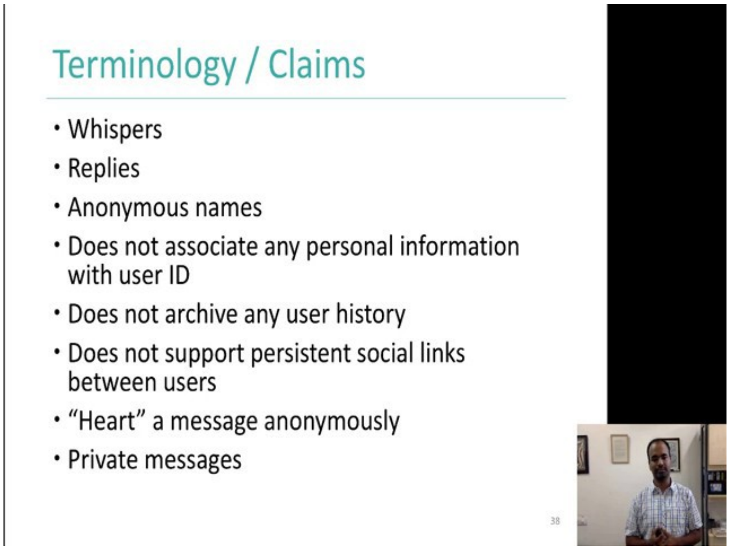


Hearts / Chat

Corkscrew
3 weeks ago
What my friends know about me
What my Whisper peeps know about me
13
Reply
Chat
Swipe up to view 1 reply
AT&T
12:05 PM
Latest
You are inconsiderate and you are naive. Please stop talking. You always make yourself look foolish.
1min
0
0
I have a serious coke addiction... Coca Cola that is... Have 2 have them just
1min
0
0
listening to my roomie and landlady argue. oh joy
My mother and I were biking when we almost got hit by a pizza delivery car from my favorite pizza shop. never going there again
Home
Explore
Chat
Me
37

Now, that we are seeing the video, the video actually talks about how whisper is being used, what kind of users get on whisper and what kind posts they do and how whisper actually works in creating some content, it actually gets merged ontoimages andyou get posts and creating memes in other terms. So, the way that people react to the post on whisper is by hearts and also you can chats on the post that you make. Again please remember all of this is going to be anonymous.

(ReferSlideTime:08:02)



Terminology / Claims

- Whispers
- Replies
- Anonymous names
- Does not associate any personal information with user ID
- Does not archive any user history
- Does not support persistent social links between users
- "Heart" a message anonymously
- Private messages

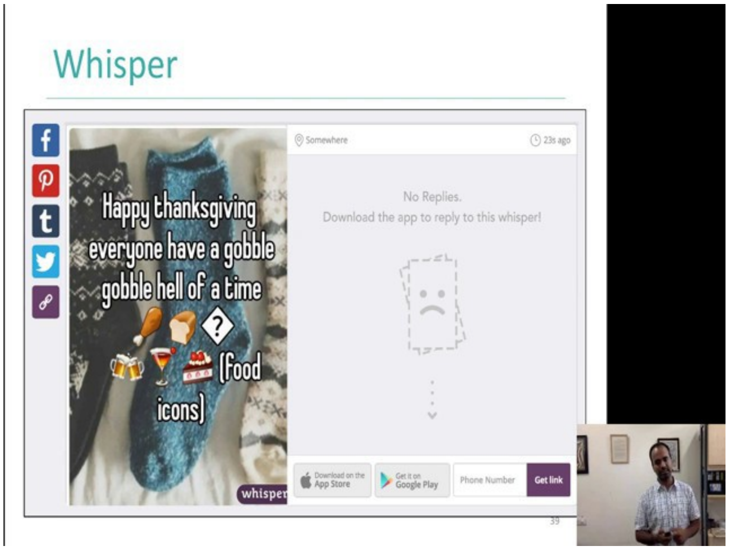
38

So, terminologies that we'll see to understand the rest of the lecture, we need to understand some terminologies, whispers or the posts, replies or I do a post and you are actually replying like a comment in Facebook or a reply in Twitter. And the posts are anonymous you really do not get to see it is ponguru . I may have an account, which is called professor, teaching computer science or anything that I wanted to keep that is the username and interestingly whisper also allows you to back with probably have seen video also, whisper also allows you to change the usernames as anonymous as you want and more number of times also. So, that makes it much more difficult to go back andlook at the person who posted the content.

And whisper does not associated any personal information of the user id, it is not collecting any information and does not archive any user history, which at least that'swhat they claim, it does not support persistent social links between users. The person who hearts at that, the person who replies it, the links of the users are not kept, where as if you remember the homework and thequestionsthat you haveseen in thepastwherein the context of facebook or twitter.

The content for all the relationship between the users are stored as a graph and you can analyze those graph, also retrieving the graph from twitter or facebook and use these graph to make some inferences. Heart a message anonymously may also use just in (ReferTime:10:02).AheartisbasicallytheonethatIshowedyouintheslide,likethe like in facebook. If in the private messages against or this in the video that I had a few minutes before, which showed private messages also you can actually post private messages between the users.

(ReferSlideTime:10:28)



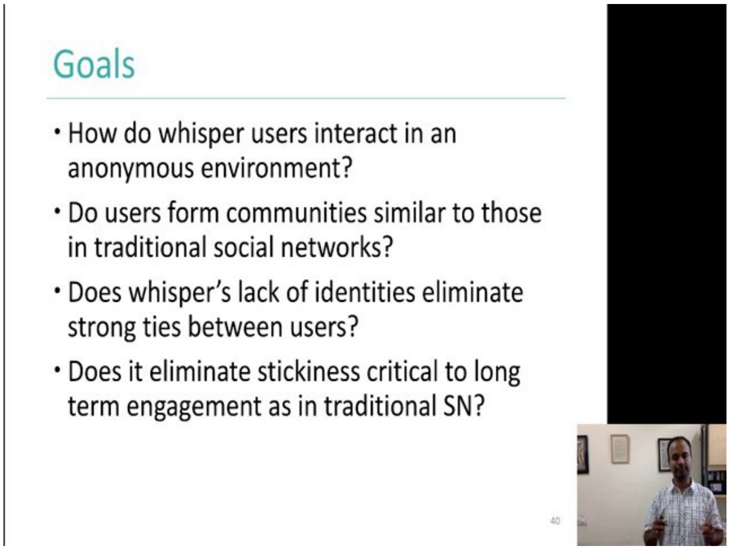
Whisper

Happy thanksgiving everyone have a gobble gobble hell of a time 🍗🍞❓🍻🍸🎂 (food icons)

No Replies.
Download the app to reply to this whisper!

Thatisthescreenshotfromwhisper.

(Refer Slide Time: 10:36)



Goals

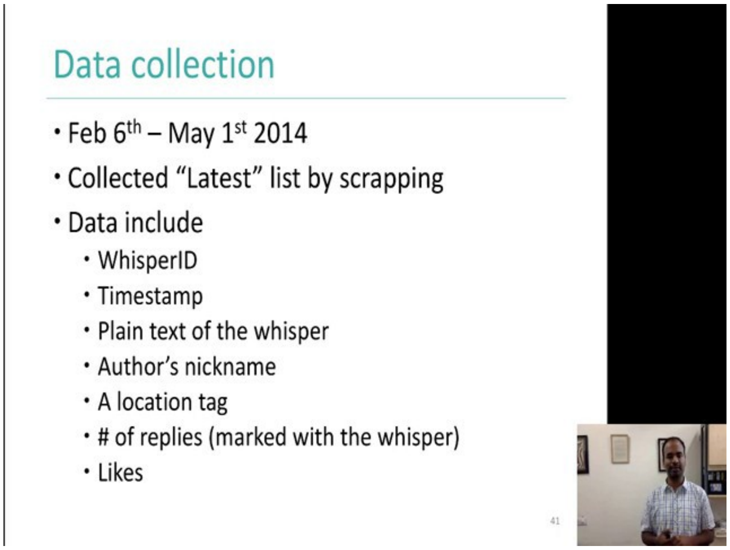
- How do whisper users interact in an anonymous environment?
- Do users form communities similar to those in traditional social networks?
- Does whisper's lack of identities eliminate strong ties between users?
- Does it eliminate stickiness critical to long term engagement as in traditional SN?

So, what we are going to look at is we are going to try an answer these four questions. How do whisper users interact in an anonymous environment, how is the interactions on whisper?Dousersformcommunitiessimilartothoseintraditionalnetworks,likefor examplepeopleinteractiononfacebook,howisthisdifferentfrompeopleinteractionson whisper or twitter. Does whisper's lack of identities eliminate strong ties between users, which is if I do not have strong ties which is if you do not know that PK is talking toyou, does it eliminate the strong relationship that you and I would have. Let's take bothof are on whisper,you doyour postand I come react toit, Ido areplyto it, if you donot know it is PK, who is the faculty at IIIT Delhi or some profile that I have, if you do not knowthatitis me,willyou continuetalkingtome? Isthereisastrongerrelationship that happens.

For example, you could also see in twitter or in facebook that some people are very stronglyconnected. Forexample, ifIdoanypostthataresome setsofpeoplewhowould always like it, who would always accurately make a comment or reply or retweet. So, thosethatarebasicallycalled strongtiesandthatdoesitexistson whisperisthequestion we have to look at.

Now also whisper, because of being anonymous does it eliminates stickiness critical to long term engagement. Stickiness is basically is a factor bywhich you are actually gluedon to the network, more and more people get connected to it or single person is actually spending more time on the network. I know that is clear those are the four goals that we have for the rest of this lecture where we will take one click one particular network inthis case whisper, we will actually try an answer for these four questions using somedata, using some inferences that we draw.

(ReferSlideTime:13:14)

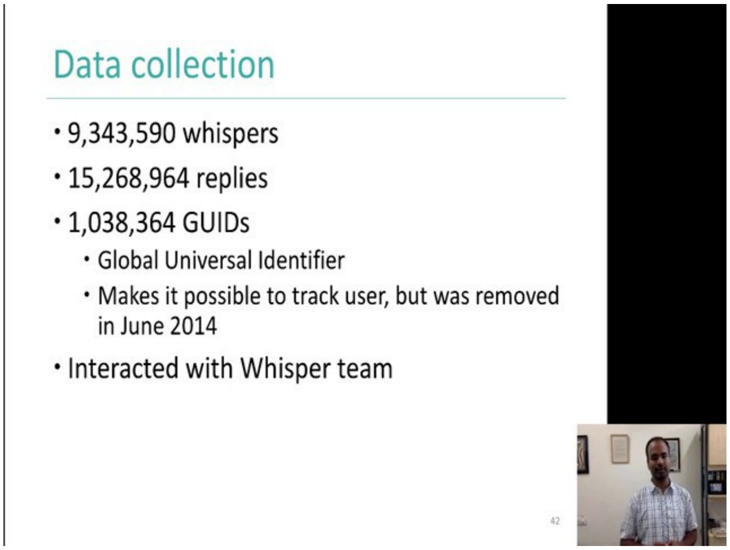


Data collection

- Feb 6th – May 1st 2014
- Collected "Latest" list by scrapping
- Data include
  - WhisperID
  - Timestamp
  - Plain text of the whisper
  - Author's nickname
  - A location tag
  - # of replies (marked with the whisper)
  - Likes

So, data that was collected for doing this analysis is from 2014. And of course, that whisper does not have an apiso, data was scraped and what all they include. They included whisper id, which is like a post id, time stamp when the post was done, plain text of the whisper - thetext that was on thepicture, author's nickname which is the your handle, names so to say in the traditional sense, alocation tag if it was available, number of replies for the whisper and of course, the likes is the hearts that we talked about. So, that is clear simple to collection I think you all of you have seen this kind of data collection the past in all the networks that we have seen. Take it a network collect some basic data, do some analysis and answer interesting questions that actually makes sense.

(ReferSlideTime:14:26)

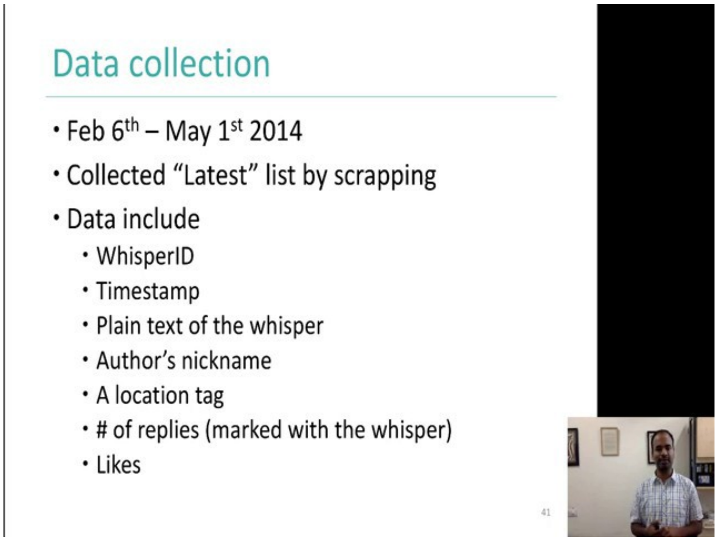


Data collection

- 9,343,590 whispers
- 15,268,964 replies
- 1,038,364 GUIDs
  - Global Universal Identifier
  - Makes it possible to track user, but was removed in June 2014
- Interacted with Whisper team

Data collection again, so 9 million whispers, 15 million replies 1 million GUIDs, which is global user universal identifier, which is the id for every user like you've seen in the twitter also.So,theusersgetoneuniqueidwhichiswhatwascollected. So,interestingly the team that worked on this work also interacted with the whisper team. Where they actually talked to them about the data that that they were collecting and about this universal identifier, which they were able to convince the whisper team thatusing this id you could actually go back and find out which user did what. So, the user GUID concept was removed on June 2014.

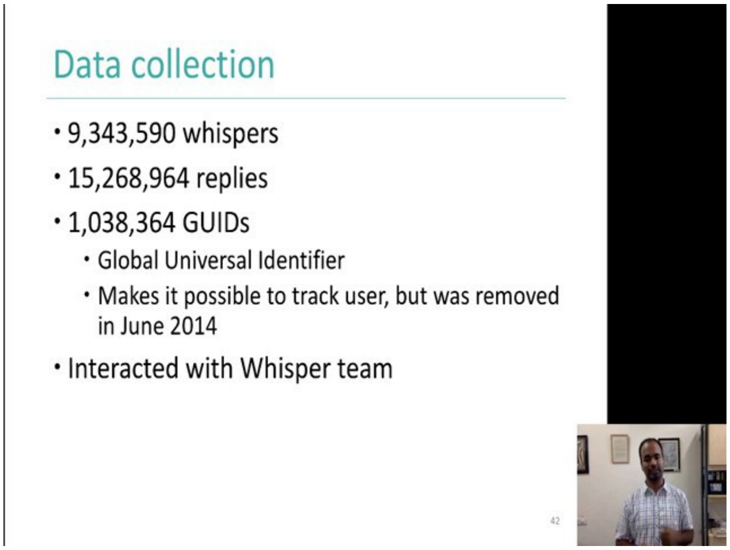
(ReferSlideTime:15:27)



Data collection

- Feb 6th – May 1st 2014
- Collected "Latest" list by scrapping
- Data include
  - WhisperID
  - Timestamp
  - Plain text of the whisper
  - Author's nickname
  - A location tag
  - # of replies (marked with the whisper)
  - Likes

(ReferSlideTime:15:34)

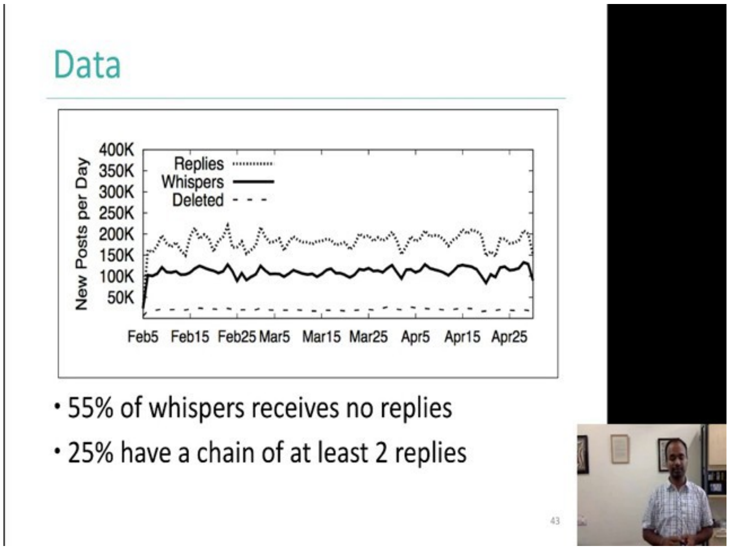


Data collection

- 9,343,590 whispers
- 15,268,964 replies
- 1,038,364 GUIDs
  - Global Universal Identifier
  - Makes it possible to track user, but was removed in June 2014
- Interacted with Whisper team

So, they collected the data, the researchers looked at what is going on whisper and then went and had discussion with whisper team to remove this. That for they actually wroteabout.

(ReferSlideTime:15:40)



Data

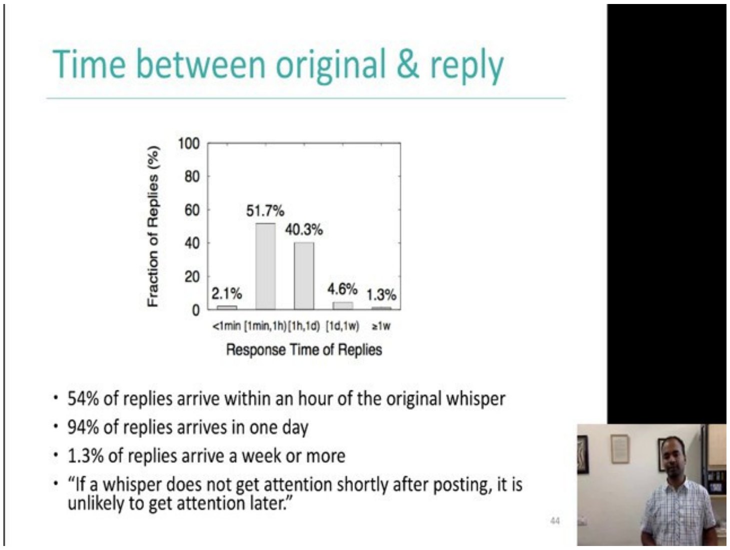
- 55% of whispers receives no replies
- 25% have a chain of at least 2 replies

So,now lookat analysis again, we havedone this inthe past soI'mgoingto goslowlyin terms of what analysis, first time using this data, what kind of inferences to be true.And allconnectedtothesefourquestionsthatwe have.Inthexaxissuchasthetime,itisthe time, inthis casebetweenFebruaryandMay,thatis wheretheycollected thedata.Andy axis is the number of posts per day.

And they actually look at 3 different types of posts, which is one as whisper, so to say what is content that is getting generated, one is the replies, which ishow many repliesare being posted for the particular whisper. And there is also third category of whispers being deleted. We get to this deletion later, which is also interesting problem, which is that, when in twitter also, we have more recent studies in 2016, people have seen that lot of content that are posted on the social network gets actually deleted for whateverreasons that the users are deciding to.

So, in this case, in whisper case 55 percent of whispers receives no replies, people just post content and nobody even replies to these posts. 25 percent have a chain of at least 2 replies.Only25percentsotosayactually weshouldreaditthatway,the25percenthave achain of atleast2 replies. 55 percentof thereplies, 55 percentof thewhispersdon'tgeta reply.

(ReferSlideTime:17:27)



Time between original & reply

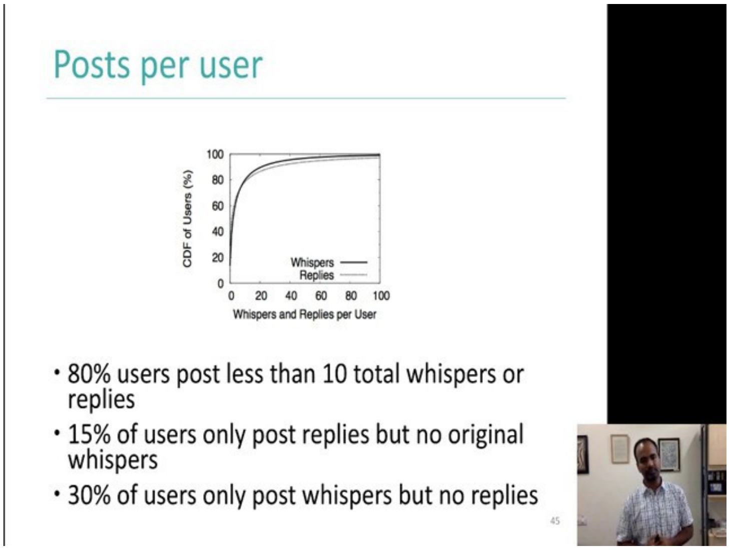
- 54% of replies arrive within an hour of the original whisper
- 94% of replies arrives in one day
- 1.3% of replies arrive a week or more
- "If a whisper does not get attention shortly after posting, it is unlikely to get attention later."

Time between original replies this is also interesting thing how quickly are theresponses towhisperthatisposted.Howquicklydothepeopleactuallylookatthepostthatisdone and how they reply and what they reply. In this case, we are only looking at the time we are not looking at the content. So, if your x axis is time again less than one minute, one minutetoonehour,onehourtooneday,onedaytooneweekandgreaterthan1week.

That is the x axis, y axis is fraction of replies, fraction of replies it shows you whatis the proportion of replies that the whisper gets. 54 percent of replies are within hour of those original whispers.

You can add the first two bars which is less than one minute and one minute to an hour, thiswillshowyou 54percent,54percentoftherepliesarrivewithinonehour,94percent within oneday.Basically, shows that if they do notget aresponsein one day they do not get it. More than half of them get a response within one hour.One point three percent of replies arrive within a week or more that is the last bar on the graph. So, essentially the conclusion is that if awhisper does notget attention shortlyafter posting, it is unlikelyto get attention later. Understandably, that because it is an anonymous people kind of post content, it gets little bit of attention and then dies off. This is similar to other networks also that we have seen.

(ReferSlideTime:07:16)



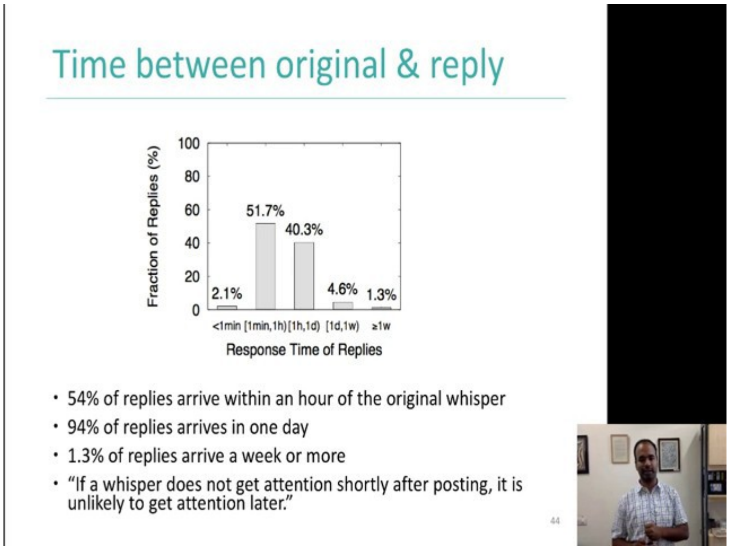
Posts per user

- 80% users post less than 10 total whispers or replies
- 15% of users only post replies but no original whispers
- 30% of users only post whispers but no replies

So,apostper user, which isjustthexaxis iswhispersandreplies peruser.Whispersand replies per user, which is how many times user is doing it and there is two lines in the graph, one the dark line, which is the whisper and the dotted line, which is the reply, they axis is the CDF you have seen many of the CDFs before cumulative frequency of user. Here it is basically showing that 80 percent of the users post less than 10 total whispers, which is actually pretty bad if you just look at the networks, 80 percent of the users post less than 10 total whispers and replies.

Which is if you flip it and see it is probably looking at 20 percent of the users areactually the people who are actually very active or in another sense less percentage of people are the ones who are actually doing maximum number of activities in thenetwork, which wehavealreadyseenin othernetworksalso. 15percentoftheusersonly post replies, but no original whispers, which basically again shows that less fraction of people posts replies that they do not create original content, only look at what users are doing and then they are replying to it.Thirty percent of the users only post whispers. But no replies, they're just the people who are creating the original content, but they actually do not reply to any of the content, reply or react to it.

(ReferSlideTime:21:32)



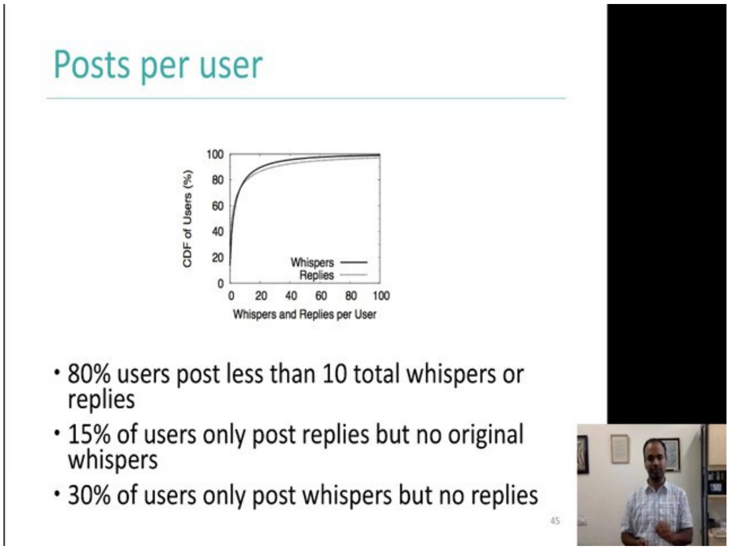
Time between original & reply

- 54% of replies arrive within an hour of the original whisper
- 94% of replies arrives in one day
- 1.3% of replies arrive a week or more
- "If a whisper does not get attention shortly after posting, it is unlikely to get attention later."

44

And that is clear, that is basically has two analysis that we saw, one is how much you attention is the content posted on whisper is getting.

(ReferSlideTime:21:44)

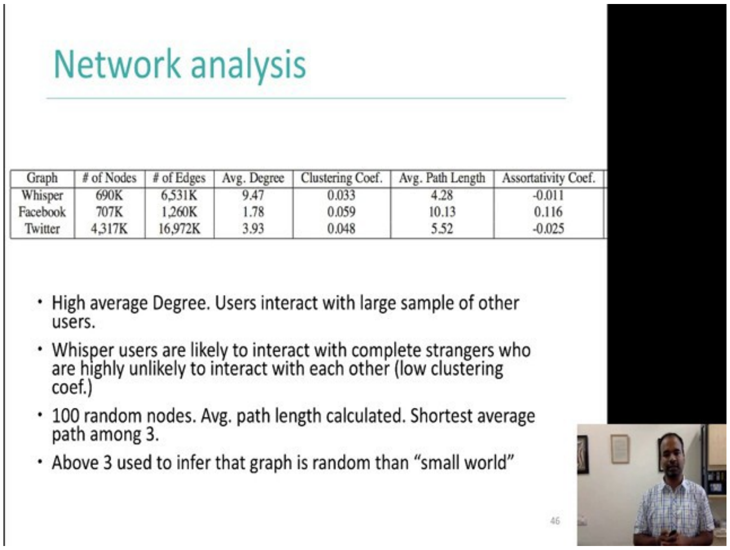


Posts per user

- 80% users post less than 10 total whispers or replies
- 15% of users only post replies but no original whispers
- 30% of users only post whispers but no replies

And how much time this it get take to get their attention and then what are the level of activity do users have on these networks.

(ReferSlideTime:21:51)



Network analysis

Graph	# of Nodes	# of Edges	Avg. Degree	Clustering Coef.	Avg. Path Length	Assortativity Coef.
Whisper	690K	6,531K	9.47	0.033	4.28	-0.011
Facebook	707K	1,260K	1.78	0.059	10.13	0.116
Twitter	4,317K	16,972K	3.93	0.048	5.52	-0.025

- High average Degree. Users interact with large sample of other users.
- Whisper users are likely to interact with complete strangers who are highly unlikely to interact with each other (low clustering coef.)
- 100 random nodes. Avg. path length calculated. Shortest average path among 3.
- Above 3 used to infer that graph is random than "small world"

So, now,we are going to look at the topic that we have seen more in the past also which is network analysis. We have also have tutorials on this topic looking at how you can actually use metrics, which are developed in network analysis to make some inferences. Here they took whisper, they also took random users from facebook, compared also to twitter.So,thefirstcolumnisgraphwhisper,facebook,twitter,secondcolumnisnumber of nodes - 690,000 nodes in whisper, 707,000 in facebook, 4,317,000 nodes in twitter, number of edges, average degree, clustering coefficient, average path length and assortativity coefficient. I will go through average degree clustering coefficient and the rest and tell you what does this is mean.

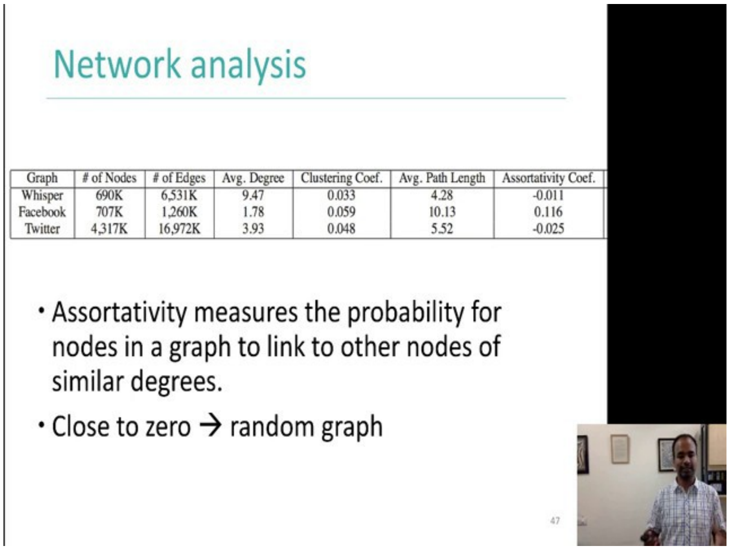
So, first if you look at the column four, which shows you average degree, the average degree is actually very high for whisper compared to facebook and twitter, what does it mean? This means that I am connecting to lot more. So, which is also connected to the clustering coefficient, but this says that users interact to the large sample of further users which means any user in whisper is not restricted only to a set of people.

But they otherwise interact, they interact with the large set of people.When you compare it to the facebook or twitter that we talk about the interactions are much closely connected, it's mostly with the followers that you have or probably people who mention you or probably the hashtag that you interested in. Facebook is mostly of friends. So, if you look at facebook, where it is only 1.78, for twitter is it is 3.93. Now let us look at whisper, it is 9.47 the degree in which they interact with the users in whisper is prettylarge.

Whisper users are likely to interact with complete strangers, look at the clustering coefficient. If you remember, what clustering coefficient tells, clustering coefficient just lets you to say how the graph looks like, whispers or whisper users are likely to interact with complete strangers who are highly unlikely to interact with each other also. So, if you look at the values it is pretty low,0.033 compared to 0.059 and 0.048 in twitter. So, theyhave also looked at 100 random nodes,average path length calculated, shortest path was the shortest average path among the 3 is actually for the whisper; if you look at the column average path length 4.28, 10.13 for facebook and 5.52 for twitter.

So, this just says that average length in the graph, if you take the whisper graph is actuallythelowestandthereisaveragepathlength.Basicallywhatdoesitmean itmeans average degree being highest, clustering coefficient being lowest, average path length being lowest is inferred that is it is the random graph. People interactions are completely random, there is no specific small world phenomenon that happens in a network like whisper. That is a good difference from the traditional networks that we have seen like facebook and twitter.

(ReferSlideTime:26:17)



Network analysis

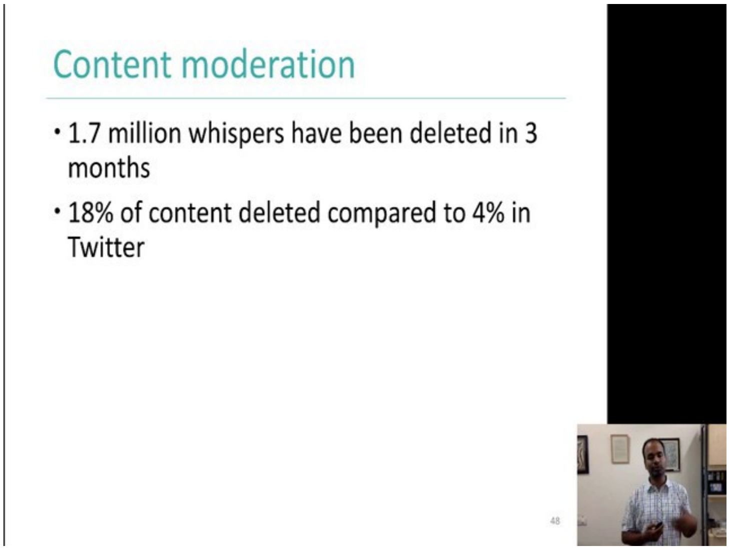
Graph	# of Nodes	# of Edges	Avg. Degree	Clustering Coef.	Avg. Path Length	Assortativity Coef.
Whisper	690K	6,531K	9.47	0.033	4.28	-0.011
Facebook	707K	1,260K	1.78	0.059	10.13	0.116
Twitter	4,317K	16,972K	3.93	0.048	5.52	-0.025

- Assortativity measures the probability for nodes in a graph to link to other nodes of similar degrees.
- Close to zero → random graph

47

Now, let us look at assortativity. Assortativity measures the probability of nodes in a graph to link to other nodes of similar degrees. So, the more the value that is closer to 0, or the less the value is, it is actually assumed that the graph is a random or you can infer the graph is a random graph. If you look at it, it is the lowest value minus 0.011 and this is the assortativity coefficient of the whisper, for all the 3 graphs. It basically says that it is a random graph. I know that is clear that. So, essentially the conclusion from this network analysis that you can draw is that whisper network is a random graph, whisper network people actually interact with the random people and the graph is actually pretty sparse.

(ReferSlideTime:27:27)



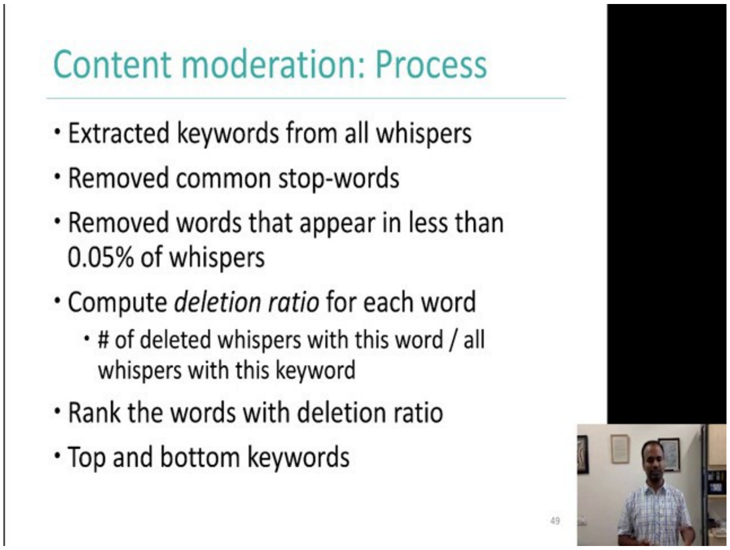
Content moderation

- 1.7 million whispers have been deleted in 3 months
- 18% of content deleted compared to 4% in Twitter

So, now another interesting thing that they did is to study what content was getting deleted on whisper. So, for this they collected the 1 point 7 million whispers, that have been deleted in 3 months, 18 percent of the content deleted. 18 percent of the total generated content was deleted from whisper where as compared to twitter, which is only 4 percent.

Andthisbegsthequestionwhich isthatfor whisperwhyisthispercentage high,because anonymous content you posted today you feel like there is some problem you feel like you created the some contents which others do not like to see or you do not want any attribution to you, even though it is an anonymous network, still you want to get itdeleted. So, there is higher proportion of content generated on whisper which is getting deleted.

(ReferSlideTime:28:34)



Content moderation: Process

- Extracted keywords from all whispers
- Removed common stop-words
- Removed words that appear in less than 0.05% of whispers
- Compute *deletion ratio* for each word
  - # of deleted whispers with this word / all whispers with this keyword
- Rank the words with deletion ratio
- Top and bottom keywords

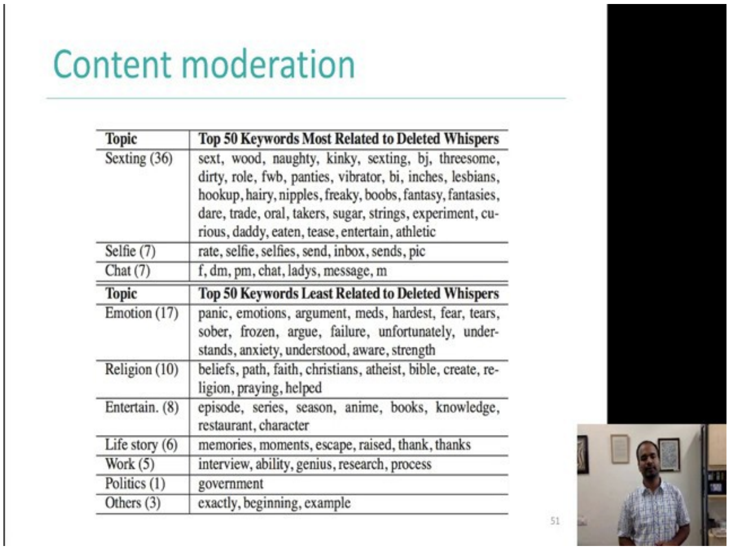
49

Content moderation, so what moderation in the context of whisper is, the analysis that they did was, they extracted keywords from all whispers, which put all the text that was created that was drawn from or collected from whisper, removed the common stop words, removed words that appear in less than 0.05 percent of whisper, that remove all the words that actually people care about or people have used it, compute deletion ratio they take calculated a value, which actually says that number of deleted.

Whispers with these words, by all whispers with this words. Which is essentially to say that what is the chance if the word appears in the post; and what is ratio for this word getting deleted; what is the ratio that whisper that has this word getting deleted. And it ranks the words with deletion ratio, they rank basically all the words which are with the deletion ratio and they looked at top ten and bottom, top keywords and the bottom keywords, here is the table which actually shows you the top keywords and the bottom keywords.

We will see it in the next slide, they ran this methods for 9 million original whispers. They saw the 1.7 million are deleted, 2324 keywords ranked by deletion ratio, manually they put them in categories to see which categories are largest amongst of deletion the lowest number they rank them in the tables here.

(ReferSlideTime:30:30)



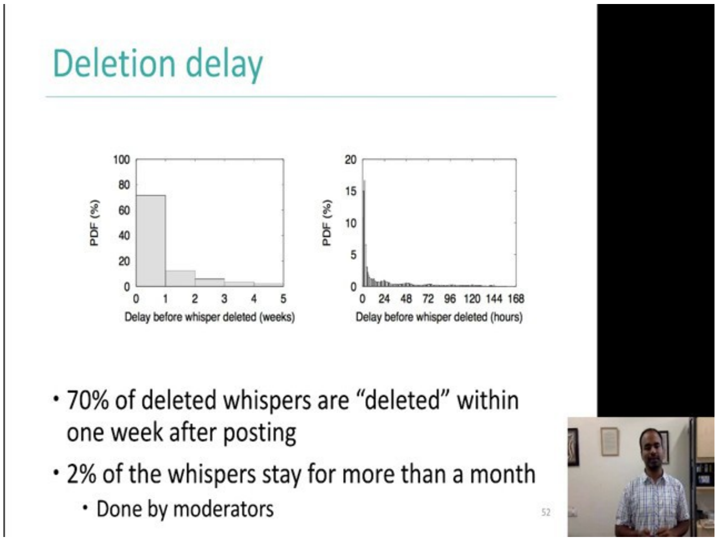
Content moderation

Topic	Top 50 Keywords Most Related to Deleted Whispers
Sexting (36)	sext, wood, naughty, kinky, sexting, bj, threesome, dirty, role, fwb, panties, vibrator, bi, inches, lesbians, hookup, hairy, nipples, freaky, boobs, fantasy, fantasies, dare, trade, oral, takers, sugar, strings, experiment, curious, daddy, eaten, tease, entertain, athletic
Selfie (7)	rate, selfie, selfies, send, inbox, sends, pic
Chat (7)	f, dm, pm, chat, ladys, message, m
Topic	**Top 50 Keywords Least Related to Deleted Whispers**
Emotion (17)	panic, emotions, argument, meds, hardest, fear, tears, sober, frozen, argue, failure, unfortunately, understands, anxiety, understood, aware, strength
Religion (10)	beliefs, path, faith, christians, atheist, bible, create, religion, praying, helped
Entertain. (8)	episode, series, season, anime, books, knowledge, restaurant, character
Life story (6)	memories, moments, escape, raised, thank, thanks
Work (5)	interview, ability, genius, research, process
Politics (1)	government
Others (3)	exactly, beginning, example

51

The categories that they had a sexting, selfie, chat, topic, emotion, the top 50 keywords most related to the deleted whispers. And the top 50 keywords least related to deleted whispers. The top points 50 keywords that are in the top, the bottom of the table gives you the bottom that was there on the deleted whispers. Essentially showing that you sexting, selfie and chat are the categories, which were most frequently deleted, and emotion, religion, entertain, life story, work, politics and others were the least deleted categories from the whispers.

(ReferSlideTime:31:34)



Deletion delay

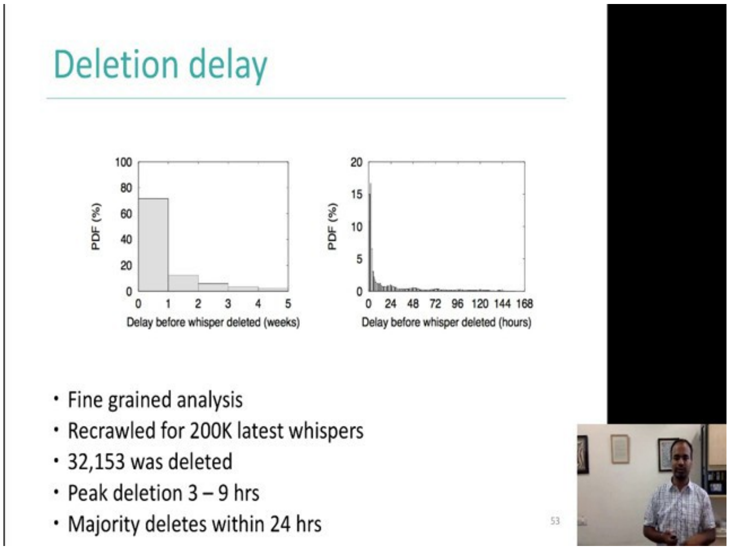
PDF (%)
100
80
60
40
20
0
0 1 2 3 4 5
Delay before whisper deleted (weeks)
PDF (%)
20
15
10
5
0
0 24 48 72 96 120 144 168
Delay before whisper deleted (hours)

- 70% of deleted whispers are "deleted" within one week after posting
- 2% of the whispers stay for more than a month
  - Done by moderators

52

This graphis actually showingyou, Imean let us lookat thewayinwhich thedeletion is happening, how much time and relationship for actually deletion. 70 percent of the deleted whispers are deleted within one week after posting. So, that is the first graphfrom the left, which is 70 percent of the post, x axis is week y axis probability of getting deleted, proportion of whispers getting deleted, 70 percent of the whispers are deleted within one week after posting. The right side shows you delay before whisper is getting deleted, that this one, 2 percent of the whispers stay for more than a month, if you see it had a four weeks that is the graph from the left.

(ReferSlideTime:32:44)



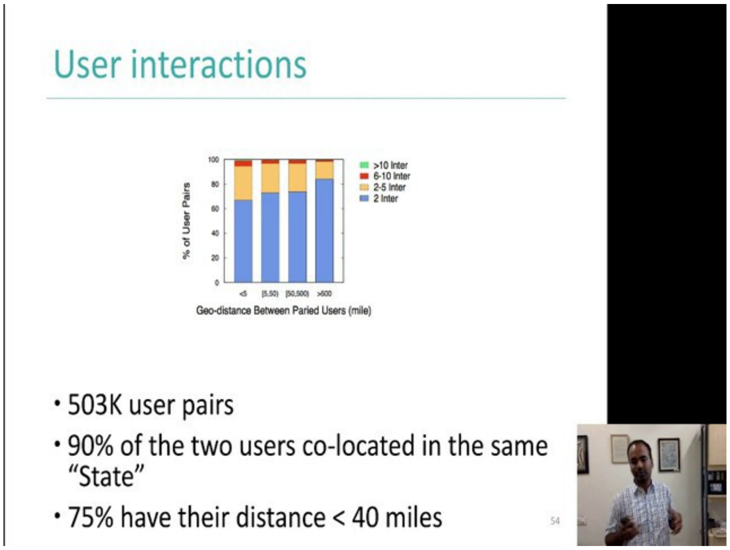
Deletion delay

PDF (%)
100
80
60
40
20
0
0
1
2
3
4
5
Delay before whisper deleted (weeks)
PDF (%)
20
15
10
5
0
0
24
48
72
96
120
144
168
Delay before whisper deleted (hours)
53

- Fine grained analysis
- Recrawled for 200K latest whispers
- 32,153 was deleted
- Peak deletion 3 – 9 hrs
- Majority deletes within 24 hrs

Now, lets look at the content analysis on the right. fine grained analysis, recrawled for 200 thousand latest whispers, they were actually interested in trying to understand how many hours, this was a week the first graph, what is the analysis in the hours that is what they are interested. They actually found that 32153 was deleted, peak deletion was between 3 and 9 hours, which is any post on, if it is was both get deleted is between 3 to 9 hours. Majority deletes within 24 hours. So, it is even if youzoom in to the data forthis one week, majority of them are actually getting deleted within the first 24 hours.

(ReferSlideTime:33:40)



User interactions

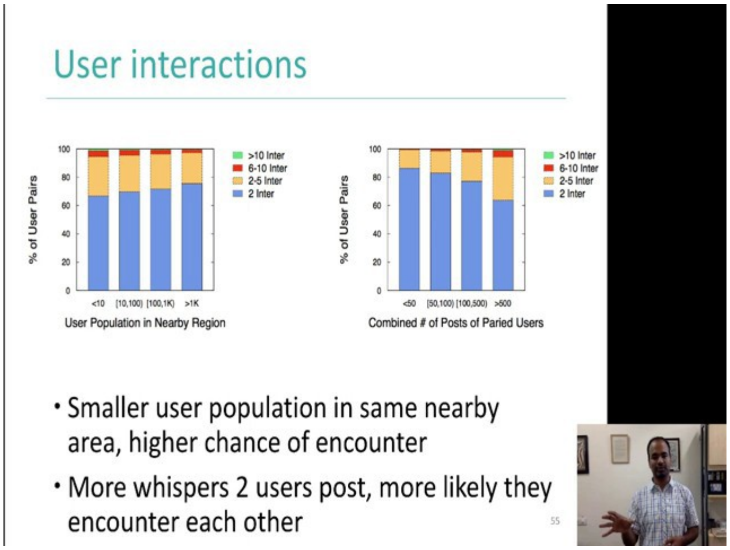
- 503K user pairs
- 90% of the two users co-located in the same "State"
- 75% have their distance < 40 miles

54

User interactions is another interesting analysis that they did, which is how frequently how users actually interact, which is 2 handles in this case are actually interacting between them. This graph is showing you on the x axis, geo distance between the pairedusers, of course these are the locations that people actually disclose, percentage of user pairs, what is a number, what is the percentage of user pairs which are actually interacting.

The colors are blues is two interactions, yellows are two to five interactions, red is 6 to10interactions, anything thatwas above10interactions, which wasactuallygiven green. Youcould alreadyguess,that thenumber ofinteractions whichis higher,is actuallyvery lowinthewhisper,whichagainusingthenetworkanalysis, usingthethingsthatwehave alreadyseen,youcouldactuallymaketheinference.Thatiswhygreenisverylowonthe graph. 90 percent of that the two users are co located in the same state, 75 percent have their distance which is less than 40 miles. So, this basically shows that the users are also co located very closely, within 75 percent have their distance less than 40 miles.

(ReferSlideTime:35:21)

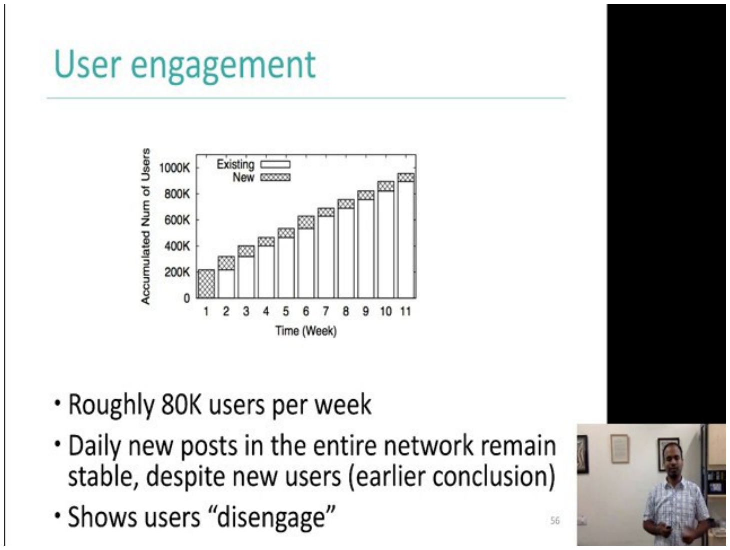


User interactions

- Smaller user population in same nearby area, higher chance of encounter
- More whispers 2 users post, more likely they encounter each other

Smaller user population in same nearby area, higher chance of encounter, so if you look at a graph less than 10 to 100, 100 to 100,000, greater than 1000 on the left and then combined post of paired user which is on the right. The left side is showing you user population in nearby region. The right is showing you combined number of posts of paired users, More whispers two users post, more likely they encounter with each other. If the users are likely to post more; they are likely to interact more also. So, that'sactually looking at a the right graph which is combined number of posts of paired users, more whispers two users post which says if you and I want to interact, if you and I actually generating more whispers that's more likely that you and I interact, that is the inference that you can draw.

(ReferSlideTime:36:37)



User engagement

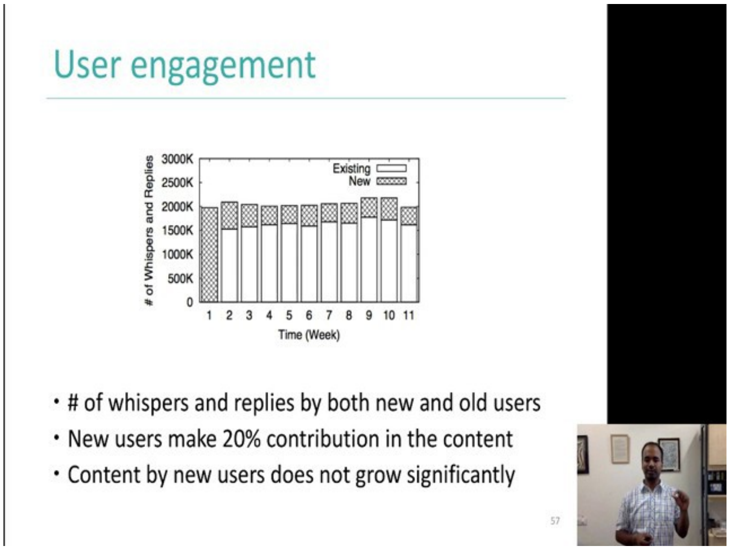
- Roughly 80K users per week
- Daily new posts in the entire network remain stable, despite new users (earlier conclusion)
- Shows users "disengage"

Now,look at how users engage in thesein this network on whisper.So, here x axis is the time of week, y axis is the accumulated number of users and the two data points that are drawninthisgraphis theexistingandthenewcontent that isgetting generated. Roughly 80,000 user per week are interacting daily,new posts in the entire network remain stable thatwewillactuallyseeinthenextgraphalso.Howthisrenames same isactually,ifyou see the new content that is generated per week on the graph, they actually seeing to be very same across.

Even though there are more users that are getting added to the network, it does not look likethenewcontentthatisgettinggeneratedisactuallyincreasing.Thatistheinteresting conclusion that you can actually see in this graph, daily new posts in the entire network remain stable despite new users joining. That you can see actually accumulated number of users is increasing.

So,thisbasicallyshows thateventhoughthe users areincreasing,whichmeans theposts should be increasing, the engagement should be increasing. But it is not, this basically shows that that are lot of people who are getting into the network, generating some content and then lot of people, who are known to be already in the network are not generating the content. That is why the proportion of the net the content is getting generated is always remaining the same. Even though there are more users are added toit.

(ReferSlideTime:38:37)

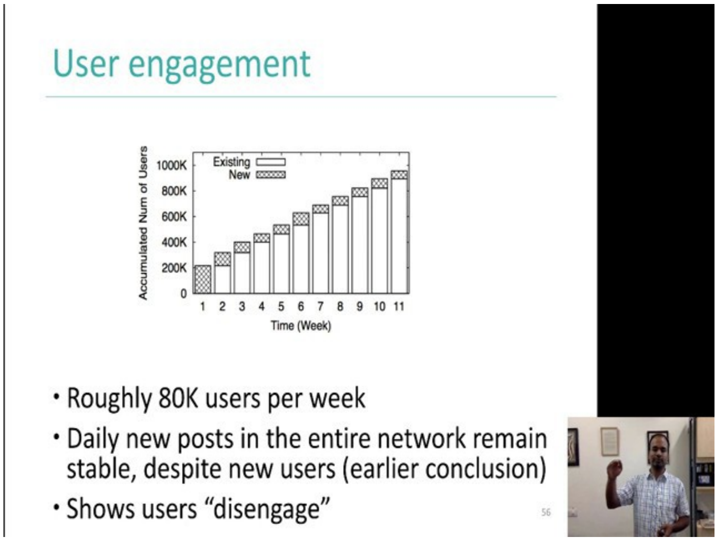


User engagement

- # of whispers and replies by both new and old users
- New users make 20% contribution in the content
- Content by new users does not grow significantly

User engagement, so, here is the graph that I will actually show you. Number ofwhispers and replies this is weeks,this is also in weeks.So, it is kind of the same kind of graph and you will see number of whispers and replies by both new and old users if you see, the top, the new users make a twenty percent of the contribution in the content. Content by new users do not grow, right. So, this is time and number of whispers and replies for that particular week.

(ReferSlideTime:39:17)

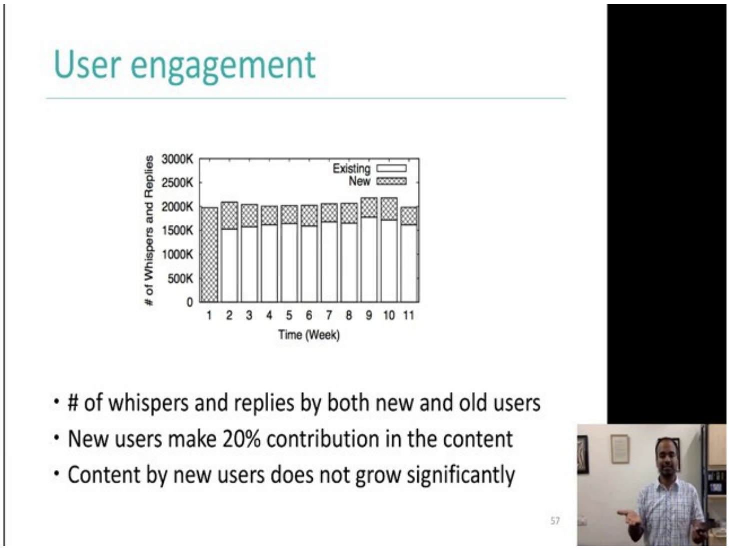


User engagement

- Roughly 80K users per week
- Daily new posts in the entire network remain stable, despite new users (earlier conclusion)
- Shows users “disengage”

The earlier graph was the cumulative number of users right, accumulated number of users. So, the existing user which is the light without the check that the graph the bar is actually, which is risen which is showing you that the number of users are actually risen.

(ReferSlideTime:39:38)



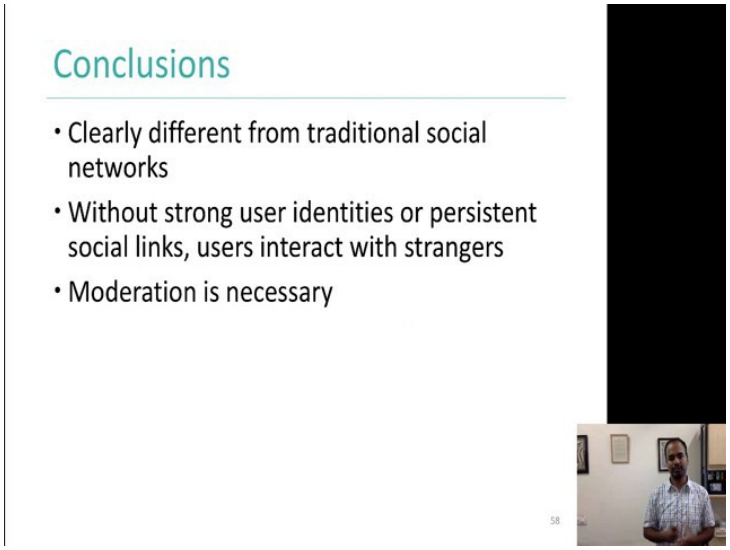
User engagement

- # of whispers and replies by both new and old users
- New users make 20% contribution in the content
- Content by new users does not grow significantly

57

The next graph is essentially showing you that the post and the replies that, that is there in the network is actually pretty constant even though the number of users are increased, correct. So, that is the conclusion that they had in the user engagement, that is basically kind of addresses the question that we started off with, which is do user's engaged differently in a network like whisper.

(ReferSlideTime:40:06)



Conclusions

- Clearly different from traditional social networks
- Without strong user identities or persistent social links, users interact with strangers
- Moderation is necessary

So, the conclusion is clearly different from traditional social networks. We saw that the average path length is different, clustering coefficient is different, we saw that the deletion is actually pretty high and inferences like that. Without strong user identities or persistent social links users interact with strangers which is also derived from the conclusion that user is interacting with any random people on the network right. There was not a persistent relationship between the users, moderation is of course necessary because content is getting deleted very highly.

So, that is all I had for this week, in this week, we saw what 8.2 we saw anonymous networks,whicharenetworkswhereyoucanpostcontentwhereyoucan maintain ahigh anonymity. In 8.1, we saw how to actually do identical resolution with multiple accounts given to us. That is we get. I will see you in next week.

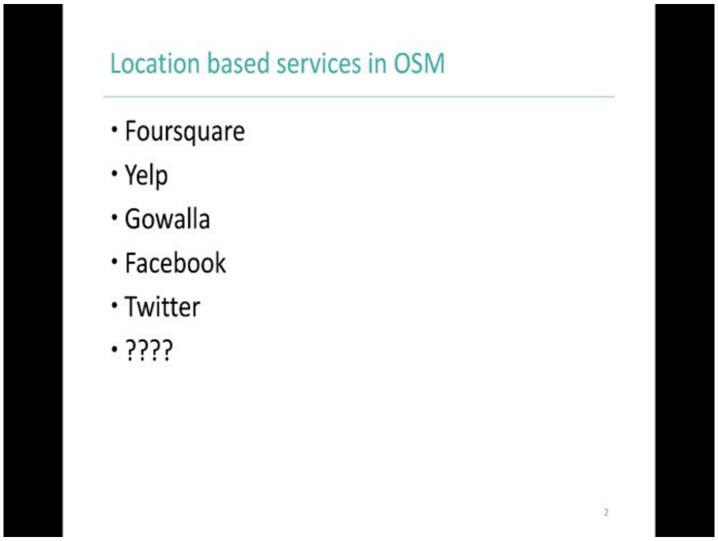
Unit-5

Privacy in Location Based Social Networks Part-I

WelcomebacktoPrivacyandSecurityinOnlineSocialMediacourseonNPTEL.Thisis week 9. So, what we will do today and in this week is that we will look at some of the research which was done in terms of just looking at the papers itself and going through the paper in terms of different techniques that are applied. The goal here is that next couple of hours, what we will do is we will get you to actually look at research papers written on the topic and we will go through the same analysis that you have done across the course for you to get a sense of how the analysis that you have done fit into actually making some interesting inferences.

Where do people start, how do theywriteapaper whatallthings fits into thepaper,what allanalysis that theyhave doneessentially,itis looking atthecontentthatyou have seen in the past, but in terms of the structure of the paper itself we will go through some of them.

(ReferSlideTime:01:22)



Location based services in OSM

- Foursquare
- Yelp
- Gowalla
- Facebook
- Twitter
- ????

2

So, location, the first topic that we will go through is location based privacy problems. Location based services on online social media, there are many actually and there are some which are very popular which are like Foursquare, Yelp, Gowalla, Facebook, Twitter these are the different social media services that are actually pretty popular in terms of giving the location based services, for example, in Foursquare you could actually see where is the next, let us take petrol pump, in the directions that you're travelling. Secondly, in the Yelp you can look at where the restaurant or places that you are interested in, what kind of reviews do they have, Facebook you can actually looked you can actually do check in into a location in Facebook.

Similarly,you can do the geo location information shared on Twitter.So, essentially you must have seen check-ins from people on Facebook like XYZ is on T3 Indira Gandhi AirportinDelhitravelingtoXYZplace.So,thatislocationbasedservice,youopenyour phone, you check into this location or post this status updatewith thelocation updated init. Foursquare, Yelp, these are very, very popular services which actually do location based services, these are called location based social networks.

(ReferSlideTime:03:18)

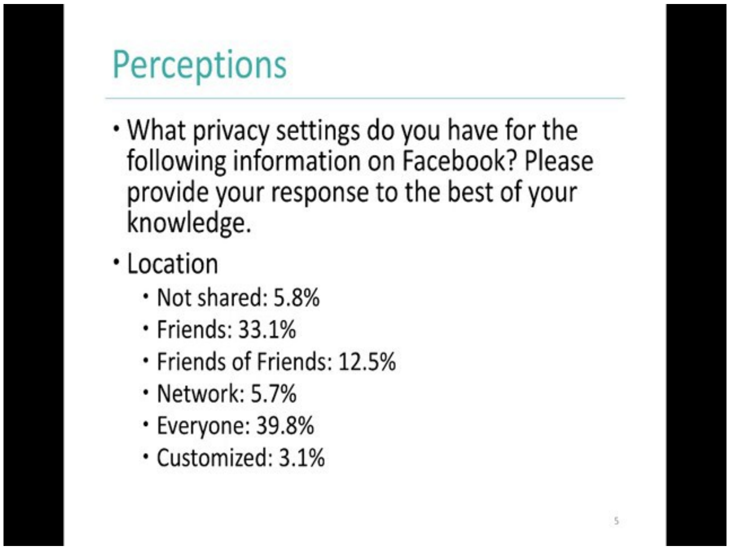
Privacy concerns in them?

- Foursquare
- Yelp
- Gowalla
- Facebook
- Twitter
- ????

So,ofcourse,allofthemhavesomesortofprivacyconcernstoo.Iam,Iamsurebynow youalreadyrealizedtheprivacyconcernsinlocationbasedserviceswhichare,wheremy location is shared, if my location is shared there are going to be concerns accordingly, that is, somebody else will get to know where I am, if the information is given public, then many more people actually, more than just your friends get access to your location at that given point in time, right.

So, that is the privacy concern and of course, every each of these social media services will have its own privacy concerns also.

(ReferSlideTime:04:08)



Perceptions

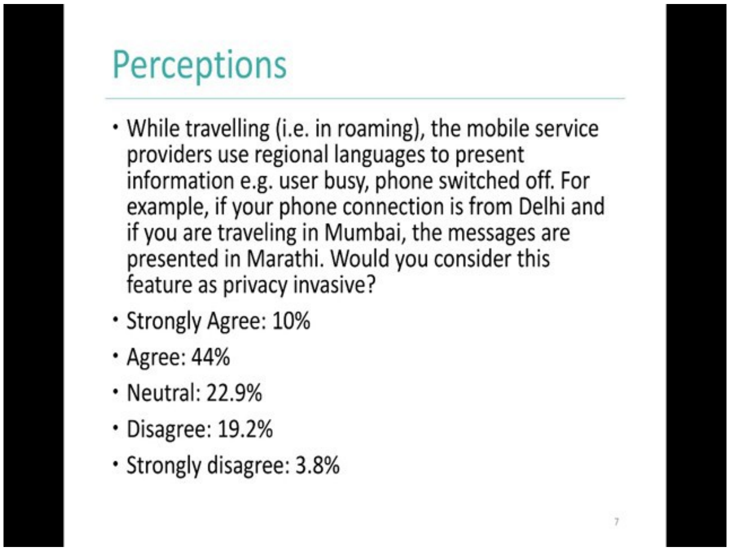
- What privacy settings do you have for the following information on Facebook? Please provide your response to the best of your knowledge.
- Location
  - Not shared: 5.8%
  - Friends: 33.1%
  - Friends of Friends: 12.5%
  - Network: 5.7%
  - Everyone: 39.8%
  - Customized: 3.1%

So,justtogiveyou asenseofwhataretheperceptionsintermsofactually theselocation based services. If you remember, there was a study that I referred earlier in the lectures also, where 10427 people were asked some questions about different aspects of privacy. One section in that was online social media, here is a question that was asked in the study, the question reads, what privacy settings do you have for the following information on Facebook?

Please provide your responses to the best of your knowledge. Location - not shared, friends, friends of friends, network, everyone, and ‘I have customized it’, those are the differentoptionsthatwasgivenforthequestion.Idonotknowwhereeachofyouwillfit in, but if you look at it, here is a distribution of responses that was got for this particular question - 33 percent to friends, friends of friends is 12.5 percent, network is 5.7, everyone is 39 percent and customized is 3.1.

It just says that location, is been. what privacy settings has been set up is for everyone, right, location information is shared maximum to everyone on Facebook, and if you put everyone in friends that is a lot of percentage of responses which where the information is been shared. So, that is a point I wanted to get across, which is that information that is shared through location can be actually used for things beyond what you think for now also.

(ReferSlideTime:06:26)



Perceptions

- While travelling (i.e. in roaming), the mobile service providers use regional languages to present information e.g. user busy, phone switched off. For example, if your phone connection is from Delhi and if you are traveling in Mumbai, the messages are presented in Marathi. Would you consider this feature as privacy invasive?
- Strongly Agree: 10%
- Agree: 44%
- Neutral: 22.9%
- Disagree: 19.2%
- Strongly disagree: 3.8%

Perceptions,again.Anotherquestioninthesamesurvey,samedatacollection,whichwas asked to get some responses from participants - While traveling the mobile service providers use regional languages to present information, that is, user busy, phone switched off.

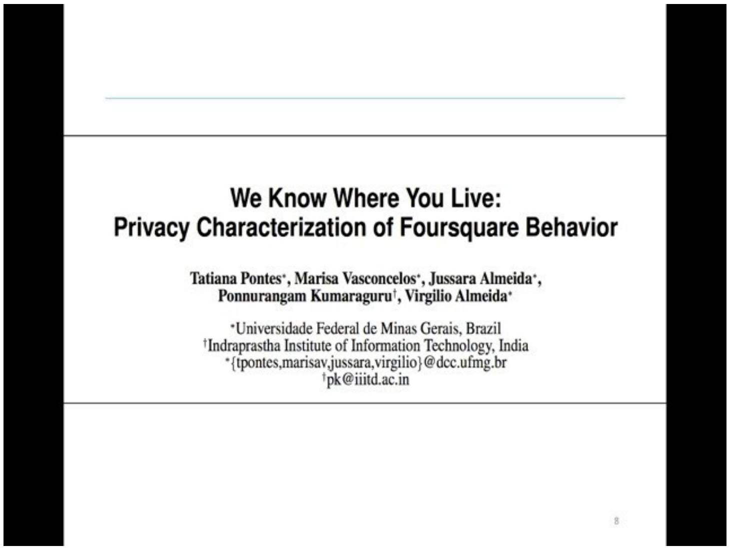
For example if your phone connection is from Delhi and if you are traveling in Mumbai the messages are presented in Marathi. Would you consider this feature as privacy invasive, I am sure you got the question, the question is simply that I have a number whichIboughtitinDelhiandIamtravelinginMumbaiandmyphoneisswitchedoff.

When somebody calls me, the messages are actually want to be saying that the phone is switched off, but it is going to be giving that message in Marathi.

The same thing changes when I go to Kerala, this number is going to be actually this messageis going tobein Malayalam, which isbasicallyrevealing theinformation where you are at when somebody is trying to call you. This can be privacy intrusive because it just says that the exact location where you are at least in the regional language the state where you are. So, if you see, would you consider this feature as privacy invasive? 44 percentof themsay,thatit is privacyinvasive, 42 percent neutral, disagree is 19 percent, strongly disagree is 3 percent, and strongly agree is 10 percent.

So, essentially if you look at just the people who are agreeing, it is 54 percent, people who are disagreeing is about 22- 23 percent and rest are in neutral that just to share that and allofthisdatawascollectedonlyinIndia.Theperceptionsherewearetalkingabout person living in India who is thinking about these problems.

(ReferSlideTime:08:34)



We Know Where You Live:
Privacy Characterization of Foursquare Behavior

Tatiana Pontes[*], Marisa Vasconcelos[*], Jussara Almeida[*], Ponnurangam Kumaraguru[†], Virgilio Almeida[*]

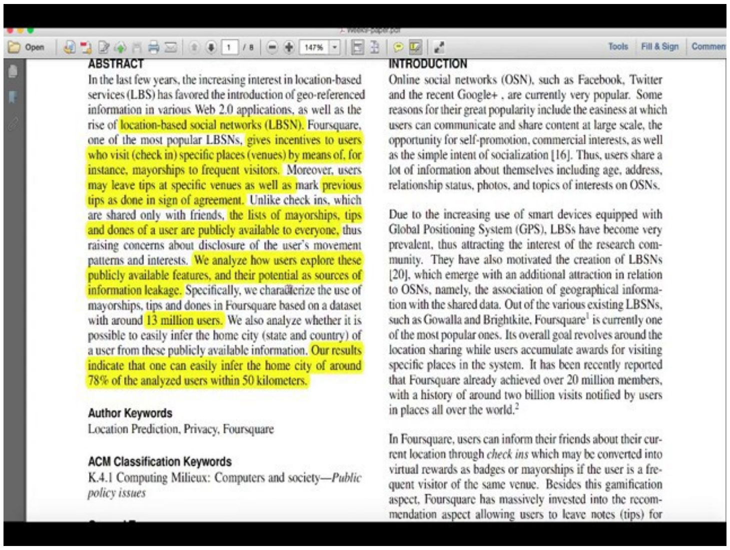
[*]Universidade Federal de Minas Gerais, Brazil
[†]Indraprastha Institute of Information Technology, India
[*]{tpontes,marisav,jussara,virgilio}@dcc.ufmg.br
[†]pk@iiitd.ac.in

8

So, that is basically a sense of what location based privacy is. Location based privacy services are what kind of issues you can have, but what we will do is as I said we will actuallylookatsomespecificresearchthathavedoneintermsofinformation.Interms of analysis, data collection, what the topic is, we will take the paper now and will go through the paper only in detail about what was done what kind of analysis was done, what kind of inferences was drawn.

This will actually it help you to get a better sense of how to use the social media data techniques that we have learnt in the in this course until now in terms of actuallymaking some inferences. So, now, here is a paper that we will look at and spend some time onthe paper.The title of the paper is 'WeKnow WhereYou Live: PrivacyCharacterizationof Foursquare behavior'.We Know Where You Live: Privacy Characterization ofFoursquare behavior

(ReferSlideTime:09:52)



ABSTRACT

In the last few years, the increasing interest in location-based services (LBS) has favored the introduction of geo-referenced information in various Web 2.0 applications, as well as the rise of location-based social networks (LBSN). Foursquare, one of the most popular LBSNs, gives incentives to users who visit (check in) specific places (venues) by means of, for instance, mayorships to frequent visitors. Moreover, users may leave tips at specific venues as well as mark previous tips as done in sign of agreement. Unlike check ins, which are shared only with friends, the lists of mayorships, tips and dones of a user are publicly available to everyone, thus raising concerns about disclosure of the user's movement patterns and interests. We analyze how users explore these publicly available features, and their potential as sources of information leakage. Specifically, we characterize the use of mayorships, tips and dones in Foursquare based on a dataset with around 13 million users. We also analyze whether it is possible to easily infer the home city (state and country) of a user from these publicly available information. Our results indicate that one can easily infer the home city of around 78% of the analyzed users within 50 kilometers.

Author Keywords

Location Prediction, Privacy, Foursquare

ACM Classification Keywords

K.4.1 Computing Milieux: Computers and society—*Public policy issues*

INTRODUCTION

Online social networks (OSN), such as Facebook, Twitter and the recent Google+ , are currently very popular. Some reasons for their great popularity include the easiness at which users can communicate and share content at large scale, the opportunity for self-promotion, commercial interests, as well as the simple intent of socialization [16]. Thus, users share a lot of information about themselves including age, address, relationship status, photos, and topics of interests on OSNs.

Due to the increasing use of smart devices equipped with Global Positioning System (GPS), LBSs have become very prevalent, thus attracting the interest of the research community. They have also motivated the creation of LBSNs [20], which emerge with an additional attraction in relation to OSNs, namely, the association of geographical information with the shared data. Out of the various existing LBSNs, such as Gowalla and Brightkite, Foursquare[1] is currently one of the most popular ones. Its overall goal revolves around the location sharing while users accumulate awards for visiting specific places in the system. It has been recently reported that Foursquare already achieved over 20 million members, with a history of around two billion visits notified by users in places all over the world.[2]

In Foursquare, users can inform their friends about their current location through *check ins* which may be converted into virtual rewards as badges or mayorships if the user is a frequent visitor of the same venue. Besides this gamification aspect, Foursquare has massively invested into the recommendation aspect allowing users to leave notes (tips) for

This is what we will do. We will go through the paper content and look at what is mentioned in the paper and go through them. I will actually give some insights about what is going on. So, as I said, we are going to look at location based social network. Foursquare is one of the popular ones. The way Foursquare works is, it gives incentives to users to use the check-ins specific places.

So, a visit is a check-in.You go to a place, for example, IIITDelhi is a venue, you come toIIITDelhi,whichisavenue,andthenyoudoa check-in,whichisyouarevisitingthe place and mayorship is for frequent visits. Mayorship is nothing, but it is the person who is being to that place for most number of times in the last 60 days, which is. he is checked-in that location for the most number of times in the last 60 days. that is the mayorship and people do mayorship for many reasons, there is actually incentives that are done.

For example, if you go to any malls, if you go to places they are actually giving you -and if you are a mayor of that location - you can actually get more discounts or you can actually get some parking spots free for a week or so. So, being a mayor is actually giving you some incentives. So, check-ins is the action to be in that place, venues is the location or the places, mayorships.

They can also leave tips, tips at specific menus are the is the information that people put in to the Foursquares saying I will checked in into this location, for examples, SaravanaBhavan, I had food, food was pretty good. That is a tip and if you agree on the tips, like the like in Facebook or like the re tweet or the like in Twitter again, there is something called as done, d o n e in Foursquare which is actually again saying that I like this tip.So, if you see check-ins are actually available only for your friends, but the list of mayorships, tips and done of users are publicly available to everyone.

So, this basically allows us to collect this information and do some interesting analysis which is what we will actually look at in this part of the lecture - collecting data from Foursquare,analyzingthemandmakingsomeinterestinginferences,particularlylooking at privacy issues in Foursquare and given the title of the paper, particularly looking at, can we actually find out where people live, from just the check-in, from just their Foursquare behavior.

This paper basically explores these publicly available features – mayorships, tips, dones, and their usage for informational leakage, but interesting part of this particular paper is that it actually uses the data of entire Foursquare. At that point in time, which is 13 million users, and the paper kind of concludes that - there are many interesting conclusions in this paper we will look at all of them in detail - but in the abstract, they talkabout,ourresultsindicatethat,oneeasily,onecaneasilyinferthehomecityof around 75 percent of the analyzed users within 58 hours. So, that is actually is privacy intrusive if I can actually tell you where you live just by looking at your check-in location and the Foursquare information then it is actually privacy intrusive.

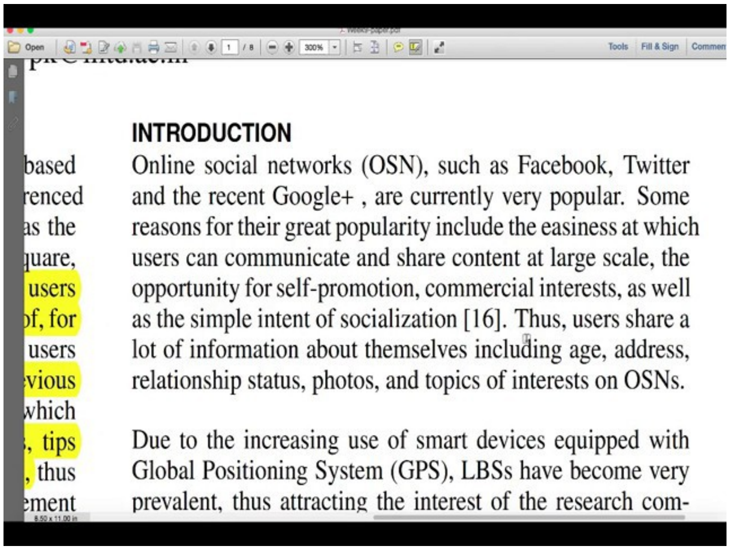
In this case, only publicly available information is used. There used to be a websitecalled please rob me dot com (pleaserobme.com), which they took it down after some time.This websitedidsomethinginteresting-whichispleaserobdotrobbedandme dot com please rob me dot com - the creators of this websites basically looked at the tweets and if a user talks about location x or if the user created the account user location that is the information that you will actually get in Twitter, with that information they were actually saying that a person created an account in Delhi and he is talking about a weather in Chennai that is a probability that he is not he is not at home.

Or if you have checked in, if you have done a post in tweet with your geolocation on from Delhi and in another hour or so, you are talking about actually weather and couple ofhoursyouaretalkingaboutweatherinChennaiagain,thatisaprobabilitythatyou are notathome,you movedfromDelhitoChennai.So,theywereusingthisinformationand the tweets that were of this category which is user from location x and weather or other information, for example, even traffic right, you are from Chennai and you actually post you're posting traffic about Delhi and the tweet is also saying that I am going to this place and that is a heavy traffic, there is a higher probability that you are actually in Delhi.

Using all this information they find out that this particular user who is posting this tweet is not in his or her home in location and therefore, they would actually take the post and show it in this website called please rob me dot com forburglars to goand robthe home. And again another leak of information, another impact of information about your location, can be actually seen from the example of please rob me dot com.

So, then I think the paper talks about in general about what Foursquare is and for the benefit of the students in the class I have highlighted the parts that have actually something that I am going to be looking at, but feel free to look at the entire paper, but for thebenefitof timeconstraintandthelecturesalsoI am goingtolookatonly theones thatIhave highlightedhere.

(ReferSlideTime:17:54)



based
renced
as the
uare,
users
of, for
users
vious
which
, tips
, thus
ment

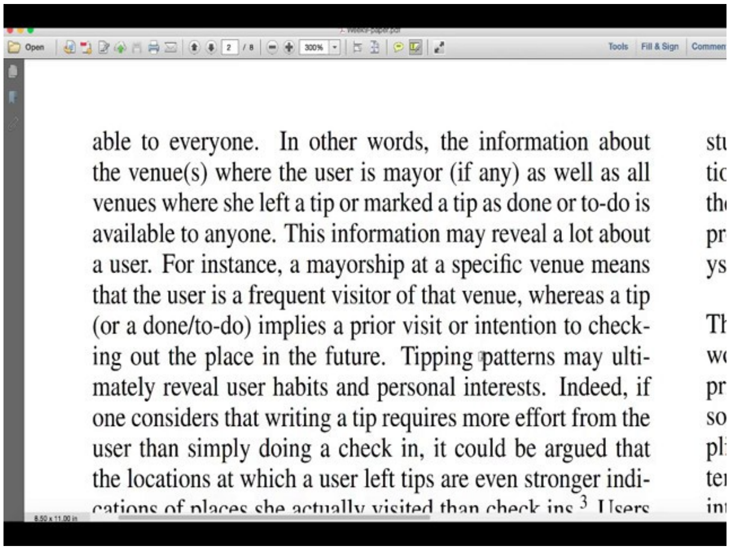
INTRODUCTION

Online social networks (OSN), such as Facebook, Twitter and the recent Google+ , are currently very popular. Some reasons for their great popularity include the easiness at which users can communicate and share content at large scale, the opportunity for self-promotion, commercial interests, as well as the simple intent of socialization [16]. Thus, users share a lot of information about themselves including age, address, relationship status, photos, and topics of interests on OSNs.

Due to the increasing use of smart devices equipped with Global Positioning System (GPS), LBSs have become very prevalent, thus attracting the interest of the research com-

So for people who are curious about writing papers like this, the structure it is – abstract, then there is introduction, introduction you generally talk about what the problem of attack in the paper is, give some background about the domain - in this case it is online social media - then talking about location based services, then giving some information about, so it has to be both, giving information about the topic - location based services, give more of quantitative numbers also saying how many people are using it, what level of impact is it making and things like that.

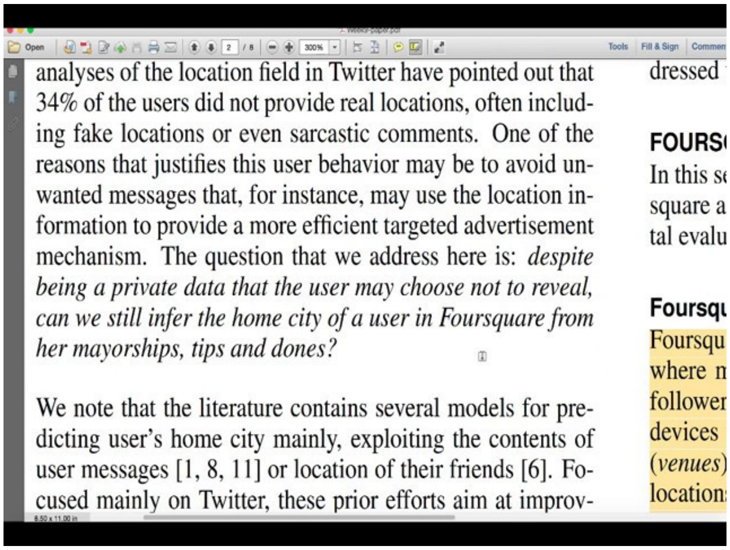
(ReferSlideTime:18:20)



able to everyone. In other words, the information about the venue(s) where the user is mayor (if any) as well as all venues where she left a tip or marked a tip as done or to-do is available to anyone. This information may reveal a lot about a user. For instance, a mayorship at a specific venue means that the user is a frequent visitor of that venue, whereas a tip (or a done/to-do) implies a prior visit or intention to checking out the place in the future. Tipping patterns may ultimately reveal user habits and personal interests. Indeed, if one considers that writing a tip requires more effort from the user than simply doing a check in, it could be argued that the locations at which a user left tips are even stronger indications of places she actually visited than check ins.[3] Users

Andthenyou quicklymention aboutthe methodologythatyou did.

(ReferSlideTime:18:45)



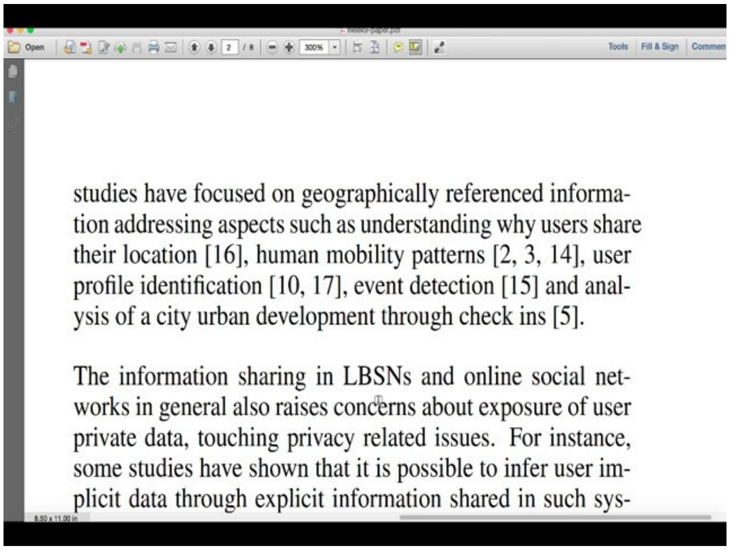
analyses of the location field in Twitter have pointed out that 34% of the users did not provide real locations, often including fake locations or even sarcastic comments. One of the reasons that justifies this user behavior may be to avoid unwanted messages that, for instance, may use the location information to provide a more efficient targeted advertisement mechanism. The question that we address here is: *despite being a private data that the user may choose not to reveal, can we still infer the home city of a user in Foursquare from her mayorships, tips and dones?*

We note that the literature contains several models for predicting user's home city mainly, exploiting the contents of user messages [1, 8, 11] or location of their friends [6]. Focused mainly on Twitter, these prior efforts aim at improv-

So, essentially introduction would be just a shorter abridged version of the rest of the paper which is covering methodology, inferences, contributions, and conclusions. And the question that the paper specifically addresses is, despite being a private data that the user may choose not to reveal, can we still infer the home city of a user in Foursquare from our mayorships, tips, and done, which are publicly available information.

So, that is the question and then in the paper there is something called as related work, which is focused on only the question that we are asking in introduction, you kind of generallymotivatefrom30,000feetheightabouttheproblemsaying,thisistheproblem, this is what we are planning to work. In related work, you focus on only the work that you are going to show in the paper and talk about past literature which are connected to that.

(ReferSlideTime:19:42)



studies have focused on geographically referenced information addressing aspects such as understanding why users share their location [16], human mobility patterns [2, 3, 14], user profile identification [10, 17], event detection [15] and analysis of a city urban development through check ins [5].

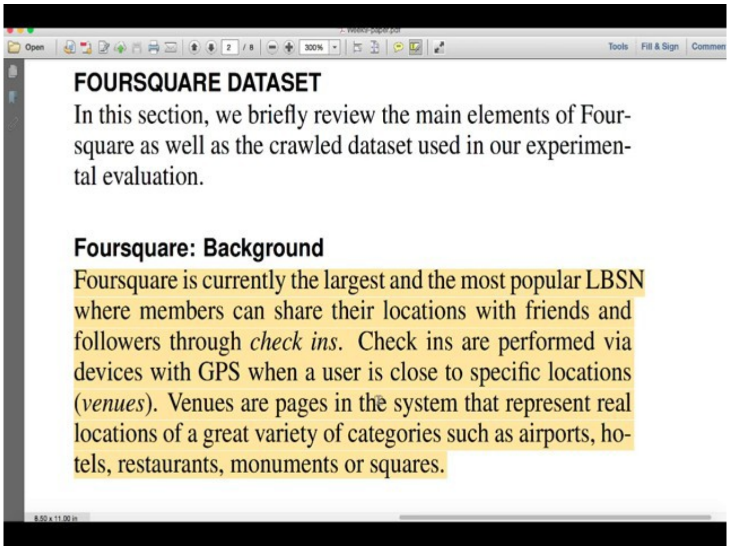
The information sharing in LBSNs and online social networks in general also raises concerns about exposure of user private data, touching privacy related issues. For instance, some studies have shown that it is possible to infer user implicit data through explicit information shared in such sys-

So,inthiscase,authorsaretalkingabouthowpeoplehaveusedhumanmobilitypatterns, even direction analysis of a city urban development, through check-ins. There is a lot of workactuallyinterms ofusingFoursquare datatofindoutthenatureofthecityandhow people have actually used the social network in a location based social network like Foursquare. People have actually used, researchers have used the information of Foursquare to design a city.

So, you can actually look at -if you are interested - you can actually look at this project called livelihoods.org. This is a project where they are actually using Foursquare information to see how people actually move in a given city and can we actually re-design a city keeping this information in mind.

Now,welookatsomethingmorecloselyinterms ofjustFoursquareDatasetitself.

(ReferSlideTime:21:13)



FOURSQUARE DATASET

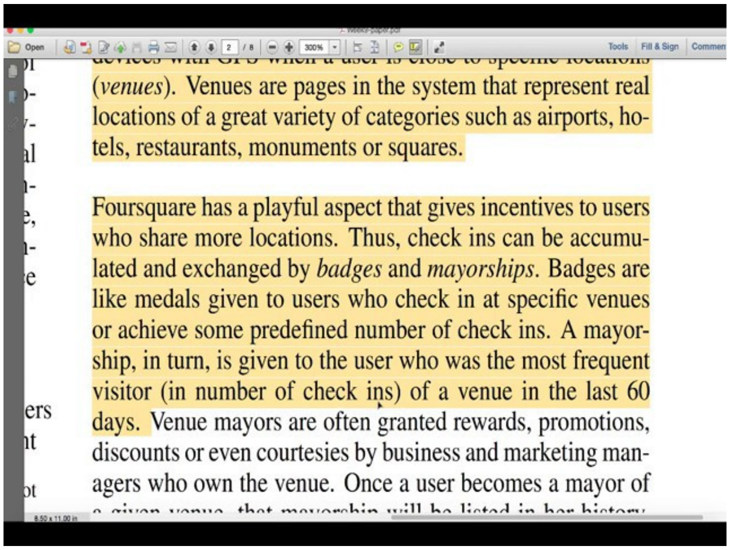
In this section, we briefly review the main elements of Foursquare as well as the crawled dataset used in our experimental evaluation.

Foursquare: Background

Foursquare is currently the largest and the most popular LBSN where members can share their locations with friends and followers through *check ins*. Check ins are performed via devices with GPS when a user is close to specific locations (*venues*). Venues are pages in the system that represent real locations of a great variety of categories such as airports, hotels, restaurants, monuments or squares.

To understand the analysis better you need to actually be clearer about some of the terminologies. So, here I am going to actually explain them, something we have already done. So, check-in is, the, where members can share the location with friends and followers through check-ins, that is the check-in. Check-ins are performed via devices with GPS when a user is close to a specific location, which is a venue, which I have said before. So,venuescanbeairport,restaurantand monuments, you comeand checkin that particular location.

(ReferSlideTime:21:55)



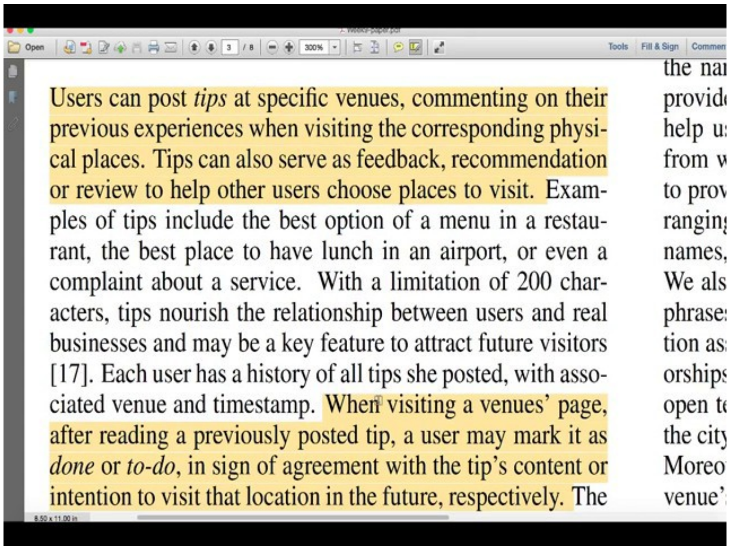
devices with GPS when a user is close to specific locations (*venues*). Venues are pages in the system that represent real locations of a great variety of categories such as airports, hotels, restaurants, monuments or squares.

Foursquare has a playful aspect that gives incentives to users who share more locations. Thus, check ins can be accumulated and exchanged by *badges* and *mayorships*. Badges are like medals given to users who check in at specific venues or achieve some predefined number of check ins. A mayorship, in turn, is given to the user who was the most frequent visitor (in number of check ins) of a venue in the last 60 days. Venue mayors are often granted rewards, promotions, discounts or even courtesies by business and marketing managers who own the venue. Once a user becomes a mayor of a given venue, that mayorship will be listed in her history

So, the interesting aspect of Foursquare is that they have made it more - gamified it - which is basically turning the platform where users actually get more incentive, more addicted through this gamification nature, which is, user get badges, mayorships here. Badges are first time person who came gets a newbie badge, so to say, and somebody who checks in to the system late in the night, early in the morning, there were different badges that were that could be actually given in system slide force for mayorship.

Secondly, works already explain right. So, the text says badges or like medals given to userswhocheckinataspecificvenuesorachievesomepredefinednumberofcheck-ins.

(ReferSlideTime:23:07)



Users can post *tips* at specific venues, commenting on their previous experiences when visiting the corresponding physical places. Tips can also serve as feedback, recommendation or review to help other users choose places to visit. Examples of tips include the best option of a menu in a restaurant, the best place to have lunch in an airport, or even a complaint about a service. With a limitation of 200 characters, tips nourish the relationship between users and real businesses and may be a key feature to attract future visitors [17]. Each user has a history of all tips she posted, with associated venue and timestamp. When visiting a venues' page, after reading a previously posted tip, a user may mark it as *done* or *to-do*, in sign of agreement with the tip's content or intention to visit that location in the future, respectively. The

the nai
provid
help u
from v
to prov
ranging
names,
We als
phrase
tion as
orships
open t
the city
Moreo
venue'

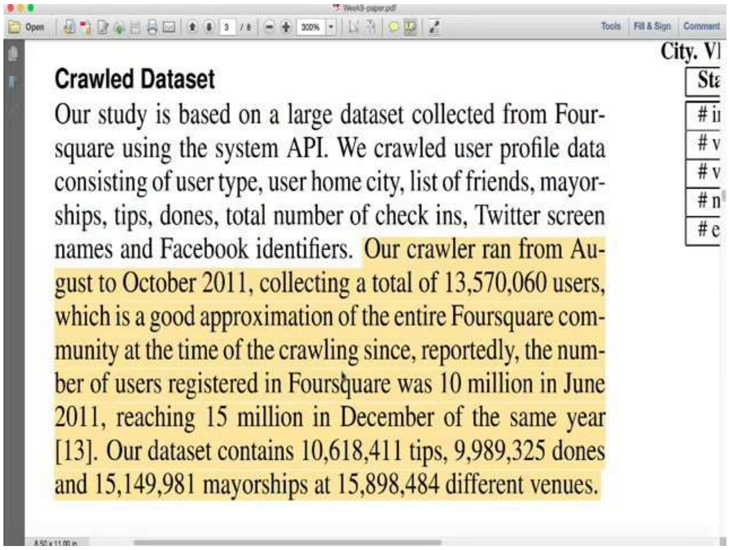
Tips as I said, users can post tips at specific venues commenting on their previous experiences when visiting the corresponding physical places. Tips can also serve as feedback recommendation or review to help others user other users choose places tovisit. So, the idea is you got to restaurant you have food you want to actually give feedback to others saying the food was good or the other way the food was bad. So,when others want to get to this restaurant they can actually use feedback to make it thatis a tip.

So, when visiting a venues page after reading a previously posted tip, the user may mark it as done or to do in sign for agreement. When the tips content or intention to visit the location in the future. So, it is essentiallysaying that I saw this tip the tip is actuallyvery useful for me. So, I take a button which is done or do you say that I want to keep it is to do which is I want to go to this place in future.

PrivacyinLocationBasedSocialNetworksPart2

Now,letuslookatthedatasetthatwascollected.

(Refer Slide Time: 00:12)



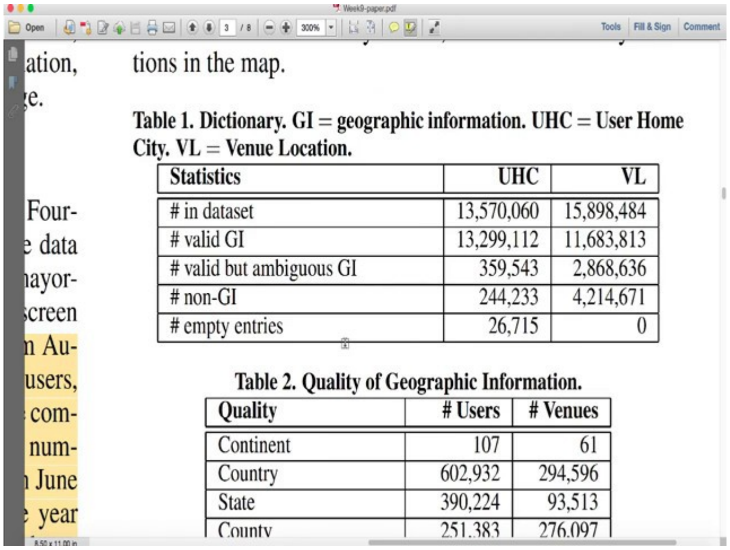
Crawled Dataset

Our study is based on a large dataset collected from Foursquare using the system API. We crawled user profile data consisting of user type, user home city, list of friends, mayorships, tips, dones, total number of check ins, Twitter screen names and Facebook identifiers. Our crawler ran from August to October 2011, collecting a total of 13,570,060 users, which is a good approximation of the entire Foursquare community at the time of the crawling since, reportedly, the number of users registered in Foursquare was 10 million in June 2011, reaching 15 million in December of the same year [13]. Our dataset contains 10,618,411 tips, 9,989,325 dones and 15,149,981 mayorships at 15,898,484 different venues.

So, the dataset is about total of about 13 million users collected ran from August to October 2011, the total of 13 million users. It is almost close to the entire Foursquare in terms of the number users that are using it, and the total dataset contains 10 million tips and what we are interested in the data particularly the questions that we are asking as I said, we are interested in mostly the tips, dones and mayorships, because that is the information that is publicly available.

What can you use this information for in terms of the actually getting the location of the particular user. It is the total number of tips that are available in the dataset is about 10 million, total number of dones is about 9million, and mayor ship is about 15 million and different venues that are available are about 15 million again.

(ReferSlideTime:01:56)



ation,
e.

Four-
e data
nayor-
creen
n Au-
users,
com-
num-
n June
e year

tions in the map.

Table 1. Dictionary. GI = geographic information. UHC = User Home City. VL = Venue Location.

Statistics	UHC	VL
# in dataset	13,570,060	15,898,484
# valid GI	13,299,112	11,683,813
# valid but ambiguous GI	359,543	2,868,636
# non-GI	244,233	4,214,671
# empty entries	26,715	0

Table 2. Quality of Geographic Information.

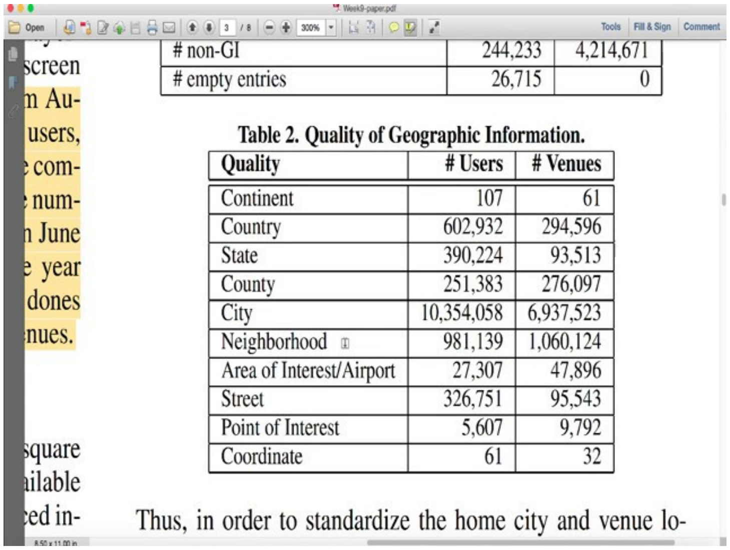
Quality	# Users	# Venues
Continent	107	61
Country	602,932	294,596
State	390,224	93,513
County	251,383	276,097

So, let us just look at just characterization of this data, which is what kind ofin generally about the dataset that is available. So, we will look at everytable and every figure in this paper.Also here if you look at the number in the dataset is 13 million UHC. UHC stands for user home city; VL stands for venue location, the number of venues that areavailable; and GI stands for geographic information right.

Of course, there is going to be some information, some locations which are not going to be valid right for example, something in the middle of sea, you are not going to get any location, and there are locations that may be generated which is somebody's heart right,h e a r t. So, these kind of locations has to be removed that is what happened between number in the dataset and valid GI; valid, but ambiguous which is it is valid, but we are not able to figure out the exact location, reverse look up, and find out the location that falls into the third row.

Non- geographic information and empty entries, so essentially the dataset was pruned to get data which the researchers can actuallyuseto do the analysis right.This is what even youwoulddoforthehomeworksthatyoudidyoucollected somedata,butyouprobably did not do the way to actually prune the data to get more accurate, more specific data.

(ReferSlideTime:04:00)



screen
m Au-
users,
e com-
e num-
n June
e year
dones
nues.

square
ailable
ced in-

# non-GI	244,233	4,214,671
# empty entries	26,715	0

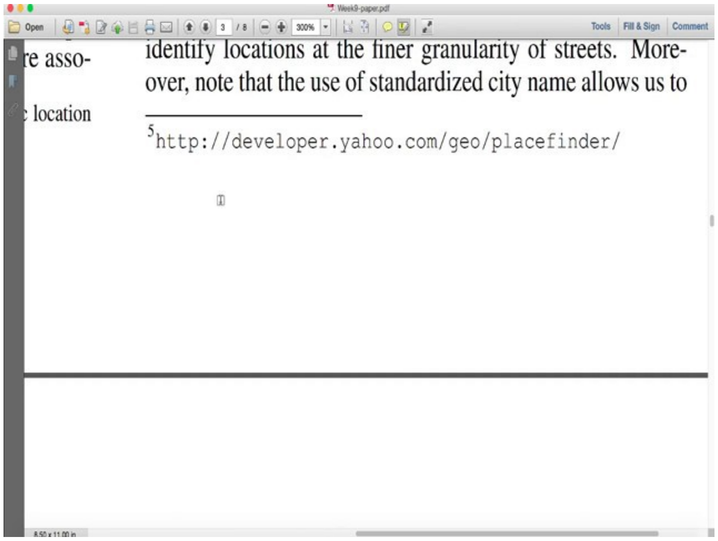
Table 2. Quality of Geographic Information.

Quality	# Users	# Venues
Continent	107	61
Country	602,932	294,596
State	390,224	93,513
County	251,383	276,097
City	10,354,058	6,937,523
Neighborhood	981,139	1,060,124
Area of Interest/Airport	27,307	47,896
Street	326,751	95,543
Point of Interest	5,607	9,792
Coordinate	61	32

Thus, in order to standardize the home city and venue lo-

Also the quality of geographic information is continent, countries, state and this information that is available in this dataset is number of users, number of venues. Country, state, county, city, neighborhood, area of interest or airport, street, point of interest and coordinates. So, all this pieces of information you will get in your json when you collect data from foursquare.And this was basically pruned to get more quality data which can be used for analysis, is that making sense. So, these are called exploratorydata analysis, here we just explaining the data itself describing the data in terms of what is available in the data that was collected.

(ReferSlideTime:05:02)



re asso-
location

identify locations at the finer granularity of streets. Moreover, note that the use of standardized city name allows us to

[5] http://developer.yahoo.com/geo/placefinder/

Multiple tools were used. So, let me just show you one of them which are developer dot yahoo dot com slash geo slash placefinder. These kind of tools lets you actually reverse look up a place and find out where they are in the map; if you give them location, it can actually give you the latitude, longitude even the other way round, you give the latitude longitude it will give you the location.

(ReferSlideTime:03:35)

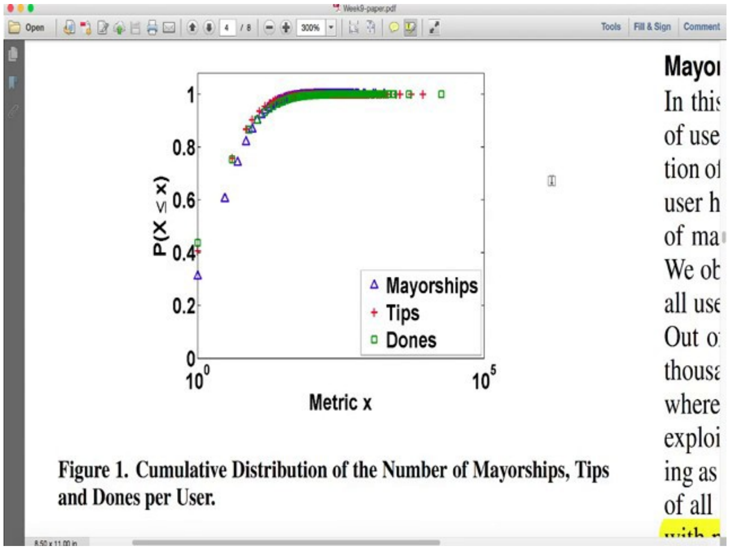
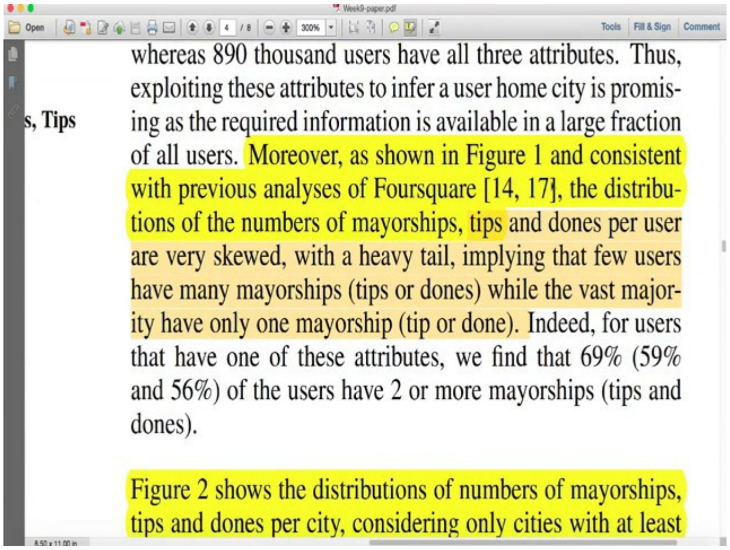


Figure 1. Cumulative Distribution of the Number of Mayorships, Tips and Dones per User.

Mayo
In this
of use
tion o
user h
of ma
We ob
all use
Out o
thous
where
exploi
ing as
of all
with r

Many of such tools were used in this particular research to find out the location of the checkins,locationoftips,andotherinformationthatwascollected.So,ifyoulookatthe rest of the analysis, so here is the two figures that we will also talk about; first let us talk about the figure 1, which is shown on the left, but let us look at the content.

(ReferSlideTime:06:03)



whereas 890 thousand users have all three attributes. Thus, exploiting these attributes to infer a user home city is promising as the required information is available in a large fraction of all users. Moreover, as shown in Figure 1 and consistent with previous analyses of Foursquare [14, 17], the distributions of the numbers of mayorships, tips and dones per user are very skewed, with a heavy tail, implying that few users have many mayorships (tips or dones) while the vast majority have only one mayorship (tip or done). Indeed, for users that have one of these attributes, we find that 69% (59% and 56%) of the users have 2 or more mayorships (tips and dones).

Figure 2 shows the distributions of numbers of mayorships, tips and dones per city, considering only cities with at least

For the figure 1, as shown in the figure 1 and consistent with previous analysis of Foursquarethedistributionofnumberofmayorships,tipsanddones.So,thisiswhatyou do first when you collect such data. So, one other things that you would have generally seen in social media analysis is to show whether the data is a power law or show overthat the data that you have collected from social media follows the pareto principle. which is to say that 20 percent of the users only actually contribute to the 80 percent of the content that is generated on the social media.

One of the similar type of graph was drawn here which is to see the distribution of mayorships, tips and dones per users and the inferences that they are skewed, with a heavy tail, implying that few users have many mayorships - tips or dones, while vast majority of them have only one mayorship, tip or done. So, that is essentially what a Pareto principle is that is essentially, what power law is also.

(ReferSlideTime:07:30)

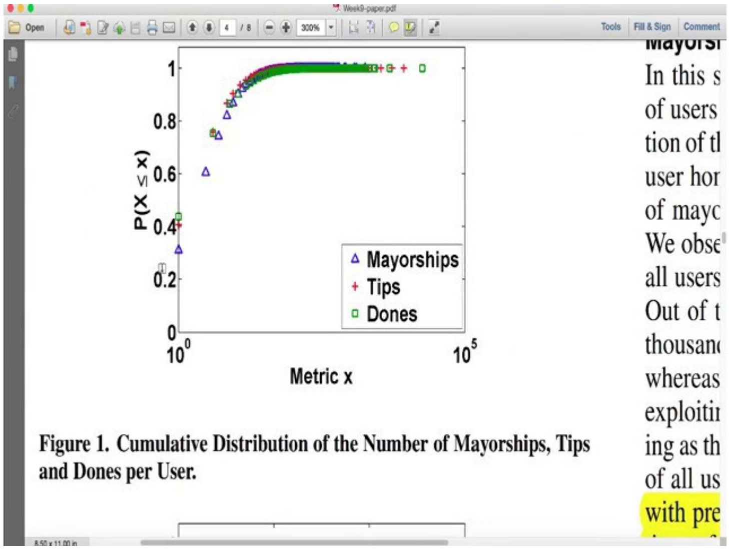
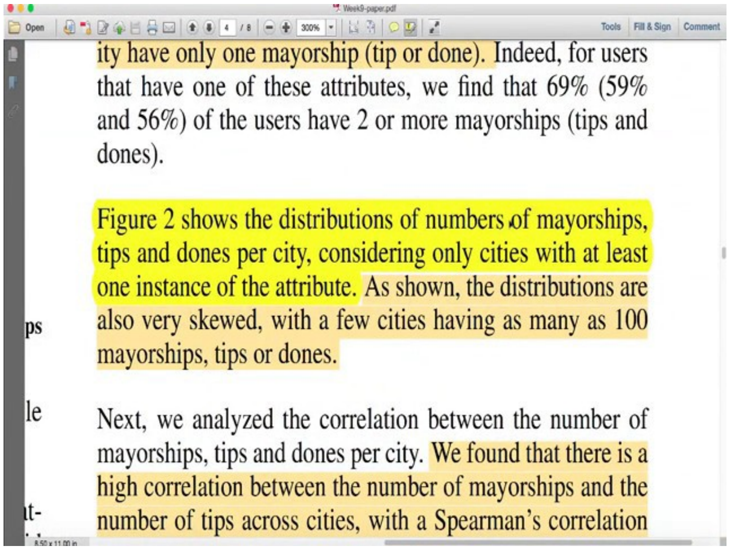


Figure 1. Cumulative Distribution of the Number of Mayorships, Tips and Dones per User.

Mayorsh
In this s
of users
tion of tl
user hor
of mayo
We obse
all users
Out of t
thousan
whereas
exploitir
ing as th
of all us
with pre

So, if you look at the graph which is on the left for the figure 1, you will see the same thing which is large amount of, so large amount of users actually have small amount of users have so that is what this here. So, let us take figure 90 percent of the data is getting generated by small set of users; majority of the users over here do not contribute to anyof the mayorships, tips or dones, so that is the graph that you would read.

(ReferSlideTime:08:07)



ity have only one mayorship (tip or done). Indeed, for users that have one of these attributes, we find that 69% (59% and 56%) of the users have 2 or more mayorships (tips and dones).

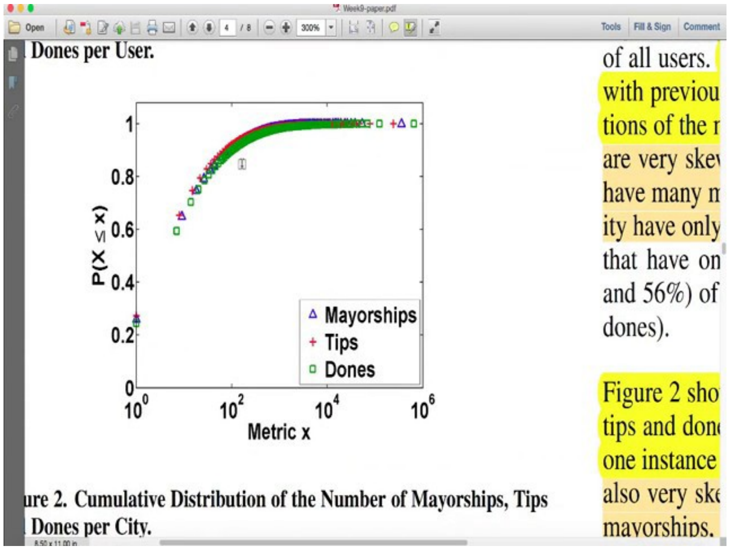
Figure 2 shows the distributions of numbers of mayorships, tips and dones per city, considering only cities with at least one instance of the attribute. As shown, the distributions are also very skewed, with a few cities having as many as 100 mayorships, tips or dones.

Next, we analyzed the correlation between the number of mayorships, tips and dones per city. We found that there is a high correlation between the number of mayorships and the number of tips across cities, with a Spearman's correlation

Similarly,let us go to figure 2.Figure 2shows the distribution ofnumber of mayorships, tipsanddonespercity,consideringonlycitieswithatleastoneinstanceoftheattribute.

So, it is actually wanted to see whether from a city, whether you are able to get a lot of mayorship, tips and dones, this can actually help some, help find out how the data is.As shown the distributions are also very skewed, with a few cities having as many as 100 mayorship tips and dones.

(ReferSlideTime:08:47)



Dones per User.

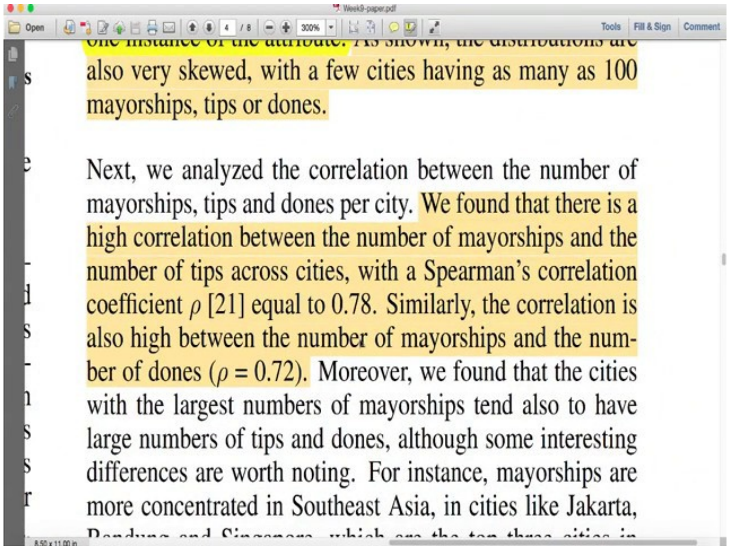
ure 2. Cumulative Distribution of the Number of Mayorships, Tips
Dones per City.

of all users.
with previou
tions of the r
are very skev
have many n
ity have only
that have on
and 56%) of
dones).

Figure 2 sho
tips and done
one instance
also very ske
mayorships.

So, again if you look at figure 2, the distribution is very similar in terms of the small number of cities having large number of mayorship, tips and dones; and large number of cities, do not have these. So, these are all large, so if you look at it somewhere around 80 or90yearssomethingthatisonlyasmallsetofcitieshere.Largeamountofcitiesdonot have mayorship tips and dones or very little as many little mayorship, tips and dones because the condition was that they were considering only cities with at least one instance of the attribute, which is they should have had one tip, mayorship or done.

(ReferSlideTime:09:39)

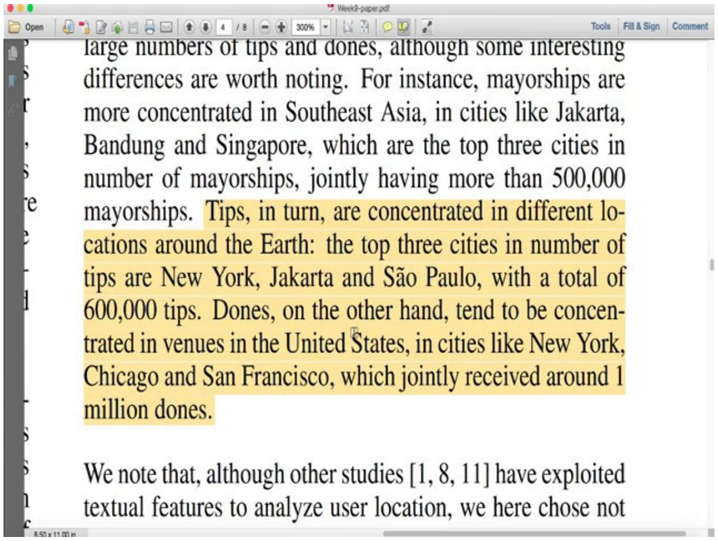


one instance of the attribute. As shown, the distributions are also very skewed, with a few cities having as many as 100 mayorships, tips or dones.

Next, we analyzed the correlation between the number of mayorships, tips and dones per city. We found that there is a high correlation between the number of mayorships and the number of tips across cities, with a Spearman's correlation coefficient ρ [21] equal to 0.78. Similarly, the correlation is also high between the number of mayorships and the number of dones ($\rho = 0.72$). Moreover, we found that the cities with the largest numbers of mayorships tend also to have large numbers of tips and dones, although some interesting differences are worth noting. For instance, mayorships are more concentrated in Southeast Asia, in cities like Jakarta, Bandung and Singapore, which are the top three cities in

So,thenlookingatcorrelationbetweenthenumberofmayorship,tipsanddonespercity, they found that a high correlation between the number of mayorship and the number tips across cities, with the coefficient of 0.78. Similarly, the correlation is also high between number of mayorships and the number of dones, which is if there are more mayorships there in a city that is high chance that there will more of tips and dones also. This is helping us to understand that where if I find cities which I have a high mayorship, I should be able to find, there should be more of tips and dones also there.

(ReferSlideTime:10:34)



large numbers of tips and dones, although some interesting differences are worth noting. For instance, mayorships are more concentrated in Southeast Asia, in cities like Jakarta, Bandung and Singapore, which are the top three cities in number of mayorships, jointly having more than 500,000 mayorships. Tips, in turn, are concentrated in different locations around the Earth: the top three cities in number of tips are New York, Jakarta and São Paulo, with a total of 600,000 tips. Dones, on the other hand, tend to be concentrated in venues in the United States, in cities like New York, Chicago and San Francisco, which jointly received around 1 million dones.

We note that, although other studies [1, 8, 11] have exploited textual features to analyze user location, we here chose not

So, if you look at the tips, tips are concentrated in different location around the Earth; which is if you remember tips of the content that people post for a particular venue. The top 3 cities in the number tips are New York, Jakarta and Sao Paulo with the total of 600,000 tips. Dones, on the other hand, tend to be concentrated in venues in the US, in cities like first New York, Chicago and San Francisco, and total they have about 1 million dones, so which is to show that some cities, some popular cities have a lot of these tips and donesNewYork being common in both tips and dones. Once again to say that cities generate a lot of these tips mayorships and dones and of course, this wouldalso probably lead in to check ins also.

(ReferSlideTime:11:55)

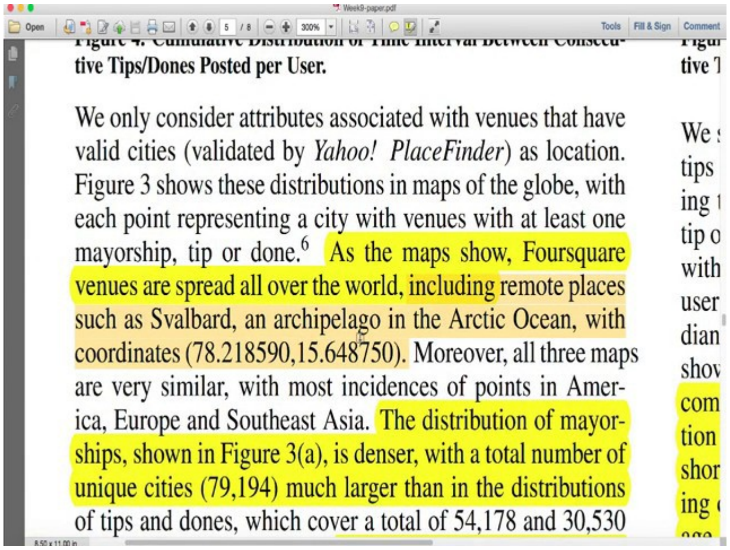
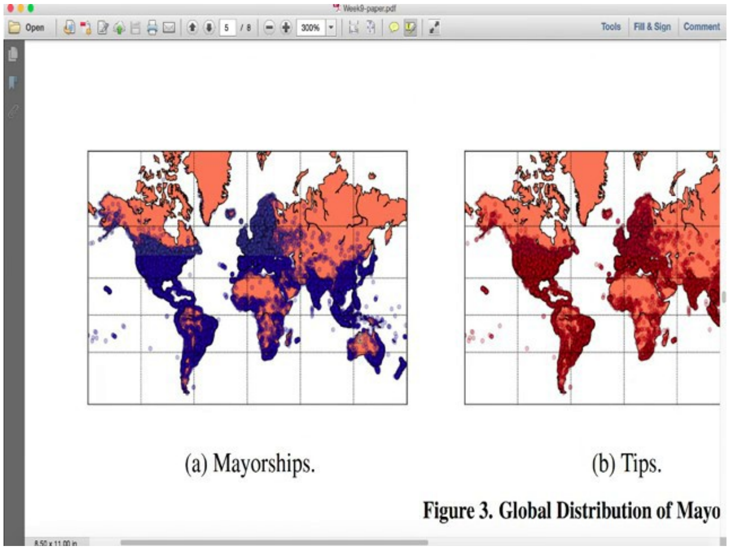


Figure 4. Cumulative Distribution of Time Interval Between Consecutive Tips/Dones Posted per User.

We only consider attributes associated with venues that have valid cities (validated by *Yahoo! PlaceFinder*) as location. Figure 3 shows these distributions in maps of the globe, with each point representing a city with venues with at least one mayorship, tip or done.[6] As the maps show, Foursquare venues are spread all over the world, including remote places such as Svalbard, an archipelago in the Arctic Ocean, with coordinates (78.218590,15.648750). Moreover, all three maps are very similar, with most incidences of points in America, Europe and Southeast Asia. The distribution of mayorships, shown in Figure 3(a), is denser, with a total number of unique cities (79,194) much larger than in the distributions of tips and dones, which cover a total of 54,178 and 30,530

So, here is another graph which we will see in figure 3 and figure 3 for that we will see. So,figure3showsthesedistributionsinmapsoftheglobewitheachpointerrepresenting a city with venues, with at least one mayorship tip and done. So, essentially until nowwe only saw per city what is happening? Now when you look at it in a map, the figure 3 actually shows the results. As the maps show, Foursquare venues are spread all over the world including remote places such as Svalbard, an archipelago in the Arctic Ocean.

(ReferSlideTime:12:41)



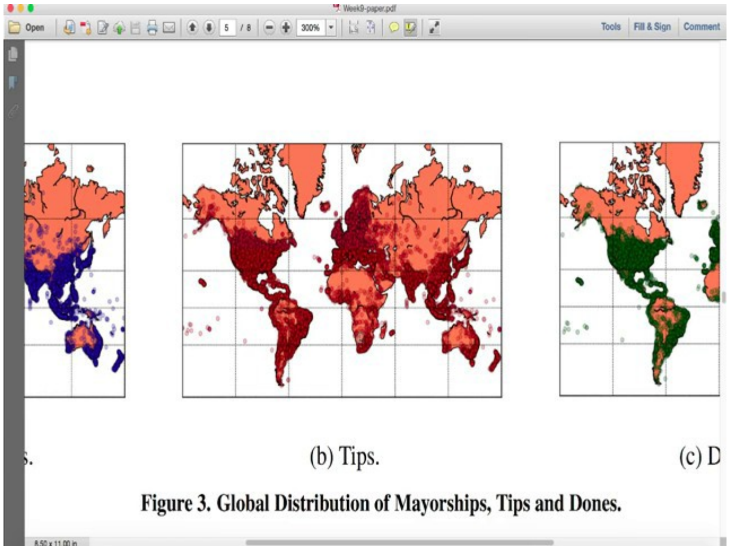
(a) Mayorships.

(b) Tips.

Figure 3. Global Distribution of Mayo

So,forexample,letusgolookatthefigure3a,bandc.So,thisisthemayorship.Thisis basically showing you every dot in this graph, every blue dot in this graph, figure 3(a) shows you the distribution of the mayorships that are available around the world, thatwere done around the world.

(ReferSlideTime:13:00)



s.

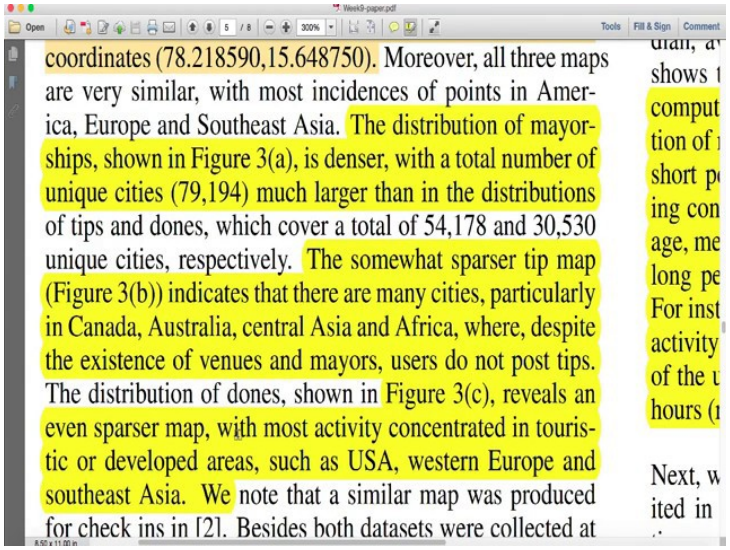
(b) Tips.

(c) D

Figure 3. Global Distribution of Mayorships, Tips and Dones.

Figure 3(b) shows you the tips that were done. So, if you see earlier, we saw that the there is high correlation between mayorship, tips and dones. So, therefore, when mayorshipsarehighthereisgoingtobetipsanddonesalsowhicharehigh.So,youcan clearly see heavy concentration on many places in the world. And the last one is dones, the figure 3(c) shows you green dot, every green dot is a done from that particular locations. So, this basically shows you mayorships of the blue dot, tips - the red dot, and the green dot being dones. So that is figure 3 is denser with the total number of unique cities.

(ReferSlideTime:13:48)



coordinates (78.218590,15.648750). Moreover, all three maps are very similar, with most incidences of points in America, Europe and Southeast Asia. The distribution of mayorships, shown in Figure 3(a), is denser, with a total number of unique cities (79,194) much larger than in the distributions of tips and dones, which cover a total of 54,178 and 30,530 unique cities, respectively. The somewhat sparser tip map (Figure 3(b)) indicates that there are many cities, particularly in Canada, Australia, central Asia and Africa, where, despite the existence of venues and mayors, users do not post tips. The distribution of dones, shown in Figure 3(c), reveals an even sparser map, with most activity concentrated in touristic or developed areas, such as USA, western Europe and southeast Asia. We note that a similar map was produced for check ins in [2]. Besides both datasets were collected at

shows t
comput
tion of
short p
ing con
age, me
long pe
For inst
activity
of the u
hours (

Next, w
ited in

So, the distribution of mayorships shown in figure 3(a) is denser with a total number of unique cities being 79,000. So, 79,000 cities have higher check ins, have mayorships in the figure 3(a) in the total data. And if you look at the figure 3(b), somewhat sparser tip map for figure 3(b) indicates that there are manycities, particularlyin Canada,Australia, and Central Asia, and Africa, where despite their insistence of venues and mayors' users do not post tips.

So, therefore, there is a chance that there are mayorships in that location, but not tips. Figure 3(c), reveals an even sparser map, with most activity concentrated in touristic or developed areas, such as USA, Western Europe and Southeast Asia. So, essentially even though there is a correlation between mayorships, tips and dones, there is actually some places which are sparser for a tips and dones.

(ReferSlideTime:15:33)

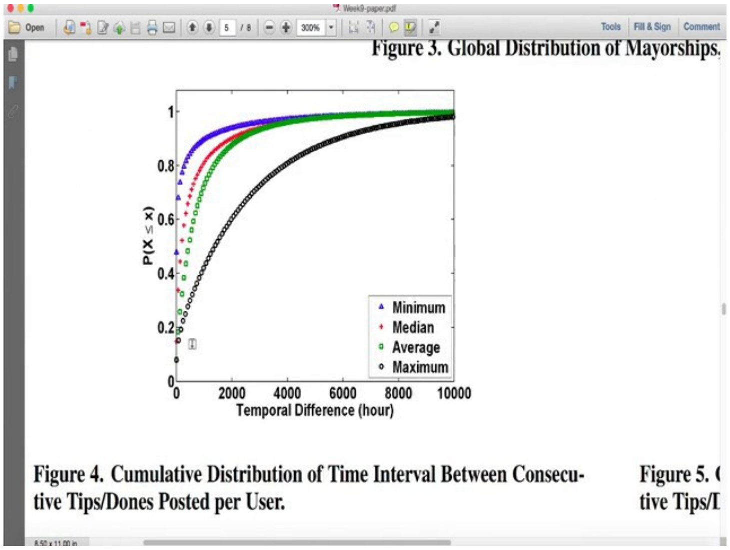


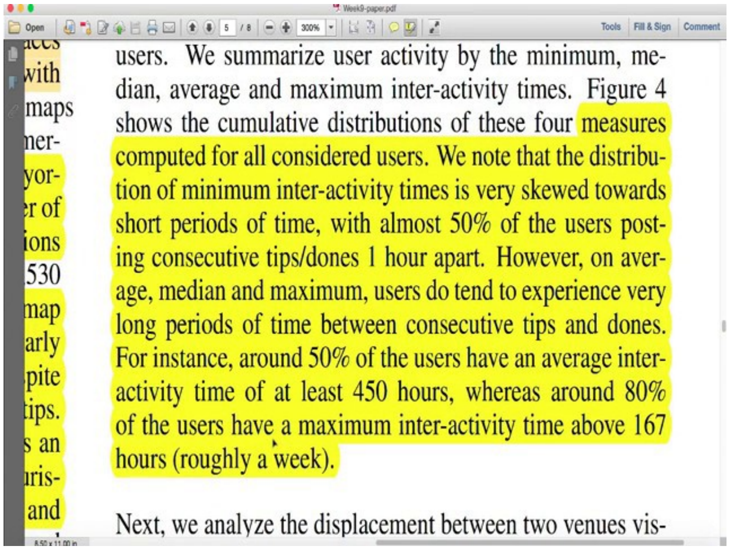
Figure 3. Global Distribution of Mayorships.

Figure 4. Cumulative Distribution of Time Interval Between Consecutive Tips/Dones Posted per User.

Figure 5. (
tive Tips/I

So, now let us look at figure 4; figure 4 here, so this is figure 4. Figure 4 is showing you the cumulative distribution of time interval between consecutive tips and dones posted per user. Why is this interesting? This is interesting to find out about the activity or the frequency of activity of the users.

(ReferSlideTime:16:07)



aces
with
maps
mer-
yor-
r of
ions
530
map
arly
pite
tips.
s an
uris-
and

users. We summarize user activity by the minimum, median, average and maximum inter-activity times. Figure 4 shows the cumulative distributions of these four measures computed for all considered users. We note that the distribution of minimum inter-activity times is very skewed towards short periods of time, with almost 50% of the users posting consecutive tips/dones 1 hour apart. However, on average, median and maximum, users do tend to experience very long periods of time between consecutive tips and dones. For instance, around 50% of the users have an average inter-activity time of at least 450 hours, whereas around 80% of the users have a maximum inter-activity time above 167 hours (roughly a week).

Next, we analyze the displacement between two venues vis-

This is figure 4, figure 4 shows the cumulative distributions of these four measures. We note that the distribution of minimum inter-activity times is very skewed towards short periodsoftimes,withthealmost50percentoftheuserspostingconsecutivetipsand dones 1 hour apart, got it. So, that is it shows the there is a lot of content that are generated, lot of tips are generated byusers,tips and dones are generated byusers within apart 1 hour. However, an average, median and maximum users do tend to experience very long periods of times between consecutive tips and dones.

So, essentially what this shows is again it is going back to the same power law concept, there are some set of users where there is consecutive tips and dones are done very frequently. There is set of population where this distribution is actually pretty skewed, which is long set of long time taken between two consecutive tips and dones. For instance, around 50 percent of the users have an average interactivitytime of at least 450 hours that is close to about 20 days, whereas around 80 percent of the users have the maximum interactivity of 167 hours - roughly a week.

(ReferSlideTime:17:51)

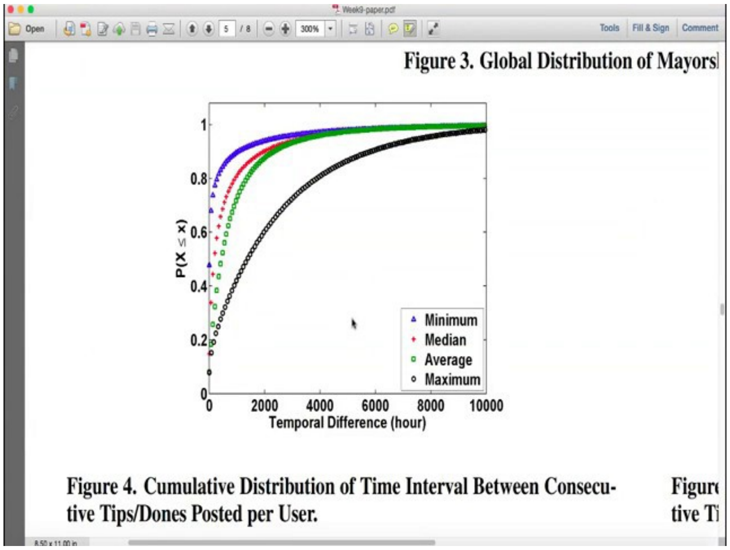


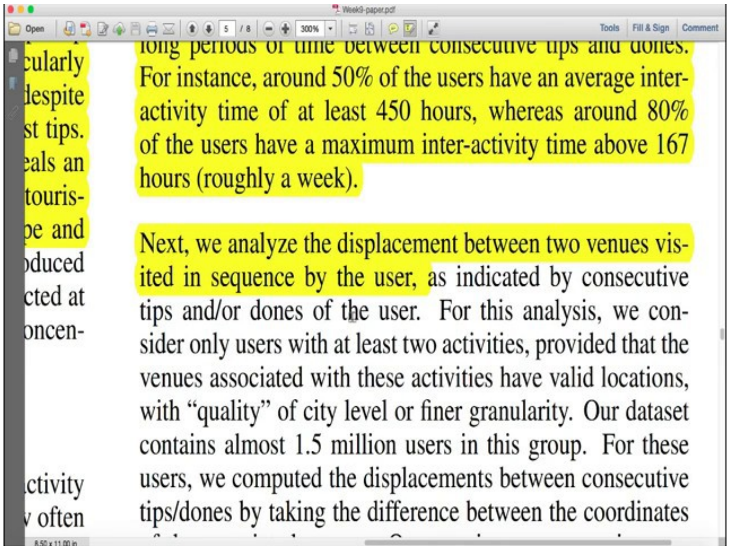
Figure 3. Global Distribution of Mayorsl

Figure 4. Cumulative Distribution of Time Interval Between Consecutive Tips/Dones Posted per User.

Figur
tive Ti

So, here is the graph for cumulative distribution of time interval between consecutivetips and dones posted per user, it's the same thing as what we saw in the text.

(ReferSlideTime:18:22)



cularly
despite
st tips.
eals an
touris-
pe and
oduced
cted at
oncen-

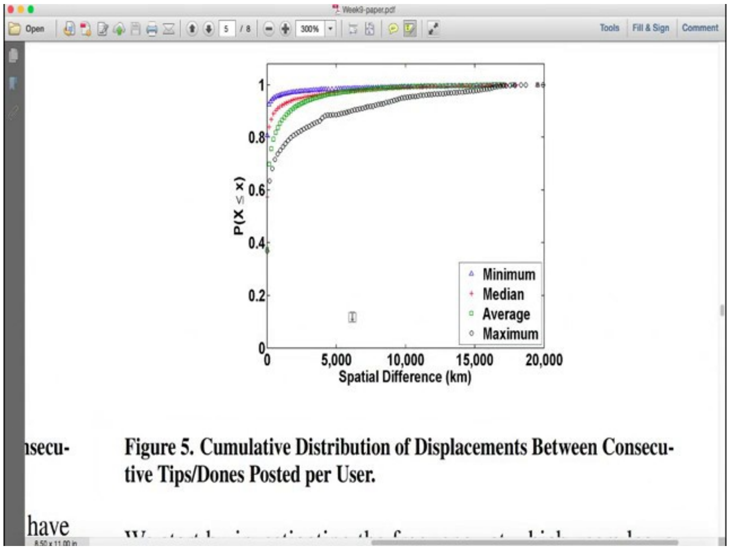
ctivity
often

long periods of time between consecutive tips and dones. For instance, around 50% of the users have an average inter-activity time of at least 450 hours, whereas around 80% of the users have a maximum inter-activity time above 167 hours (roughly a week).

Next, we analyze the displacement between two venues visited in sequence by the user, as indicated by consecutive tips and/or dones of the user. For this analysis, we consider only users with at least two activities, provided that the venues associated with these activities have valid locations, with "quality" of city level or finer granularity. Our dataset contains almost 1.5 million users in this group. For these users, we computed the displacements between consecutive tips/dones by taking the difference between the coordinates

Now, let us look at the next figure, figure 5. So, this part we already saw. Figure 5 is basically looking at the same question which is now we are analyzing the displacement between two venues. The last figure that we saw was looking at two different timing in whichthepostwasdone.Nowwearelookingattwodifferentvenues,whicharedoneby the same user consecutively.

(ReferSlideTime:18:58)



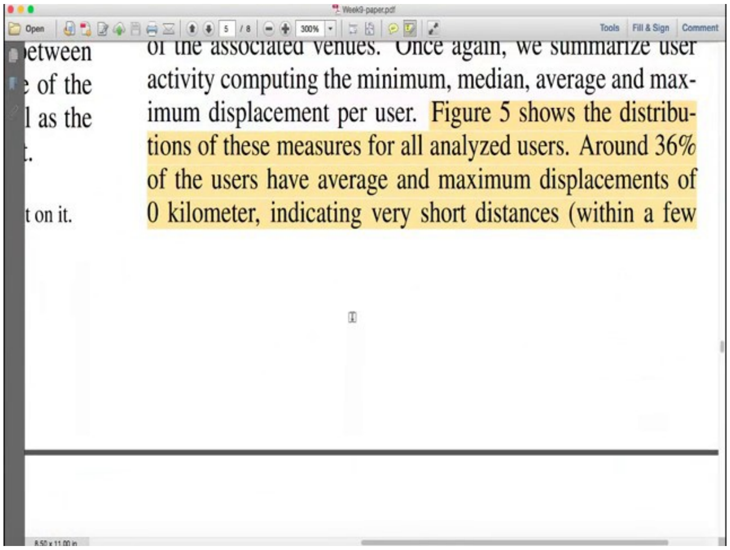
nsecu-

have

Figure 5. Cumulative Distribution of Displacements Between Consecutive Tips/Dones Posted per User.

So, let us look at the figure 5, first and I will tell you what the figure 5 all means. So, figure5isthecumulativedistributionofdisplacementbetweenconsecutivetipsand dones posted per user.So, on thex-axis itis showing you thedistance; on they axis, itis showing you the number of the distribution.

(ReferSlideTime:19:27)



etween
of the
l as the
.
t on it.

of the associated venues. Once again, we summarize user activity computing the minimum, median, average and maximum displacement per user. Figure 5 shows the distributions of these measures for all analyzed users. Around 36% of the users have average and maximum displacements of 0 kilometer, indicating very short distances (within a few

Figure 5 shows the distributions of these measures for all analyzed users. Around 36 percent of the users have average and maximum displacements of about 0 kilometers, right, indicating very short distances - within a few meters.

(ReferSlideTime:19:49)

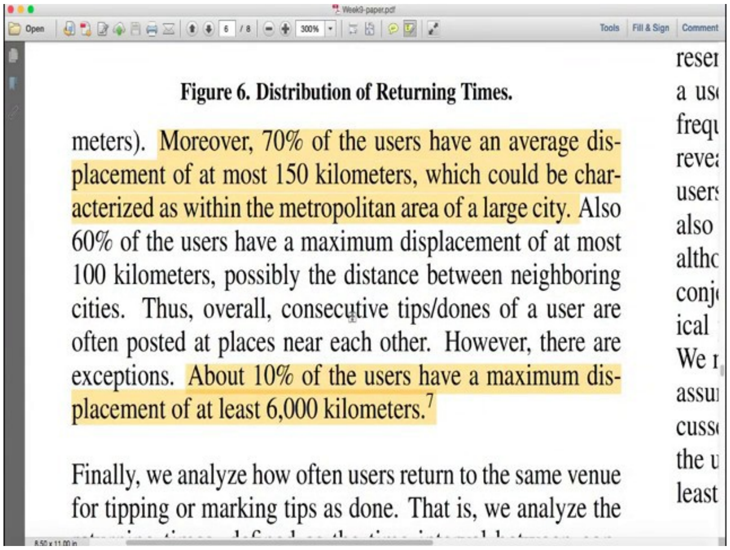


Figure 6. Distribution of Returning Times.

meters). Moreover, 70% of the users have an average displacement of at most 150 kilometers, which could be characterized as within the metropolitan area of a large city. Also 60% of the users have a maximum displacement of at most 100 kilometers, possibly the distance between neighboring cities. Thus, overall, consecutive tips/dones of a user are often posted at places near each other. However, there are exceptions. About 10% of the users have a maximum displacement of at least 6,000 kilometers.[7]

Finally, we analyze how often users return to the same venue for tipping or marking tips as done. That is, we analyze the

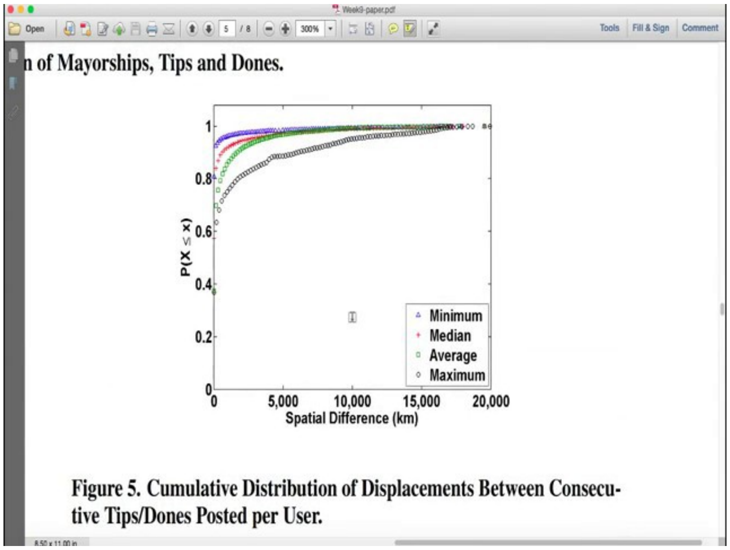
reser
a us
frequ
revea
users
also
altho
conj
ical
We r
assu
cuss
the u
least

70 percentof theusershavean averagedisplacementofatmost150 kilometers, which is basicallysomebodymovingbetweencities.So,IprobablyfromDelhiIgotoAgra,and

then I do it check in all Mathura, I do check in Mathura, I do a tip or a done, that is what is actually capturing within 150 kilometers.

And about ten percent of the users have a maximum displacement of about 6000 kilometers,thisisprobablyinternationaltravelbetweentwoconsecutivetipsordones,so that shows what is the distribution of the users who we have in the dataset, the consecutive tips and dones that they do on Foursquare. Let us go to the figure again. So, this is basically showing you that 70 percent of the users are about 150 kilometers andthe 10 percent is about the 7000 kilometers that is what you will see in this figure.

(ReferSlideTime:21:09)

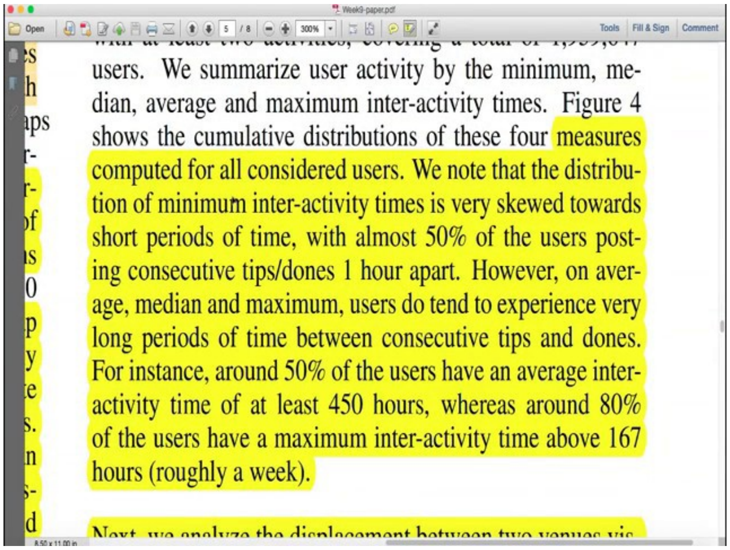


n of Mayorships, Tips and Dones.

Figure 5. Cumulative Distribution of Displacements Between Consecutive Tips/Dones Posted per User.

So, if you seen here 7000 kilometers, so it is about the last 10 percent of the users, who are about 7000 kilometers and then very short is about 70 percent.Average 70 percent is that the green one is the average, the green square is the average. The red plus symbol is amedian,triangleistheminimum,andthecircleis themaximumright,sothatgivesyou a sense of.

(ReferSlideTime:21:51)



users. We summarize user activity by the minimum, median, average and maximum inter-activity times. Figure 4 shows the cumulative distributions of these four measures computed for all considered users. We note that the distribution of minimum inter-activity times is very skewed towards short periods of time, with almost 50% of the users posting consecutive tips/dones 1 hour apart. However, on average, median and maximum, users do tend to experience very long periods of time between consecutive tips and dones. For instance, around 50% of the users have an average inter-activity time of at least 450 hours, whereas around 80% of the users have a maximum inter-activity time above 167 hours (roughly a week).

Next, we analyze the displacement between two venues vis-

So, now,what all we have seen, we have seen the time, consecutive, tips or dones that is done with respect to time, consecutive tips and dones with respect to distance.

(ReferSlideTime:22:18)

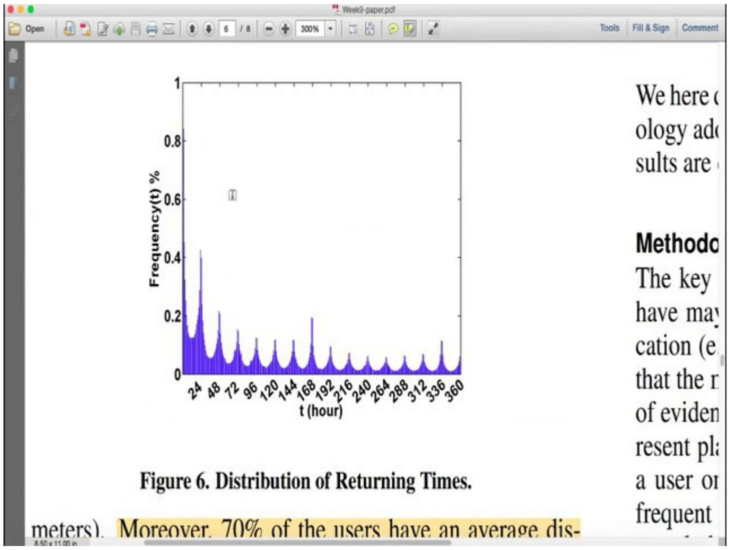


Figure 6. Distribution of Returning Times.

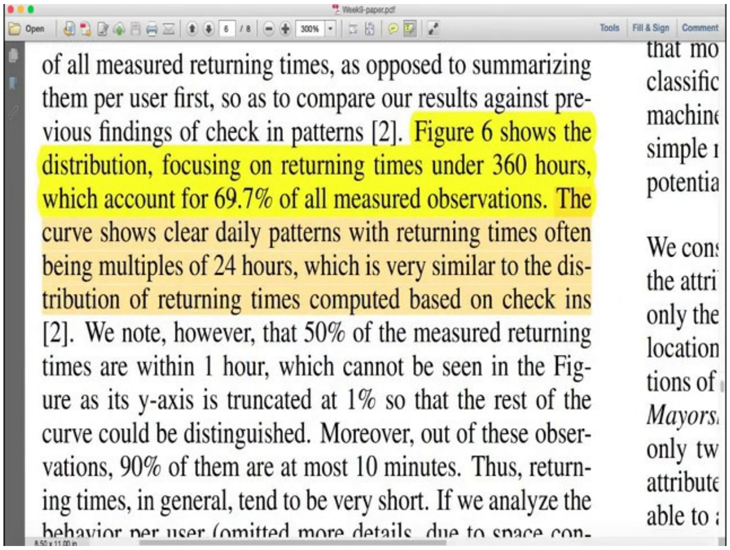
meters). Moreover, 70% of the users have an average dis-

We here c
ology adc
sults are

Methodo
The key
have may
cation (e
that the r
of eviden
resent pla
a user or
frequent

Now we will see the next figure, next figure is actually very interesting. Next analysis is actually a very interesting analysis, where they saw how frequently that the check ins,the tips or the dones are coming back for that particular location. So, this is distribution of the returning time. So, if I do tip or a done inIIIT, how frequently do I actually do a tip or a done in that location.

(ReferSlideTime:22:52)



of all measured returning times, as opposed to summarizing them per user first, so as to compare our results against previous findings of check in patterns [2]. Figure 6 shows the distribution, focusing on returning times under 360 hours, which account for 69.7% of all measured observations. The curve shows clear daily patterns with returning times often being multiples of 24 hours, which is very similar to the distribution of returning times computed based on check ins [2]. We note, however, that 50% of the measured returning times are within 1 hour, which cannot be seen in the Figure as its y-axis is truncated at 1% so that the rest of the curve could be distinguished. Moreover, out of these observations, 90% of them are at most 10 minutes. Thus, returning times, in general, tend to be very short. If we analyze the behavior per user (omitted more details due to space con-

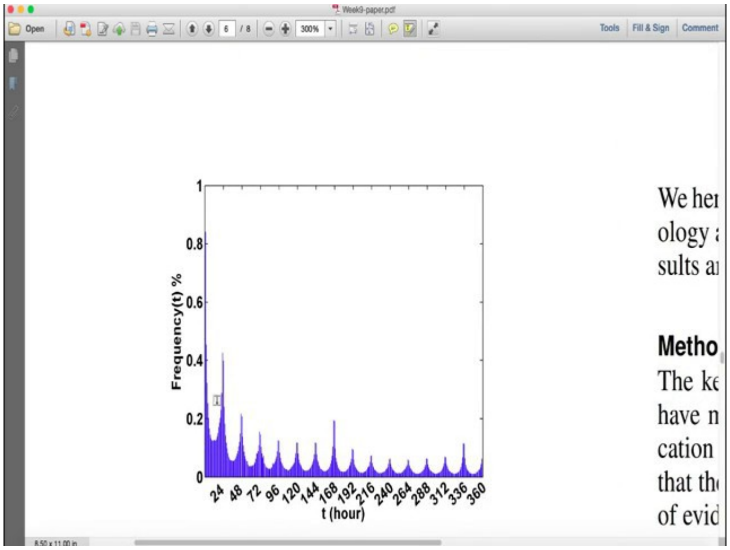
that mo
classific
machin
simple
potentia

We con
the attri
only the
location
tions of
Mayors.
only tw
attribute
able to

And the figure 6 shows the distribution focusing on returning times under 360 hours, which account for 69.7 percent of all the measured observations. The curve showsclearly daily patterns with returning times often being multiples of 24 hours which isvery similar to the distribution of returning times computed based on the check ins.

So, if you really look at what does it mean why is it 24 hours? If somebody checks into office in the morning today that they have got into the office, shop, institute and everything they do the same check in the next day, so that is what this means right. Checkins,thereferencesisgiventoanotherresearchwherethecheckinswhereseenbut in this case we are also looking at the tips or the dones. That is a very interesting conclusion to know or basically it is complimenting the real world behavior that you could expect from the users.

(ReferSlideTime:24:02)

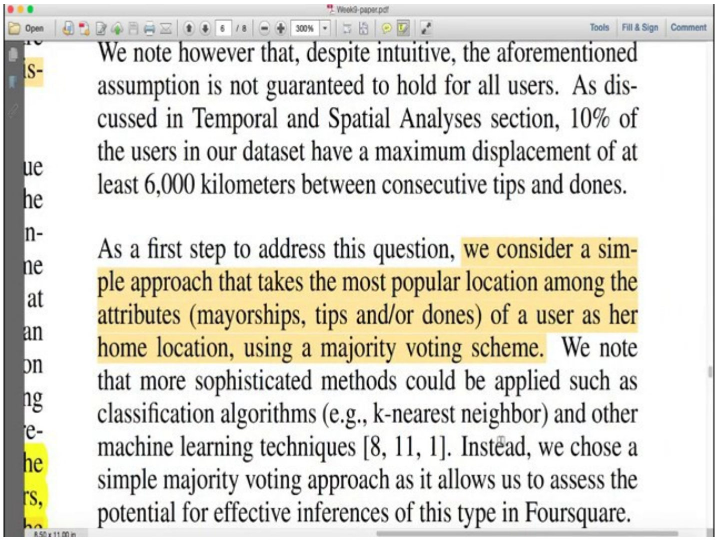


We her
ology a
sults a

Metho
The ke
have n
cation
that th
of evid

Here is a graph, which you see here this is about 360 hours.And you can clearly see that it's coming back. So, this is for everyday24, 48, 72, 96 120, so it is kind of coming back every 24 hours and sometimes the frequency is also increasing for something happens. So, this is 168 should be the week. So, therefore, there is a slighting increase from the day that which is which is the 7th day of tips and done. I hope that is making sense essentially the conclusion there is that people come back to that same location with the examples like me doing it in IIIT Delhi, that is figure 6 that is an interesting analysis.

(ReferSlideTime:25:09)



We note however that, despite intuitive, the aforementioned assumption is not guaranteed to hold for all users. As discussed in Temporal and Spatial Analyses section, 10% of the users in our dataset have a maximum displacement of at least 6,000 kilometers between consecutive tips and dones.

As a first step to address this question, we consider a simple approach that takes the most popular location among the attributes (mayorships, tips and/or dones) of a user as her home location, using a majority voting scheme. We note that more sophisticated methods could be applied such as classification algorithms (e.g., k-nearest neighbor) and other machine learning techniques [8, 11, 1]. Instead, we chose a simple majority voting approach as it allows us to assess the potential for effective inferences of this type in Foursquare.

So, now what we will do is, now we will attack the question that we started off with which is to find out how much can be actually inferred about the users' home using this data. So, in this case, we are going to actually use data for most popular location among mayorships, tips and dones of a user or home location using a majority voting scheme. I am going to explain to you at this voting scheme is there are two tables that we will look at and,

(ReferSlideTime:26:01)

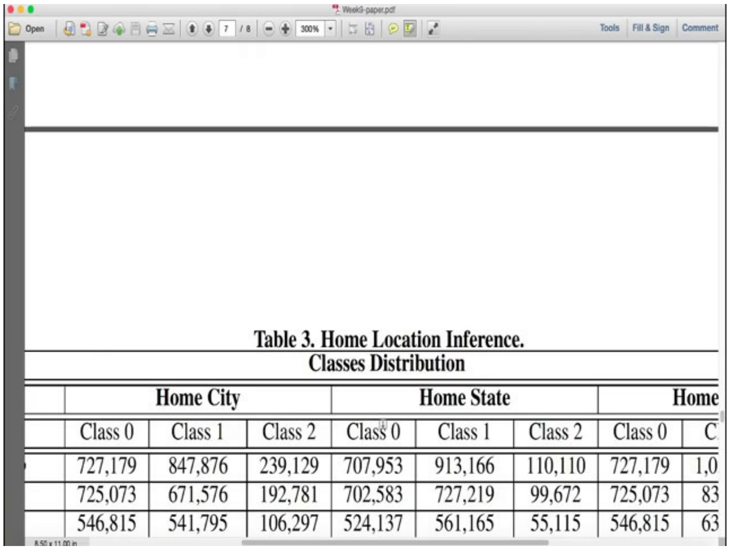


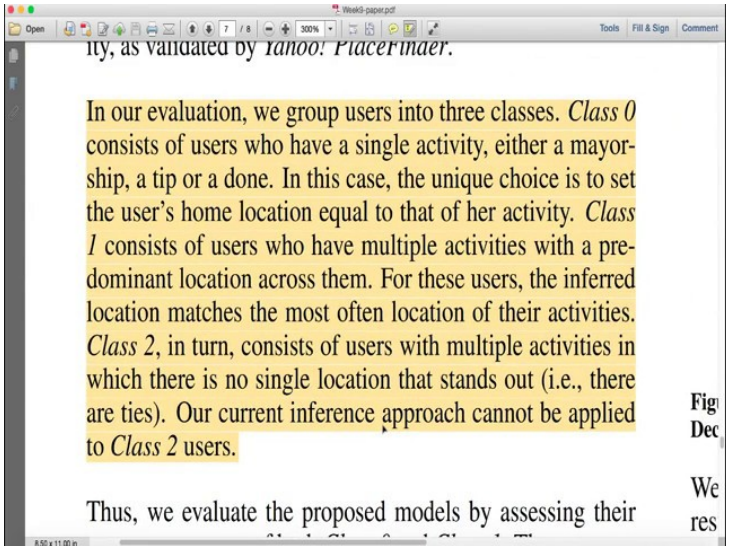
Table 3. Home Location Inference.

Classes Distribution

	Home City			Home State			Home	
	Class 0	Class 1	Class 2	Class 0	Class 1	Class 2	Class 0	C
	727,179	847,876	239,129	707,953	913,166	110,110	727,179	1,0
	725,073	671,576	192,781	702,583	727,219	99,672	725,073	83
	546,815	541,795	106,297	524,137	561,165	55,115	546,815	63

We will also see what mechanisms the authors followed in terms of generating this information about what is the possible location that this person's home would be.

(ReferSlideTime:26:15)



ity, as validated by *Yahoo! PlaceFinder*.

In our evaluation, we group users into three classes. *Class 0* consists of users who have a single activity, either a mayorship, a tip or a done. In this case, the unique choice is to set the user's home location equal to that of her activity. *Class 1* consists of users who have multiple activities with a predominant location across them. For these users, the inferred location matches the most often location of their activities. *Class 2*, in turn, consists of users with multiple activities in which there is no single location that stands out (i.e., there are ties). Our current inference approach cannot be applied to *Class 2* users.

Thus, we evaluate the proposed models by assessing their

Fig
Dec

We
res

Here is the scheme that they followed. They actually put the data into 3 buckets, class 0, class 1 and class 2. The class 0 is of the users who have single activity either the mayorship, the tip or done. They only have one activity in the dataset, whereas class 1 consistsofuserswhohavemultipleactivitieswithpredominantlocationacrossthem.So, for example, I have multiple tips and dones, mayorships in my account, but there is one which is very, very high which is IIIT Delhi for that matter. Class 2 is consists of users with multiple activities in which there is no single location that stands out.

So, again just to understand, if you understand this I think the inferences become much simple, the logic the authors followed is that take all the users who are been doing this tip,donesandmayorshipsinthedata.Class0orthepeoplewhohaveonlysingleactivity either a tip, or a mayorship or a done, class 1 is set of people who were where this one majority location that shows up for them. Class 2 is a set of people where multiple activities are done, butno single location is actually predominant in their activity,so that is the kind of classification that they made with the users.

(ReferSlideTime:28:01)

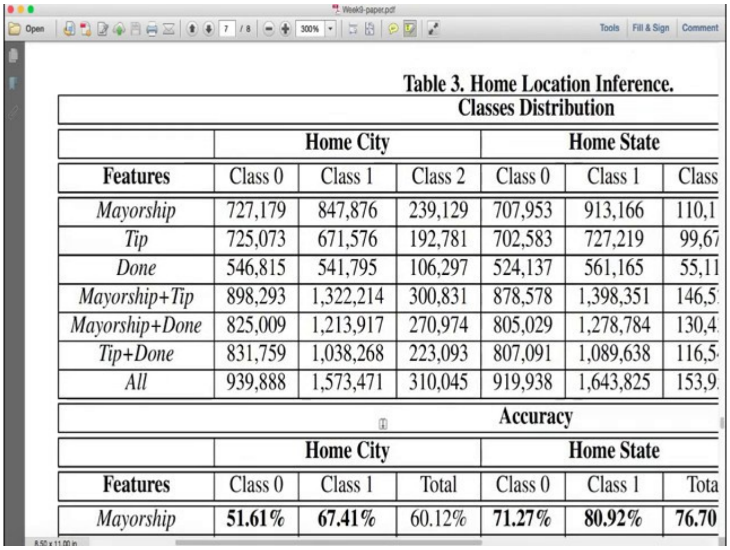


Table 3. Home Location Inference.

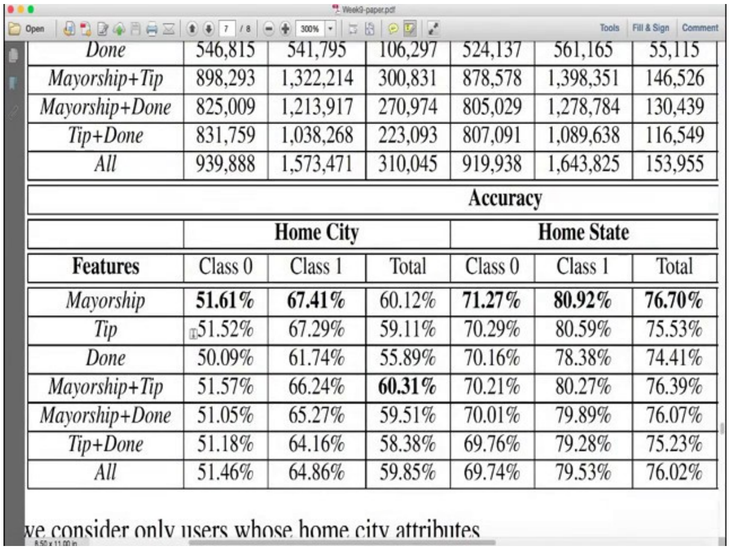
Classes Distribution						
	Home City			Home State		
Features	Class 0	Class 1	Class 2	Class 0	Class 1	Class
Mayorship	727,179	847,876	239,129	707,953	913,166	110,1
Tip	725,073	671,576	192,781	702,583	727,219	99,67
Done	546,815	541,795	106,297	524,137	561,165	55,11
Mayorship+Tip	898,293	1,322,214	300,831	878,578	1,398,351	146,5
Mayorship+Done	825,009	1,213,917	270,974	805,029	1,278,784	130,4
Tip+Done	831,759	1,038,268	223,093	807,091	1,089,638	116,5
All	939,888	1,573,471	310,045	919,938	1,643,825	153,9

Accuracy						
	Home City			Home State		
Features	Class 0	Class 1	Total	Class 0	Class 1	Tota
Mayorship	**51.61%**	**67.41%**	60.12%	**71.27%**	**80.92%**	**76.70**

Now I will show you two tables, one is the number of people, number of the data points from this dataset. If you just consider the class 0, class 1 and class 2, how many people are actually where you can infer home city,home state, and home country. So, how do I read this graph this how do you read this table. This table is this column it showing you features mayorships, tip and done, mayorship plus tip, mayorship plus done, tip plusdone and all of them, right. So, which is if you take onlythe mayorship, what is the class 0,which is onlyifIconsider mayorshipand thereis only1activitybythis user,there are about 727,000 data points in class 0; 847,000 where that are users where one location is actually predominant; and 239,000 there is no location that is predominant, that is how you read this table, correct.

So, if you look at mayorships 127,000; mayorship plus tip 898,000; obviously,mayorship plus tip will be higher, all will be higher for all of them, bigger than all of them and that is for the city.So, for the state 700,000 is for class 0; 900,000 are for class 100,000 is for class 2. Similarly, for the country, so that is giving you a sense of in the data points or the pieces of information that is available for each of these features.

(ReferSlideTime:30:09)



Done	546,815	541,795	106,297	524,137	561,165	55,115
Mayorship+Tip	898,293	1,322,214	300,831	878,578	1,398,351	146,526
Mayorship+Done	825,009	1,213,917	270,974	805,029	1,278,784	130,439
Tip+Done	831,759	1,038,268	223,093	807,091	1,089,638	116,549
All	939,888	1,573,471	310,045	919,938	1,643,825	153,955

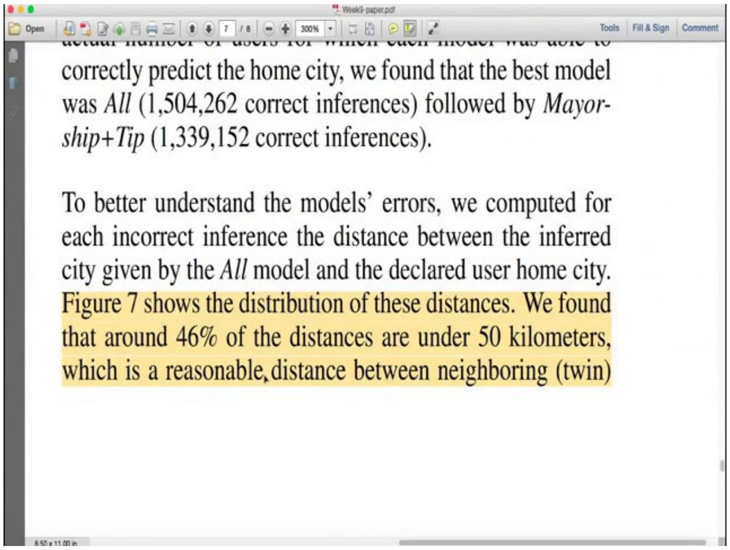
Accuracy						
	Home City			**Home State**		
Features	Class 0	Class 1	Total	Class 0	Class 1	Total
Mayorship	**51.61%**	**67.41%**	60.12%	**71.27%**	**80.92%**	**76.70%**
Tip	51.52%	67.29%	59.11%	70.29%	80.59%	75.53%
Done	50.09%	61.74%	55.89%	70.16%	78.38%	74.41%
Mayorship+Tip	51.57%	66.24%	**60.31%**	70.21%	80.27%	76.39%
Mayorship+Done	51.05%	65.27%	59.51%	70.01%	79.89%	76.07%
Tip+Done	51.18%	64.16%	58.38%	69.76%	79.28%	75.23%
All	51.46%	64.86%	59.85%	69.74%	79.53%	76.02%

ve consider only users whose home city attributes

Now, let us look at accuracy, which is if I were to use this information and find out that mayorship, only using the mayorship, I am able to find the home city 51.6 percentage of the times, wherever the percentage is higher it is been actually kept in bold. So, if you look at mayorship it is class 0 on class 1, we cannot do class 2 because it is actually the places where a particular location cannot be, one single location cannot be inferred, which is because our goal is to infer actually the home location.

So, if I have multiple locations, I am not going to use that column, that is why class 2 does not exist in high accuracy. So, which 67 percentage of the times, home city can be inferred, if I look at class 1 category of people. Which is I do a lot of tips and dones, but my predominant place where I do a tip or a done, tip or a done is actually my home location. Home state becomes higher, seventy percent and home countries even higher,of course, the percentage for the country is going to be, so the percentage of city will always be lesser than state, will always be lesser than country. Because here to get the country that I am from India more difficult to get that I'm from Tamilnadu as a state it is even more difficult to get that IamfromChennai, that is about the inference of the home location.

(ReferSlideTime:32:09)



actual number of users for which each model was able to correctly predict the home city, we found that the best model was *All* (1,504,262 correct inferences) followed by *Mayorship+Tip* (1,339,152 correct inferences).

To better understand the models' errors, we computed for each incorrect inference the distance between the inferred city given by the *All* model and the declared user home city. Figure 7 shows the distribution of these distances. We found that around 46% of the distances are under 50 kilometers, which is a reasonable distance between neighboring (twin)

There's another interesting graph that authors have which is figure 7, which I show you the graph and then I will try to explain it.

(ReferSlideTime:32:17)

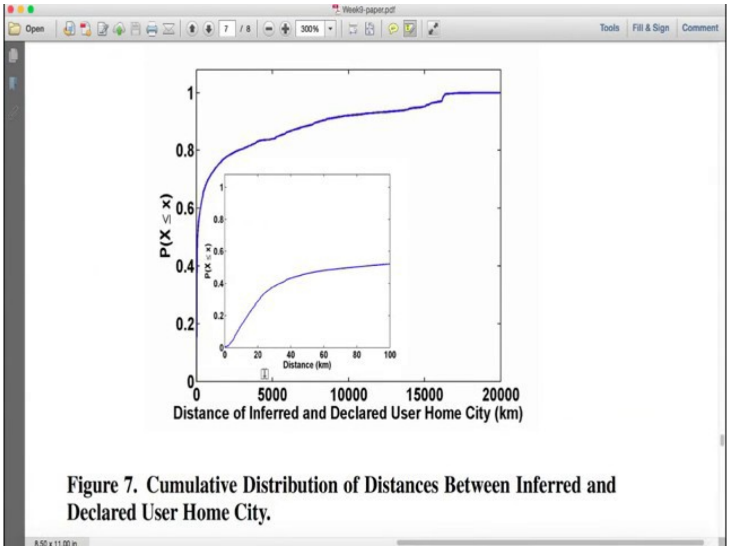
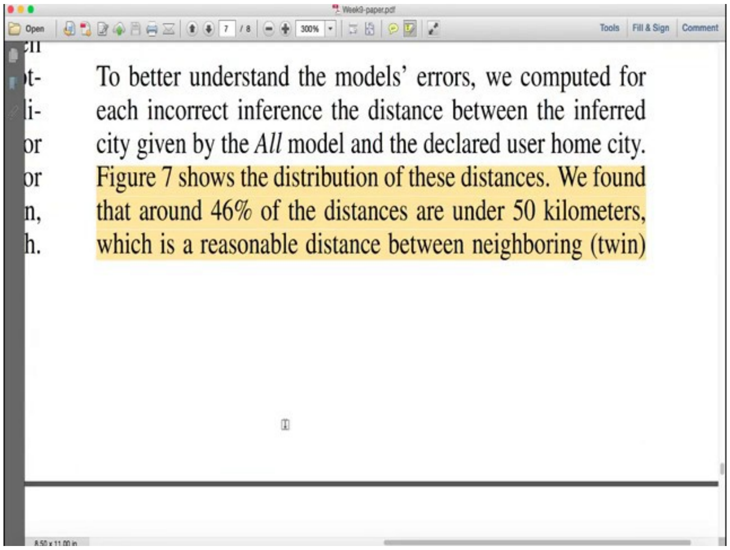


Figure 7. Cumulative Distribution of Distances Between Inferred and Declared User Home City.

So, here figure 7 is cumulative distribution of distances between inferred and thedeclared home city, which is that because people actually declared that what the home location in these services also. If you go to myaccount, you will find that is, if you go to my facebook account I probably say that I am from the current location is from New DelhiandmyhomelocationisfromChennai,sothatinformationyoucanusetomake the difference which is what did we predict from the table that I showed you know,which is prediction of my home location with class 0 or class 1.

And then use it for finding the difference between what did I say and what I actually have, that is the graph here. So, x-axis is the distance of inferred and declared user home city, and y-axis is the probability. So, this is how you will read the graph.

(ReferSlideTime:33:28)



To better understand the models' errors, we computed for each incorrect inference the distance between the inferred city given by the *All* model and the declared user home city. Figure 7 shows the distribution of these distances. We found that around 46% of the distances are under 50 kilometers, which is a reasonable distance between neighboring (twin)

Which is, distribution of these distances are shown in figure 7. Wefound that 46 percent of the distances are under 50 kilometers that is what the authors did.

(ReferSlideTime:33:40)

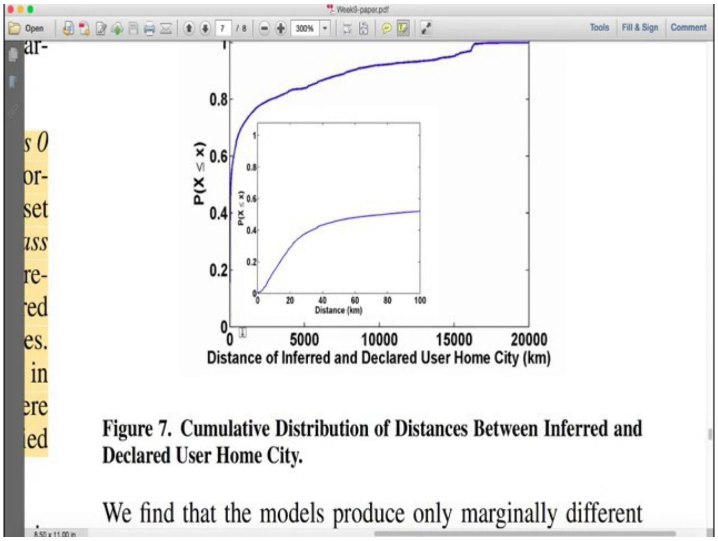
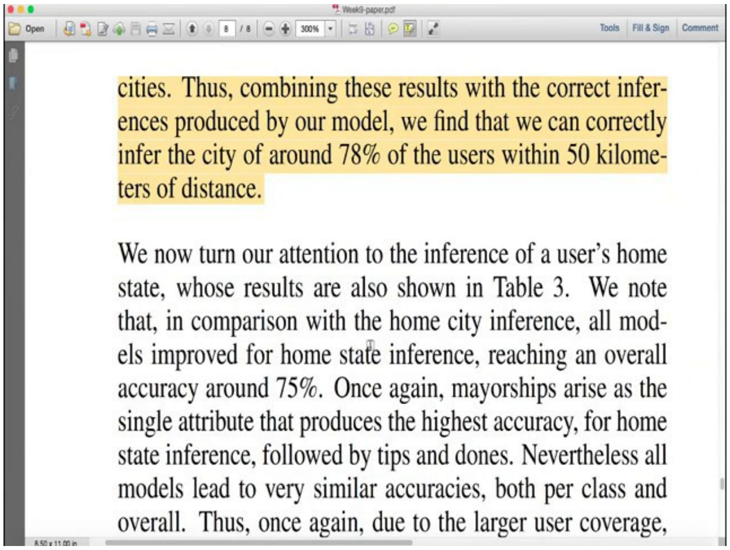


Figure 7. Cumulative Distribution of Distances Between Inferred and Declared User Home City.

We find that the models produce only marginally different

They actually zoomed in. So, this is meaning, this is 5000 kilometers, but as if you look at this graph the inside graph is only for 100 kilometers. So, you can clearly see that 46 percent of the distances are 150 kilometers. So, here is 50 kilometers and if you see 46 percent should be somewhere here correct. So, 46 percent of the users are actuallyhaving the error between finding the home location and the actual location is about 50 kilometers, that is pretty small, if I were able to actually use this information with only50 kilometers of error which is I just getting it from the general behavior tips and dones, that's quite effective.

(ReferSlideTime:34:37)

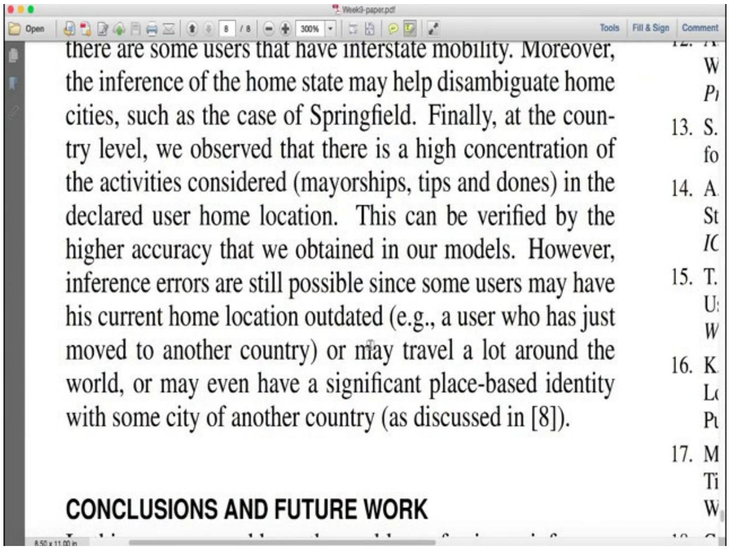


cities. Thus, combining these results with the correct inferences produced by our model, we find that we can correctly infer the city of around 78% of the users within 50 kilometers of distance.

We now turn our attention to the inference of a user's home state, whose results are also shown in Table 3. We note that, in comparison with the home city inference, all models improved for home state inference, reaching an overall accuracy around 75%. Once again, mayorships arise as the single attribute that produces the highest accuracy, for home state inference, followed by tips and dones. Nevertheless all models lead to very similar accuracies, both per class and overall. Thus, once again, due to the larger user coverage,

So, if you just take this model, and then if you look at theauthor's claim that 78 percent of the users within 50 kilometers of distance, which is whattheyaresaying is combining these results with the correct inferences produced by our model, we find that we can correctly infer the city of around 78 percentage of the users within 50 kilometers. So, whatever your city is we will be able to make an inference of that city of about 78percent within the 50 kilometers of distance, correct.

(ReferSlideTime:35:27)

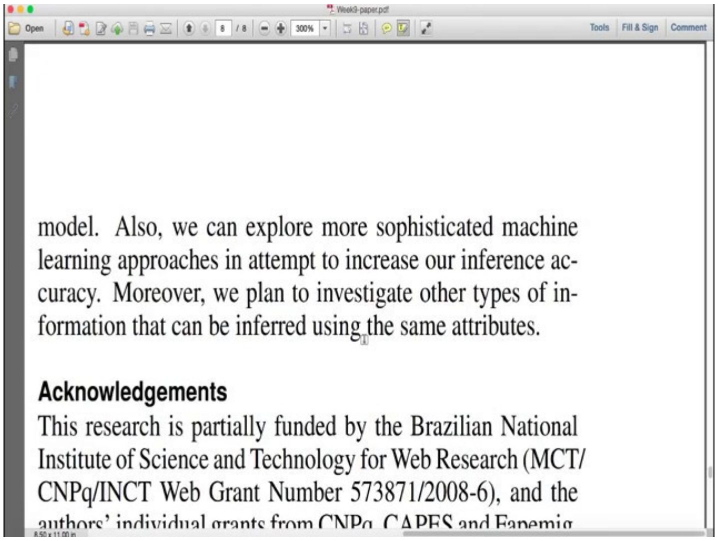


there are some users that have interstate mobility. Moreover, the inference of the home state may help disambiguate home cities, such as the case of Springfield. Finally, at the country level, we observed that there is a high concentration of the activities considered (mayorships, tips and dones) in the declared user home location. This can be verified by the higher accuracy that we obtained in our models. However, inference errors are still possible since some users may have his current home location outdated (e.g., a user who has just moved to another country) or may travel a lot around the world, or may even have a significant place-based identity with some city of another country (as discussed in [8]).

CONCLUSIONS AND FUTURE WORK

So, that gives you a sense of how also if you go if you remember even in the abstract I showed you this 78 percent of accuracy that the authors claimed that you can find the home location. And of course, in the paper structure, you finish off with the conclusions and future work and probably have some limitations if there are any data limitations in data methodology, any limitations in the paper, right.

(ReferSlideTime:35:49)



model. Also, we can explore more sophisticated machine learning approaches in attempt to increase our inference accuracy. Moreover, we plan to investigate other types of information that can be inferred using the same attributes.

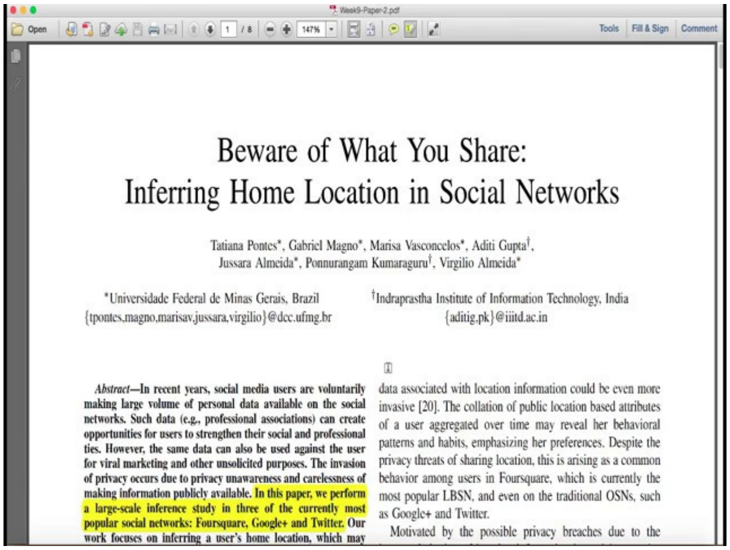
Acknowledgements

This research is partially funded by the Brazilian National Institute of Science and Technology for Web Research (MCT/CNPq/INCT Web Grant Number 573871/2008-6), and the authors' individual grants from CNPq, CAPES and Fapemig

So, that gives you a sense of how a paper meaning the things that you have seen in the class until now which is to look at take some data do some analysis, make some inferences, how these inferences are put into paper is what we saw in this particular lecture. And the focus was actually taking foursquare and finding the home location. With that, I will stop here for this paper, and I will see you soon.

Beware of What You Share Inferring Home Location in Social Networks- On the dynamics of username change behavior on Twitter

(ReferSlideTime:00:12)



Beware of What You Share:
Inferring Home Location in Social Networks

Tatiana Pontes*, Gabriel Magno*, Marisa Vasconcelos*, Aditi Gupta†,
Jussara Almeida*, Ponnurangam Kumaraguru†, Virgilio Almeida*

*Universidade Federal de Minas Gerais, Brazil
{tpontes,magno,marisav,jussara,virgilio}@dcc.ufmg.br

†Indraprastha Institute of Information Technology, India
{aditig.pk}@iiitd.ac.in

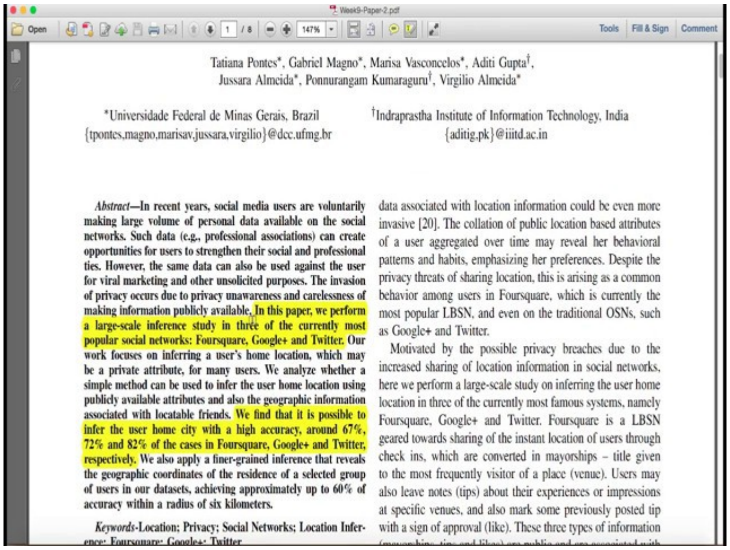
Abstract—In recent years, social media users are voluntarily making large volume of personal data available on the social networks. Such data (e.g., professional associations) can create opportunities for users to strengthen their social and professional ties. However, the same data can also be used against the user for viral marketing and other unsolicited purposes. The invasion of privacy occurs due to privacy unawareness and carelessness of making information publicly available. In this paper, we perform a large-scale inference study in three of the currently most popular social networks: Foursquare, Google+ and Twitter. Our work focuses on inferring a user's home location, which may

data associated with location information could be even more invasive [20]. The collation of public location based attributes of a user aggregated over time may reveal her behavioral patterns and habits, emphasizing her preferences. Despite the privacy threats of sharing location, this is arising as a common behavior among users in Foursquare, which is currently the most popular LBSN, and even on the traditional OSNs, such as Google+ and Twitter.

Motivated by the possible privacy breaches due to the

Welcome back to the course Privacy and Security in Online Social Media. Continuingthe trend that I did last lecture, I am going to continue actually looking at some papers which are basically addressing the problem of privacy leakages from location based services.Ifyou rememberlastlecture,wehadpaper whichlooked atFoursquare,andthe paper analyzed, how they can actually identify, where a person lives. That was onlyusing the Foursquare mayorship, tips and dones. And what we are going to see now is almost the same topic, but we are going to actually compare it with different social networks.

(ReferSlideTime:01:11)



Tatiana Pontes*, Gabriel Magno*, Marisa Vasconcelos*, Aditi Gupta†,
Jussara Almeida*, Ponnurangam Kumaraguru†, Virgilio Almeida*

*Universidade Federal de Minas Gerais, Brazil
{tpontes,magno,marisav,jussara,virgilio}@dcc.ufmg.br

†Indraprastha Institute of Information Technology, India
{aditig,pk}@iiitd.ac.in

***Abstract*—In recent years, social media users are voluntarily making large volume of personal data available on the social networks. Such data (e.g., professional associations) can create opportunities for users to strengthen their social and professional ties. However, the same data can also be used against the user for viral marketing and other unsolicited purposes. The invasion of privacy occurs due to privacy unawareness and carelessness of making information publicly available. In this paper, we perform a large-scale inference study in three of the currently most popular social networks: Foursquare, Google+ and Twitter. Our work focuses on inferring a user's home location, which may be a private attribute, for many users. We analyze whether a simple method can be used to infer the user home location using publicly available attributes and also the geographic information associated with locatable friends. We find that it is possible to infer the user home city with a high accuracy, around 67%, 72% and 82% of the cases in Foursquare, Google+ and Twitter, respectively. We also apply a finer-grained inference that reveals the geographic coordinates of the residence of a selected group of users in our datasets, achieving approximately up to 60% of accuracy within a radius of six kilometers.**

Keywords-Location; Privacy; Social Networks; Location Inference; Foursquare; Google+; Twitter

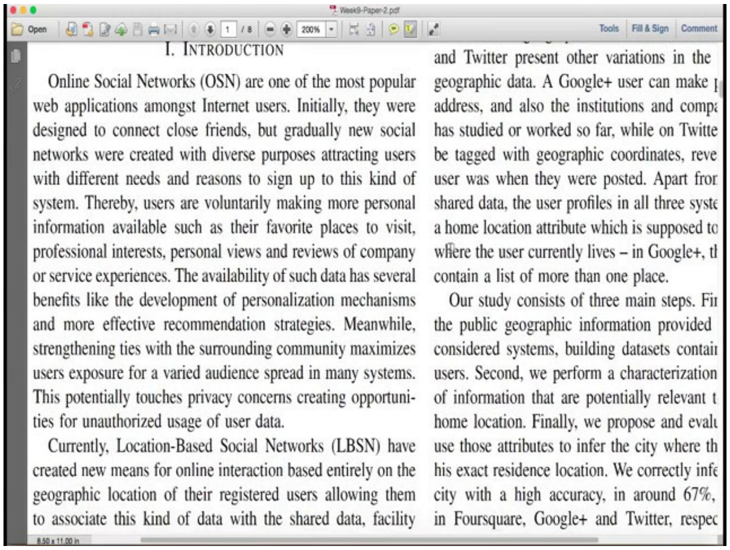
data associated with location information could be even more invasive [20]. The collation of public location based attributes of a user aggregated over time may reveal her behavioral patterns and habits, emphasizing her preferences. Despite the privacy threats of sharing location, this is arising as a common behavior among users in Foursquare, which is currently the most popular LBSN, and even on the traditional OSNs, such as Google+ and Twitter.

Motivated by the possible privacy breaches due to the increased sharing of location information in social networks, here we perform a large-scale study on inferring the user home location in three of the currently most famous systems, namely Foursquare, Google+ and Twitter. Foursquare is a LBSN geared towards sharing of the instant location of users through check ins, which are converted in mayorships – title given to the most frequently visitor of a place (venue). Users may also leave notes (tips) about their experiences or impressions at specific venues, and also mark some previously posted tip with a sign of approval (like). These three types of information

So, ifyou see here inthis paper the authors performa large scale inference studyinthree of the currently most popular social networks like Foursquare, Google plus and Twitter. So, the goal in this paper is verysimilar to the paper that wesaw last time, but it is going to be looking at different social networks not just only Foursquare. So, in this authors lookedatFoursquare,GoogleplusandTwitter.Youknowaspartofthiscourse,youhave already seen all three social networks in terms of their content, in terms of the data collection that is done and information that you can actually collect from the social networks.

The authors actually find that it is possible to infer the user home city with the high accuracyaround 67percent, 72percent and 82percent in the case ofFoursquare, Google plus and Twitter, which is 67 percent for Foursquare, 72 percent for Google plus and 82 percent for Twitter. I am sure as we move along; you will actually understand why Twitter is actually high in terms of finding out the home location.

(ReferSlideTime:02:55)



I. Introduction

Online Social Networks (OSN) are one of the most popular web applications amongst Internet users. Initially, they were designed to connect close friends, but gradually new social networks were created with diverse purposes attracting users with different needs and reasons to sign up to this kind of system. Thereby, users are voluntarily making more personal information available such as their favorite places to visit, professional interests, personal views and reviews of company or service experiences. The availability of such data has several benefits like the development of personalization mechanisms and more effective recommendation strategies. Meanwhile, strengthening ties with the surrounding community maximizes users exposure for a varied audience spread in many systems. This potentially touches privacy concerns creating opportunities for unauthorized usage of user data.

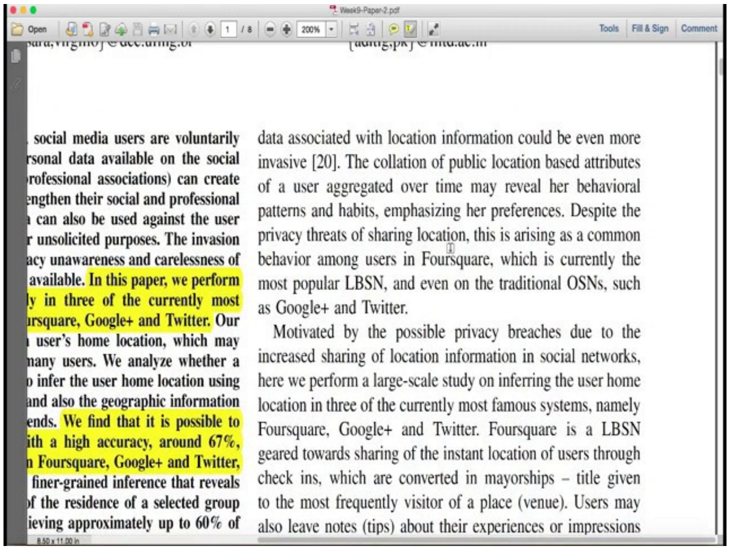
Currently, Location-Based Social Networks (LBSN) have created new means for online interaction based entirely on the geographic location of their registered users allowing them to associate this kind of data with the shared data, facility

and Twitter present other variations in the
geographic data. A Google+ user can make
address, and also the institutions and comp
has studied or worked so far, while on Twitte
be tagged with geographic coordinates, reve
user was when they were posted. Apart fror
shared data, the user profiles in all three syste
a home location attribute which is supposed to
where the user currently lives – in Google+, tl
contain a list of more than one place.

Our study consists of three main steps. Fir
the public geographic information provided
considered systems, building datasets contai
users. Second, we perform a characterization
of information that are potentially relevant t
home location. Finally, we propose and evalu
use those attributes to infer the city where th
his exact residence location. We correctly infe
city with a high accuracy, in around 67%,
in Foursquare, Google+ and Twitter, respec

So, now let us look at the paper in terms of the same structure as we saw last time, introduction, talking about what a location based social networks are.

(ReferSlideTime:03:04)



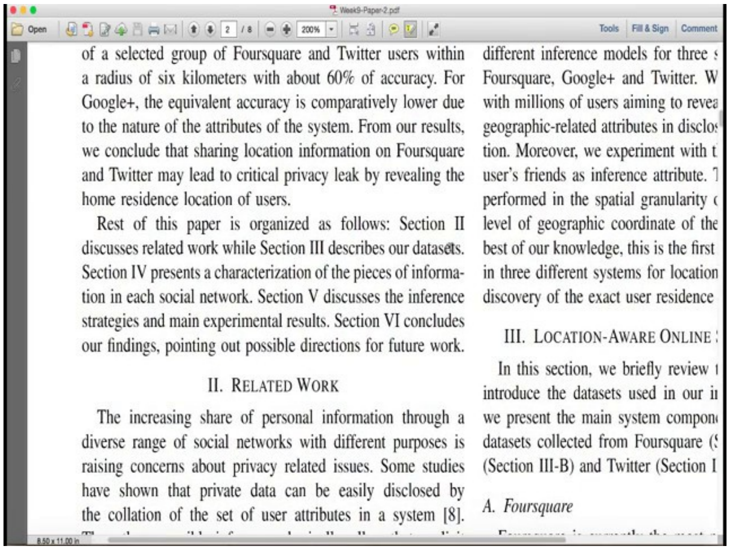
social media users are voluntarily
rsonal data available on the social
rofessional associations) can create
engthen their social and professional
can also be used against the user
r unsolicited purposes. The invasion
acy unawareness and carelessness of
available. In this paper, we perform
ly in three of the currently most
ursquare, Google+ and Twitter. Our
user's home location, which may
many users. We analyze whether a
o infer the user home location using
and also the geographic information
ends. We find that it is possible to
ith a high accuracy, around 67%,
n Foursquare, Google+ and Twitter,
finer-grained inference that reveals
of the residence of a selected group
ieving approximately up to 60% of

data associated with location information could be even more invasive [20]. The collation of public location based attributes of a user aggregated over time may reveal her behavioral patterns and habits, emphasizing her preferences. Despite the privacy threats of sharing location, this is arising as a common behavior among users in Foursquare, which is currently the most popular LBSN, and even on the traditional OSNs, such as Google+ and Twitter.

Motivated by the possible privacy breaches due to the increased sharing of location information in social networks, here we perform a large-scale study on inferring the user home location in three of the currently most famous systems, namely Foursquare, Google+ and Twitter. Foursquare is a LBSN geared towards sharing of the instant location of users through check ins, which are converted in mayorships – title given to the most frequently visitor of a place (venue). Users may also leave notes (tips) about their experiences or impressions

Talking about collecting data from a different social media services like Google plus Foursquare and Twitter and different research that are done in the context of Foursquare Google plus and Twitter.And I am talking about what information was collected, and a little bit of conclusions of the paper itself, and then talking about how the paper is out maxed.

(ReferSlideTime:03:32)



of a selected group of Foursquare and Twitter users within a radius of six kilometers with about 60% of accuracy. For Google+, the equivalent accuracy is comparatively lower due to the nature of the attributes of the system. From our results, we conclude that sharing location information on Foursquare and Twitter may lead to critical privacy leak by revealing the home residence location of users.

Rest of this paper is organized as follows: Section II discusses related work while Section III describes our datasets. Section IV presents a characterization of the pieces of information in each social network. Section V discusses the inference strategies and main experimental results. Section VI concludes our findings, pointing out possible directions for future work.

II. RELATED WORK

The increasing share of personal information through a diverse range of social networks with different purposes is raising concerns about privacy related issues. Some studies have shown that private data can be easily disclosed by the collation of the set of user attributes in a system [8].

different inference models for three s
Foursquare, Google+ and Twitter. W
with millions of users aiming to revea
geographic-related attributes in disclos
tion. Moreover, we experiment with t
user's friends as inference attribute. T
performed in the spatial granularity c
level of geographic coordinate of the
best of our knowledge, this is the first
in three different systems for location
discovery of the exact user residence

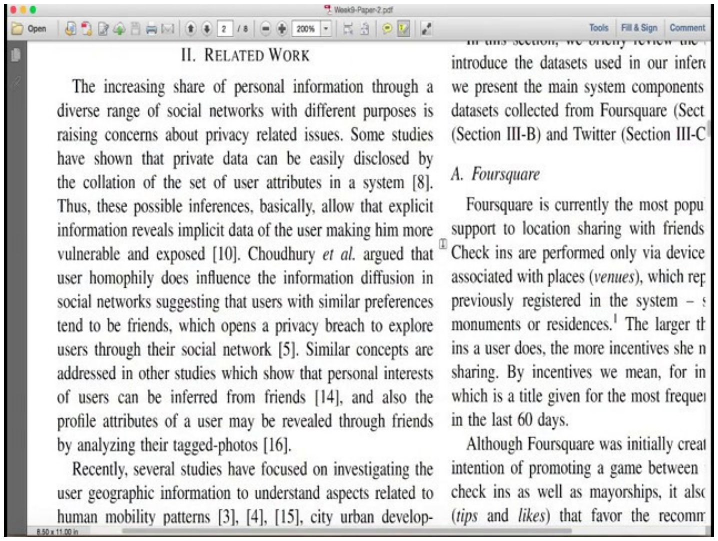
III. LOCATION-AWARE ONLINE

In this section, we briefly review t
introduce the datasets used in our in
we present the main system compon
datasets collected from Foursquare (S
(Section III-B) and Twitter (Section I

A. *Foursquare*

So, this is generally the structure that we saw even the last time, meaning almost all papers appears see the structure would be the same a paragraph about the 30,000 feet high view of the problem. Then the paragraph about the current problem and what is missing, then the paragraph about what is, what was done in this paper and then some kind ofacontributionfromthis paper.Relatedworkagain Iamnot goingtodetail inthis particular related work.

(ReferSlideTime:04:02)



II. RELATED WORK

The increasing share of personal information through a diverse range of social networks with different purposes is raising concerns about privacy related issues. Some studies have shown that private data can be easily disclosed by the collation of the set of user attributes in a system [8]. Thus, these possible inferences, basically, allow that explicit information reveals implicit data of the user making him more vulnerable and exposed [10]. Choudhury *et al.* argued that user homophily does influence the information diffusion in social networks suggesting that users with similar preferences tend to be friends, which opens a privacy breach to explore users through their social network [5]. Similar concepts are addressed in other studies which show that personal interests of users can be inferred from friends [14], and also the profile attributes of a user may be revealed through friends by analyzing their tagged-photos [16].

Recently, several studies have focused on investigating the user geographic information to understand aspects related to human mobility patterns [3], [4], [15], city urban develop-

introduce the datasets used in our infer
we present the main system components
datasets collected from Foursquare (Sect
(Section III-B) and Twitter (Section III-C

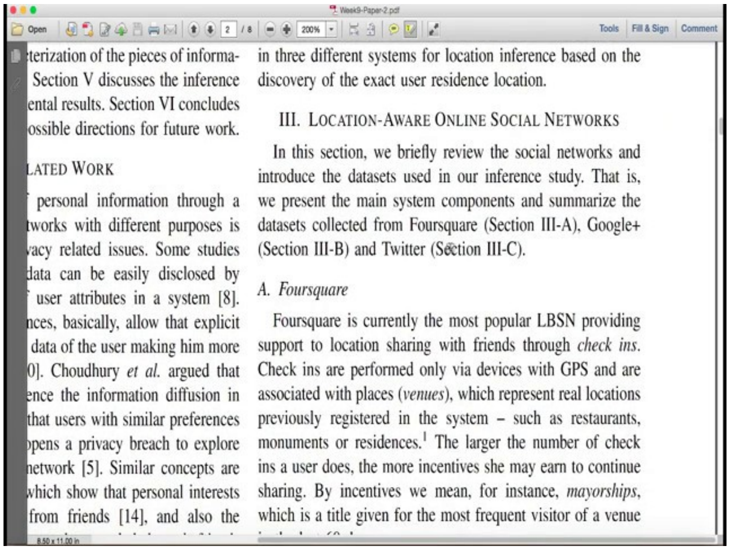
A. *Foursquare*

Foursquare is currently the most popu
support to location sharing with friends
Check ins are performed only via device
associated with places (*venues*), which rep
previously registered in the system – s
monuments or residences.[1] The larger th
ins a user does, the more incentives she n
sharing. By incentives we mean, for in
which is a title given for the most freque
in the last 60 days.

Although Foursquare was initially creat
intention of promoting a game between
check ins as well as mayorships, it also
(*tips* and *likes*) that favor the recomm

But related work generally talks about these kind of privacy leakages from locationbased services and work done on collecting data from these three social networks and inferences that were done.

(ReferSlideTime:04:19)



terization of the pieces of informa-
Section V discusses the inference
ental results. Section VI concludes
ossible directions for future work.

LATED WORK

personal information through a
tworks with different purposes is
acy related issues. Some studies
data can be easily disclosed by
user attributes in a system [8].
nces, basically, allow that explicit
data of the user making him more
0]. Choudhury *et al.* argued that
ence the information diffusion in
that users with similar preferences
pens a privacy breach to explore
network [5]. Similar concepts are
which show that personal interests
from friends [14], and also the

in three different systems for location inference based on the discovery of the exact user residence location.

III. LOCATION-AWARE ONLINE SOCIAL NETWORKS

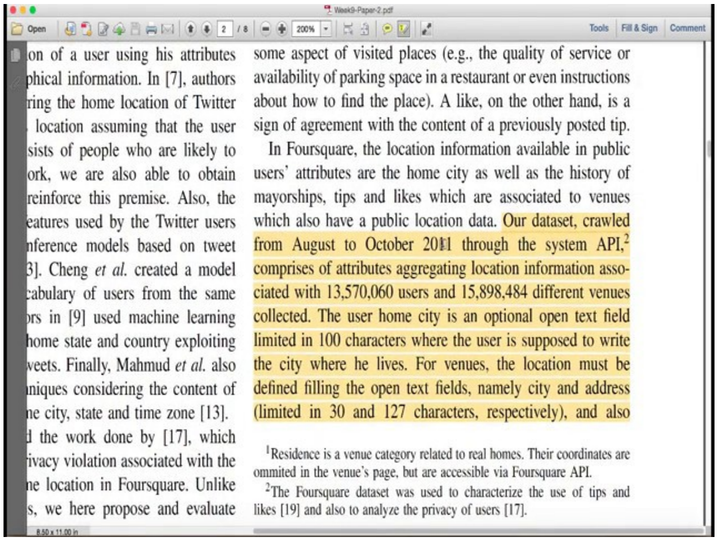
In this section, we briefly review the social networks and introduce the datasets used in our inference study. That is, we present the main system components and summarize the datasets collected from Foursquare (Section III-A), Google+ (Section III-B) and Twitter (Section III-C).

A. Foursquare

Foursquare is currently the most popular LBSN providing support to location sharing with friends through *check ins*. Check ins are performed only via devices with GPS and are associated with places (*venues*), which represent real locations previously registered in the system – such as restaurants, monuments or residences.[1] The larger the number of check ins a user does, the more incentives she may earn to continue sharing. By incentives we mean, for instance, *mayorships*, which is a title given for the most frequent visitor of a venue

Thenmanyatimesresearchersactuallytendtowritedetailsaboutthesocialnetworkthat isbeing discussed interms ofjustintroducingtheterminologieswhich wesawin thelast paper also. Here it is talking about Foursquare then there would be about Google plus and then Twitter.

(ReferSlideTime:04:42)



on of a user using his attributes
phical information. In [7], authors
ring the home location of Twitter
location assuming that the user
sists of people who are likely to
ork, we are also able to obtain
reinforce this premise. Also, the
eatures used by the Twitter users
nference models based on tweet
3]. Cheng *et al.* created a model
cabulary of users from the same
ors in [9] used machine learning
home state and country exploiting
weets. Finally, Mahmud *et al.* also
niques considering the content of
ne city, state and time zone [13].
d the work done by [17], which
ivacy violation associated with the
ne location in Foursquare. Unlike
s, we here propose and evaluate

some aspect of visited places (e.g., the quality of service or availability of parking space in a restaurant or even instructions about how to find the place). A like, on the other hand, is a sign of agreement with the content of a previously posted tip.

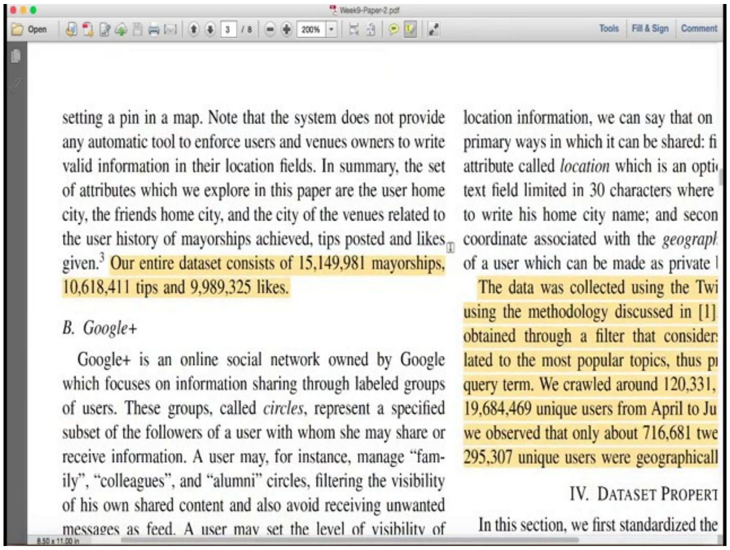
In Foursquare, the location information available in public users' attributes are the home city as well as the history of mayorships, tips and likes which are associated to venues which also have a public location data. Our dataset, crawled from August to October 2011 through the system API,[2] comprises of attributes aggregating location information associated with 13,570,060 users and 15,898,484 different venues collected. The user home city is an optional open text field limited in 100 characters where the user is supposed to write the city where he lives. For venues, the location must be defined filling the open text fields, namely city and address (limited in 30 and 127 characters, respectively), and also

[1] Residence is a venue category related to real homes. Their coordinates are ommited in the venue's page, but are accessible via Foursquare API.

[2] The Foursquare dataset was used to characterize the use of tips and likes [19] and also to analyze the privacy of users [17].

So, here if you look at it, the dataset that was used in the study is the same as the last paper.Datasetcrawled betweenOctober 2011,through thesystemanditcomprisesof13 millionusersandaboutcloseto16milliondifferentvenues.Andtheuserhomecityisan optional open text field limited to about 100 characters. For venue, is the location must be defined filling the open text fields, namely city and address limited to 30 and 127 characters respectively.That is the kind of data and that is the kind of information that is available when you collect these data for venue, tips and dones.

(ReferSlideTime:05:42)



setting a pin in a map. Note that the system does not provide any automatic tool to enforce users and venues owners to write valid information in their location fields. In summary, the set of attributes which we explore in this paper are the user home city, the friends home city, and the city of the venues related to the user history of mayorships achieved, tips posted and likes given.[3] Our entire dataset consists of 15,149,981 mayorships, 10,618,411 tips and 9,989,325 likes.

B. Google+

Google+ is an online social network owned by Google which focuses on information sharing through labeled groups of users. These groups, called *circles*, represent a specified subset of the followers of a user with whom she may share or receive information. A user may, for instance, manage "family", "colleagues", and "alumni" circles, filtering the visibility of his own shared content and also avoid receiving unwanted messages as feed. A user may set the level of visibility of

location information, we can say that on
primary ways in which it can be shared: fi
attribute called *location* which is an opti
text field limited in 30 characters where
to write his home city name; and secon
coordinate associated with the *geograph*
of a user which can be made as private

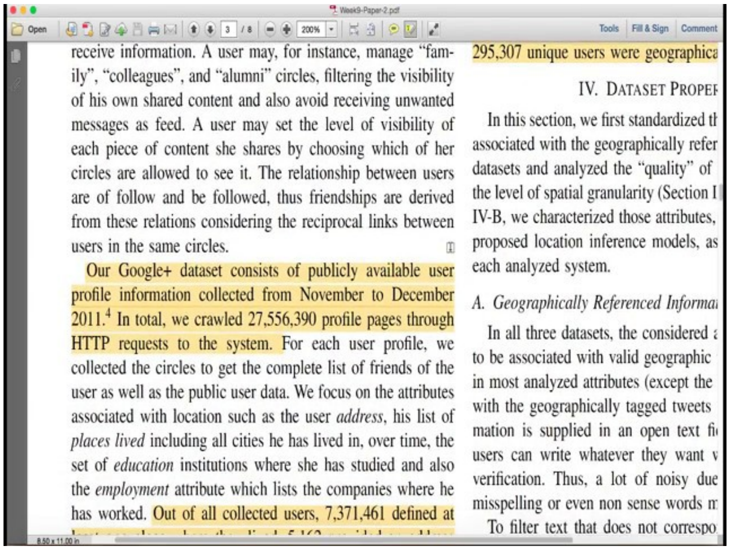
The data was collected using the Twi
using the methodology discussed in [1]
obtained through a filter that consider
lated to the most popular topics, thus p
query term. We crawled around 120,331,
19,684,469 unique users from April to Ju
we observed that only about 716,681 twe
295,307 unique users were geographicall

IV. DATASET PROPERT

In this section, we first standardized the

So, the entire dataset is about 15 million mayorships, close to 11 million tips, and closeto ten million likes. All right? So, likes is basically dones in terms of Foursquare terminology.

(ReferSlideTime:06:05)



receive information. A user may, for instance, manage "family", "colleagues", and "alumni" circles, filtering the visibility of his own shared content and also avoid receiving unwanted messages as feed. A user may set the level of visibility of each piece of content she shares by choosing which of her circles are allowed to see it. The relationship between users are of follow and be followed, thus friendships are derived from these relations considering the reciprocal links between users in the same circles.

Our Google+ dataset consists of publicly available user profile information collected from November to December 2011.[4] In total, we crawled 27,556,390 profile pages through HTTP requests to the system. For each user profile, we collected the circles to get the complete list of friends of the user as well as the public user data. We focus on the attributes associated with location such as the user *address*, his list of *places lived* including all cities he has lived in, over time, the set of *education* institutions where she has studied and also the *employment* attribute which lists the companies where he has worked. Out of all collected users, 7,371,461 defined at

295,307 unique users were geographica

IV. DATASET PROPEF

In this section, we first standardized tl associated with the geographically refer datasets and analyzed the "quality" of the level of spatial granularity (Section I IV-B, we characterized those attributes, proposed location inference models, as each analyzed system.

A. Geographically Referenced Informa

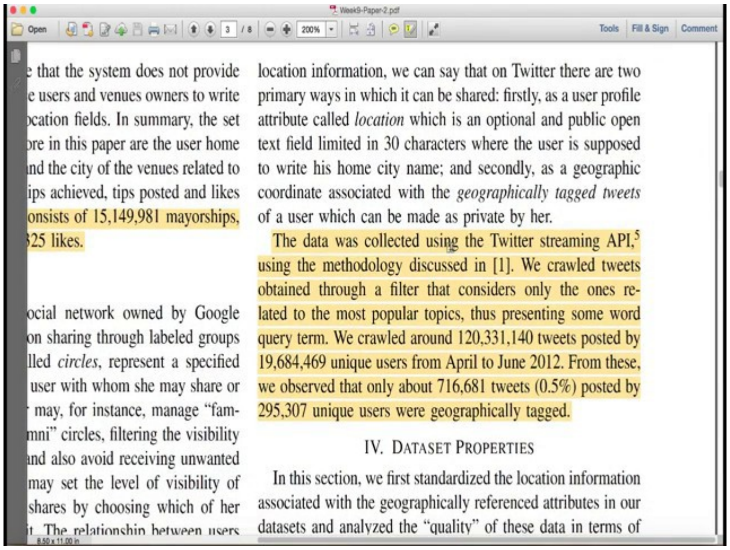
In all three datasets, the considered a to be associated with valid geographic in most analyzed attributes (except the with the geographically tagged tweets mation is supplied in an open text fi users can write whatever they want v verification. Thus, a lot of noisy due misspelling or even non sense words m

To filter text that does not correspo

So, now in terms of Google plus data, Google plus is basically a network that is very similar to, I mean, if you have a Gmail account, you essentially have a Google plus account. In total 27 million profile pages through HTTP request were crawled, and 7 million defined at least one place where they lived, and 5000 provided address information and about 7 million filled their education and about close to 6 million filled their employment.

So, these are details, meaning, if you remember, if you just recollect the social network that you use more often, which is like Facebook, you have all these details at the right places that you live, education, colleges that you study, places that you worked, places that you have lived, all of these information are taken from the users and that is what is mentioned here. Which is 7 million people have explicitly stated their education and about 6 million people have explicitly stated their employment details which is I work at IIIT Delhi.

(ReferSlideTime:07:34)



e that the system does not provide
e users and venues owners to write
ocation fields. In summary, the set
ore in this paper are the user home
nd the city of the venues related to
ips achieved, tips posted and likes
onsists of 15,149,981 mayorships,
325 likes.

ocial network owned by Google
on sharing through labeled groups
lled *circles*, represent a specified
user with whom she may share or
may, for instance, manage "fam-
mni" circles, filtering the visibility
nd also avoid receiving unwanted
may set the level of visibility of
shares by choosing which of her
t. The relationship between users

location information, we can say that on Twitter there are two primary ways in which it can be shared: firstly, as a user profile attribute called *location* which is an optional and public open text field limited in 30 characters where the user is supposed to write his home city name; and secondly, as a geographic coordinate associated with the *geographically tagged tweets* of a user which can be made as private by her.

The data was collected using the Twitter streaming API,[5] using the methodology discussed in [1]. We crawled tweets obtained through a filter that considers only the ones related to the most popular topics, thus presenting some word query term. We crawled around 120,331,140 tweets posted by 19,684,469 unique users from April to June 2012. From these, we observed that only about 716,681 tweets (0.5%) posted by 295,307 unique users were geographically tagged.

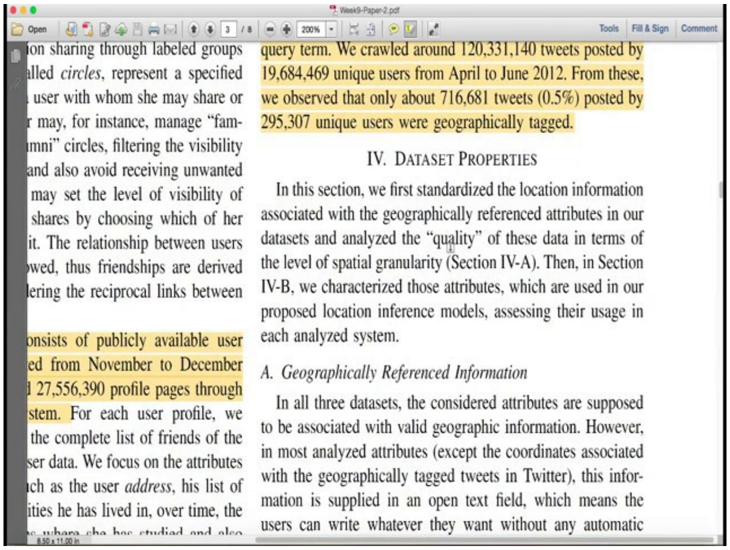
IV. Dataset Properties

In this section, we first standardized the location information associated with the geographically referenced attributes in our datasets and analyzed the "quality" of these data in terms of

Now,letuslookatthedatafromTwitter.So,thedatathedatafromTwitterwascollected using Streaming API, which all of you are aware of. And the crawl was done for 120 milliontweetspostedbyaboutcloseto20millionuniqueusersfromApriltoJune2012. 0.5percentofthepostsposted byuniqueusersweregeographically tagged.

So, what does this mean, this means that there are only 0.5 percent of the total posts that were collected where there is geo tagged information for the post which is geo tagged information for the users also. There were about 700,000 tweets and about 300,000 unique users. That is the exact location, which we have discussed in the past, which is latitude, longitude of the post from where the post is coming.

(ReferSlideTime:08:47)



ion sharing through labeled groups
alled *circles*, represent a specified
user with whom she may share or
r may, for instance, manage "fam-
mni" circles, filtering the visibility
and also avoid receiving unwanted
may set the level of visibility of
shares by choosing which of her
it. The relationship between users
wed, thus friendships are derived
lering the reciprocal links between

onsists of publicly available user
ed from November to December
l 27,556,390 profile pages through
stem. For each user profile, we
the complete list of friends of the
ser data. We focus on the attributes
ch as the user *address*, his list of
ities he has lived in, over time, the
s where she has studied and also

query term. We crawled around 120,331,140 tweets posted by 19,684,469 unique users from April to June 2012. From these, we observed that only about 716,681 tweets (0.5%) posted by 295,307 unique users were geographically tagged.

IV. Dataset Properties

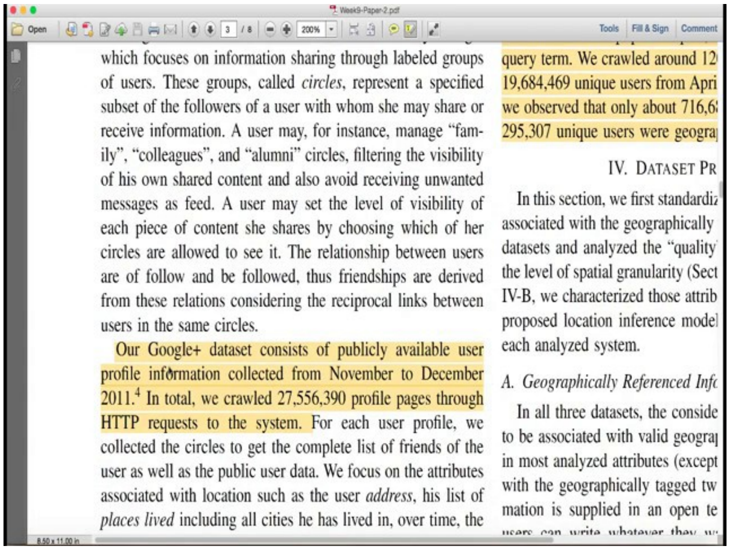
In this section, we first standardized the location information associated with the geographically referenced attributes in our datasets and analyzed the "quality" of these data in terms of the level of spatial granularity (Section IV-A). Then, in Section IV-B, we characterized those attributes, which are used in our proposed location inference models, assessing their usage in each analyzed system.

A. Geographically Referenced Information

In all three datasets, the considered attributes are supposed to be associated with valid geographic information. However, in most analyzed attributes (except the coordinates associated with the geographically tagged tweets in Twitter), this information is supplied in an open text field, which means the users can write whatever they want without any automatic

So, that is the background about the dataset. Essentially all of them are talking about in millions in Twitter it's about 0.5 million geographically tagged tweets geo tag tweets.

(ReferSlideTime:09:00)



which focuses on information sharing through labeled groups of users. These groups, called *circles*, represent a specified subset of the followers of a user with whom she may share or receive information. A user may, for instance, manage "family", "colleagues", and "alumni" circles, filtering the visibility of his own shared content and also avoid receiving unwanted messages as feed. A user may set the level of visibility of each piece of content she shares by choosing which of her circles are allowed to see it. The relationship between users are of follow and be followed, thus friendships are derived from these relations considering the reciprocal links between users in the same circles.

Our Google+ dataset consists of publicly available user profile information collected from November to December 2011.[4] In total, we crawled 27,556,390 profile pages through HTTP requests to the system. For each user profile, we collected the circles to get the complete list of friends of the user as well as the public user data. We focus on the attributes associated with location such as the user *address*, his list of *places lived* including all cities he has lived in, over time, the

query term. We crawled around 12
19,684,469 unique users from Apri
we observed that only about 716,6
295,307 unique users were geogra

IV. Dataset Pr

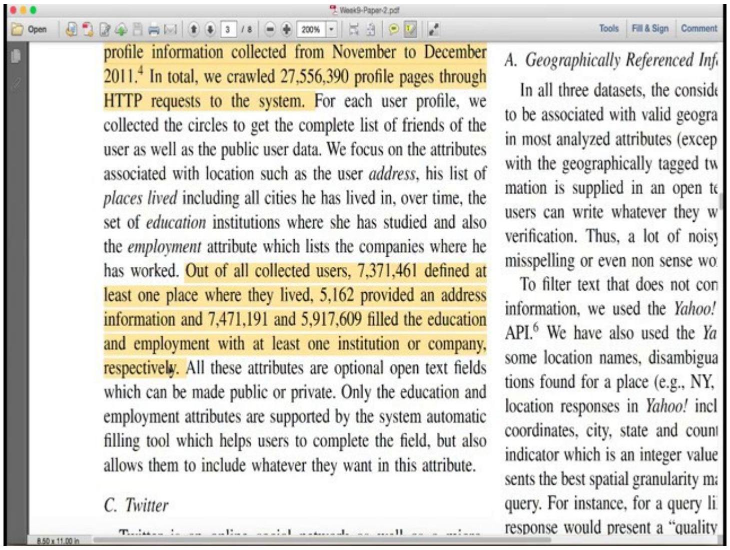
In this section, we first standardiz
associated with the geographically
datasets and analyzed the "quality
the level of spatial granularity (Sect
IV-B, we characterized those attrib
proposed location inference model
each analyzed system.

A. Geographically Referenced Infc

In all three datasets, the conside
to be associated with valid geogra
in most analyzed attributes (except
with the geographically tagged tw
mation is supplied in an open te
users can write whatever they w

In Google plus that are about 27 million profiles that were crawled and about 7 millioneducation and 6 million employment.

(ReferSlideTime:09:07)



profile information collected from November to December 2011.[4] In total, we crawled 27,556,390 profile pages through HTTP requests to the system. For each user profile, we collected the circles to get the complete list of friends of the user as well as the public user data. We focus on the attributes associated with location such as the user *address*, his list of *places lived* including all cities he has lived in, over time, the set of *education* institutions where she has studied and also the *employment* attribute which lists the companies where he has worked. Out of all collected users, 7,371,461 defined at least one place where they lived, 5,162 provided an address information and 7,471,191 and 5,917,609 filled the education and employment with at least one institution or company, respectively. All these attributes are optional open text fields which can be made public or private. Only the education and employment attributes are supported by the system automatic filling tool which helps users to complete the field, but also allows them to include whatever they want in this attribute.

C. Twitter

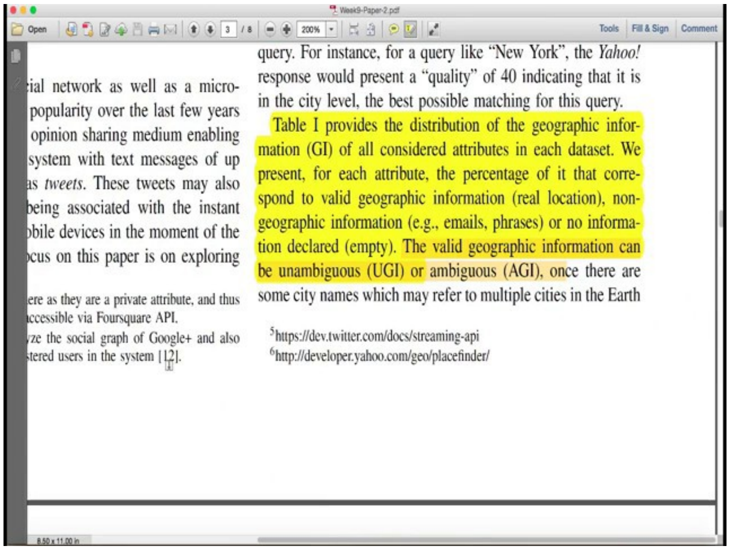
A. Geographically Referenced Inf

In all three datasets, the consid
to be associated with valid geogra
in most analyzed attributes (excep
with the geographically tagged tw
mation is supplied in an open te
users can write whatever they w
verification. Thus, a lot of noisy
misspelling or even non sense wo

To filter text that does not con
information, we used the *Yahoo!*
API.[6] We have also used the *Ya*
some location names, disambigua
tions found for a place (e.g., NY,
location responses in *Yahoo!* incl
coordinates, city, state and coun
indicator which is an integer value
sents the best spatial granularity m
query. For instance, for a query li
response would present a "quality

And Foursquare has details of about 16 million mayorship; 11 million tips and close to 10 million likes. That is the dataset we are going to play around with to do the analysis, to find inferences about the home location of the person.

(ReferSlideTime:09:38)



ial network as well as a micro-
popularity over the last few years
opinion sharing medium enabling
system with text messages of up
as *tweets*. These tweets may also
being associated with the instant
obile devices in the moment of the
ocus on this paper is on exploring

ere as they are a private attribute, and thus
ccessible via Foursquare API.
yze the social graph of Google+ and also
stered users in the system [12].

query. For instance, for a query like "New York", the *Yahoo!* response would present a "quality" of 40 indicating that it is in the city level, the best possible matching for this query.

Table I provides the distribution of the geographic information (GI) of all considered attributes in each dataset. We present, for each attribute, the percentage of it that correspond to valid geographic information (real location), non-geographic information (e.g., emails, phrases) or no information declared (empty). The valid geographic information can be unambiguous (UGI) or ambiguous (AGI), once there are some city names which may refer to multiple cities in the Earth

[5] https://dev.twitter.com/docs/streaming-api
[6] http://developer.yahoo.com/geo/placefinder/

So, in any data set, when you analyze, first you know you want to actually provide exploratory data analysis, and you want to provide what the data set looks like, because this will help reproducibility of that research. This will help others to actually collect data, if they were to, the point here is that if others want to collect the data which is very similar to what you collected; and if others want to do the same analysis that you did the results should be the same. That is the idea for reproducibility of the research. So, explaining how you did collected the data, explaining what the data looks like is extremely important in terms of actually writing these research papers.

(ReferSlideTime:10:45)

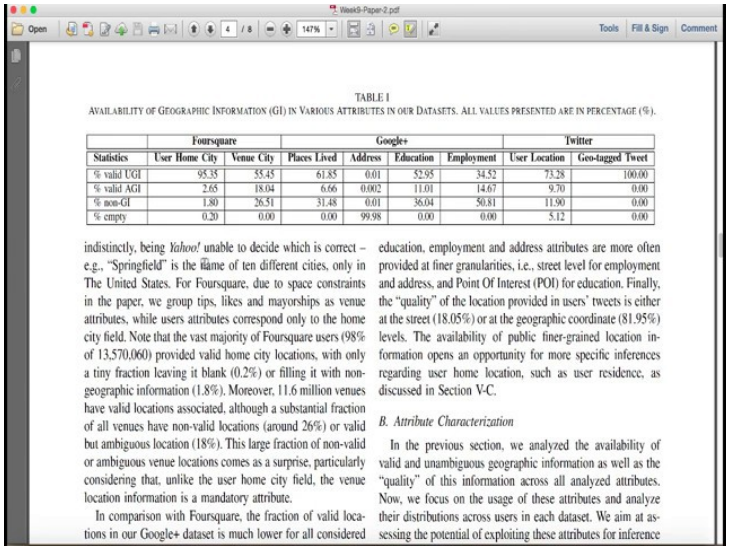


TABLE I
AVAILABILITY OF GEOGRAPHIC INFORMATION (GI) IN VARIOUS ATTRIBUTES IN OUR DATASETS. ALL VALUES PRESENTED ARE IN PERCENTAGE (%).

	Foursquare		Google+				Twitter	
Statistics	User Home City	Venue City	Places Lived	Address	Education	Employment	User Location	Geo-tagged Tweet
% valid UGI	95.35	55.45	61.85	0.01	52.95	34.52	73.28	100.00
% valid AGI	2.65	18.04	6.66	0.002	11.01	14.67	9.70	0.00
% non-GI	1.80	26.51	31.48	0.01	36.04	50.81	11.90	0.00
% empty	0.20	0.00	0.00	99.98	0.00	0.00	5.12	0.00

indistinctly, being *Yahoo!* unable to decide which is correct – e.g., "Springfield" is the name of ten different cities, only in The United States. For Foursquare, due to space constraints in the paper, we group tips, likes and mayorships as venue attributes, while users attributes correspond only to the home city field. Note that the vast majority of Foursquare users (98% of 13,570,060) provided valid home city locations, with only a tiny fraction leaving it blank (0.2%) or filling it with non-geographic information (1.8%). Moreover, 11.6 million venues have valid locations associated, although a substantial fraction of all venues have non-valid locations (around 26%) or valid but ambiguous location (18%). This large fraction of non-valid or ambiguous venue locations comes as a surprise, particularly considering that, unlike the user home city field, the venue location information is a mandatory attribute.

In comparison with Foursquare, the fraction of valid locations in our Google+ dataset is much lower for all considered education, employment and address attributes are more often provided at finer granularities, i.e., street level for employment and address, and Point Of Interest (POI) for education. Finally, the "quality" of the location provided in users' tweets is either at the street (18.05%) or at the geographic coordinate (81.95%) levels. The availability of public finer-grained location information opens an opportunity for more specific inferences regarding user home location, such as user residence, as discussed in Section V-C.

B. Attribute Characterization

In the previous section, we analyzed the availability of valid and unambiguous geographic information as well as the "quality" of this information across all analyzed attributes. Now, we focus on the usage of these attributes and analyze their distributions across users in each dataset. We aim at assessing the potential of exploiting these attributes for inference

So, table one provides the distribution of geographic information of all considered attributes in each dataset in each dataset. Wepresent for each attribute the percentage of it that corresponds to the valid geographic. Let us look at the tables. So, this table is the one that is referred. This table talks about availability of geographic information in various attributes in the datasets. So, if you look at the second column, which is Foursquare. So, the columns are referred three different networks Foursquare, Google plus and Twitter.And if you look at the statistics which are in the rows, it is valid UGI which is user geographic information, valid AGI, valid geographic information and that is empty.

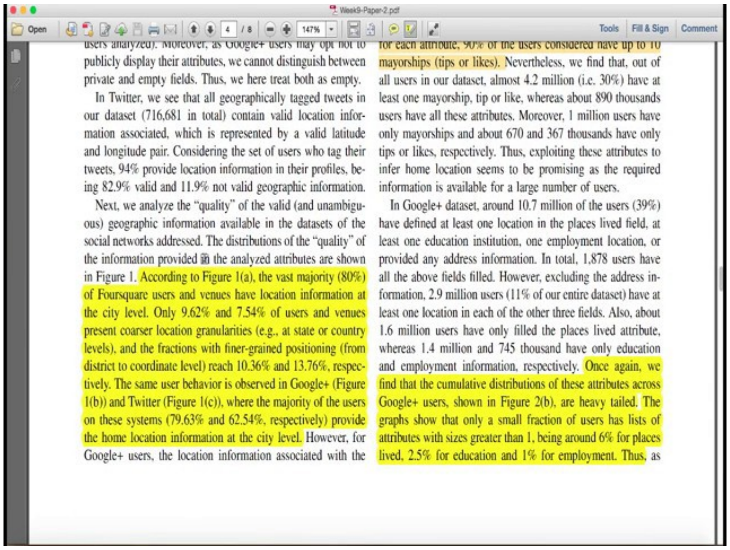
So, this basically would help you to find out, what is the amount of data that is available which is valid geographic information, valid and ambiguous geographic information, valid ambiguous geographic information and valid non geographic information and empty. What is this all mean I will I will try to explain this. Valid and ambiguous it is actually latitude or longitudinal, it is actually New Delhi; there is no ambiguity in it. Validambiguityitisnotclear,soitsaysnearTajMahalornearGovindpurimetro station, so these things are ambiguous.And non-GI – non-geographic information which could be I think as I said before, it could be somebody's heart, heart.

And information like that is actually it is not geographic information at all, and sometimes it could be actually empty. So, essentially that is what is been given in the values. They are all percentages which says user home city is about 95 percent, ambiguous is about 2.6 percent, and non-GI is 1.8 percent, and empty is about 0.2 percent.

So, in Foursquare, it is user home city and venue city. In Google plus, it is places lived address and education and employment. In Twitter, it is the user location geo tagged tweet right. So, this basically tells you different types of information are collected from different networks; I mean that is a whole body of research in terms of actually using these different sets of information from different social networks.

Then Foursquare it is user home city venue and venue. In Google plus, it is places lived address education and employment; in Twitters, it is user location and geo tagged tweet. So, if you look at the unambiguous geographic information for geo tagged tweet from Twitter it is about 100 percent. It is because all the tweets that where collected where actually at the 0.5 percent had geo tagged information in it, so that basically gives you a sense of what kind of data is collected. In terms of Google plus, 53 percent has an unambiguous education, so I studied that Carnegie Mellon University, so that is very very precise, there is unambiguity, there is no problem and actually recoding it or decoding it to a specific university.

(ReferSlideTime:14:57)



users analyzed). Moreover, as Google+ users may opt not to publicly display their attributes, we cannot distinguish between private and empty fields. Thus, we here treat both as empty.

In Twitter, we see that all geographically tagged tweets in our dataset (716,681 in total) contain valid location information associated, which is represented by a valid latitude and longitude pair. Considering the set of users who tag their tweets, 94% provide location information in their profiles, being 82.9% valid and 11.9% not valid geographic information.

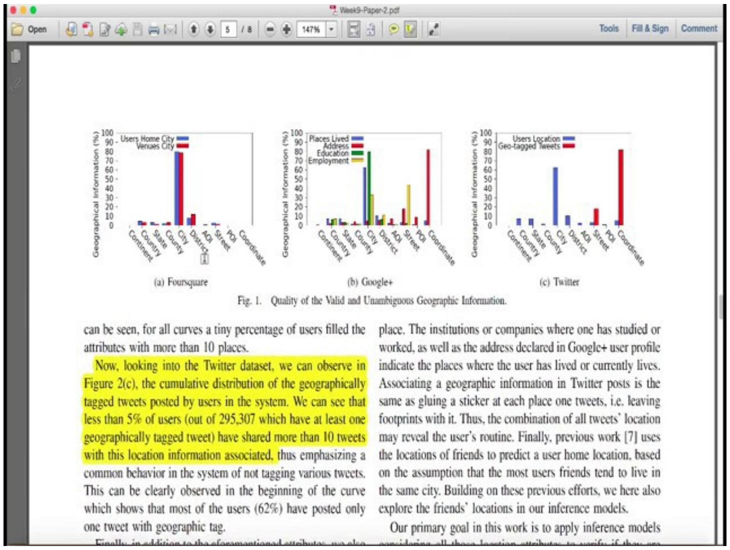
Next, we analyze the "quality" of the valid (and unambiguous) geographic information available in the datasets of the social networks addressed. The distributions of the "quality" of the information provided in the analyzed attributes are shown in Figure 1. According to Figure 1(a), the vast majority (80%) of Foursquare users and venues have location information at the city level. Only 9.62% and 7.54% of users and venues present coarser location granularities (e.g., at state or country levels), and the fractions with finer-grained positioning (from district to coordinate level) reach 10.36% and 13.76%, respectively. The same user behavior is observed in Google+ (Figure 1(b)) and Twitter (Figure 1(c)), where the majority of the users on these systems (79.63% and 62.54%, respectively) provide the home location information at the city level. However, for Google+ users, the location information associated with the

for each attribute, 90% of the users considered have up to 10 mayorships (tips or likes). Nevertheless, we find that, out of all users in our dataset, almost 4.2 million (i.e. 30%) have at least one mayorship, tip or like, whereas about 890 thousands users have all these attributes. Moreover, 1 million users have only mayorships and about 670 and 367 thousands have only tips or likes, respectively. Thus, exploiting these attributes to infer home location seems to be promising as the required information is available for a large number of users.

In Google+ dataset, around 10.7 million of the users (39%) have defined at least one location in the places lived field, at least one education institution, one employment location, or provided any address information. In total, 1,878 users have all the above fields filled. However, excluding the address information, 2.9 million users (11% of our entire dataset) have at least one location in each of the other three fields. Also, about 1.6 million users have only filled the places lived attribute, whereas 1.4 million and 745 thousand have only education and employment information, respectively. Once again, we find that the cumulative distributions of these attributes across Google+ users, shown in Figure 2(b), are heavy tailed. The graphs show that only a small fraction of users has lists of attributes with sizes greater than 1, being around 6% for places lived, 2.5% for education and 1% for employment. Thus, as

So, let us look at different figures, different analysis that is been done using this data. Figure 1 the vast majority 80 percent of Foursquare users and venues have location information at the city level. 9.6 percent and 7.4 percent of users and venues present coarser location granularities at a state or the country levels.

(ReferSlideTime:15:27)



(a) Foursquare

(b) Google+

(c) Twitter

Fig. 1. Quality of the Valid and Unambiguous Geographic Information.

can be seen, for all curves a tiny percentage of users filled the attributes with more than 10 places.

Now, looking into the Twitter dataset, we can observe in Figure 2(c), the cumulative distribution of the geographically tagged tweets posted by users in the system. We can see that less than 5% of users (out of 295,307 which have at least one geographically tagged tweet) have shared more than 10 tweets with this location information associated, thus emphasizing a common behavior in the system of not tagging various tweets. This can be clearly observed in the beginning of the curve which shows that most of the users (62%) have posted only one tweet with geographic tag.

Finally, in addition to the aforementioned attributes, we also

place. The institutions or companies where one has studied or worked, as well as the address declared in Google+ user profile indicate the places where the user has lived or currently lives. Associating a geographic information in Twitter posts is the same as gluing a sticker at each place one tweets, i.e. leaving footprints with it. Thus, the combination of all tweets' location may reveal the user's routine. Finally, previous work [7] uses the locations of friends to predict a user home location, based on the assumption that the most users friends tend to live in the same city. Building on these previous efforts, we here also explore the friends' locations in our inference models.

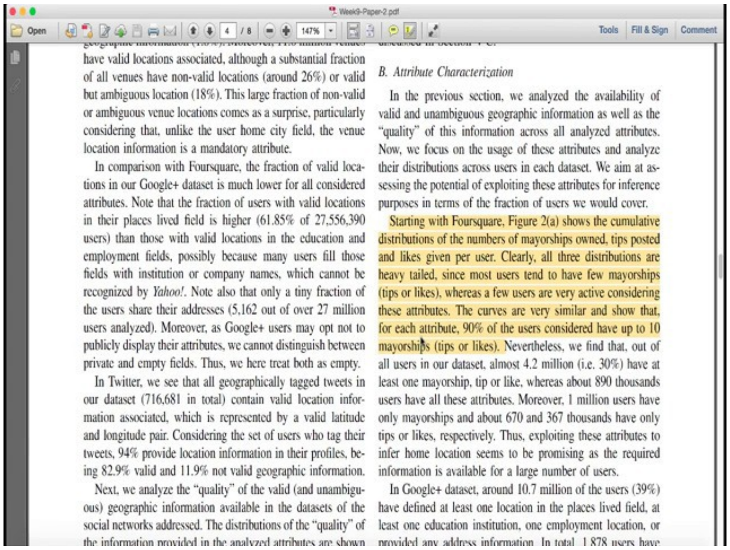
Our primary goal in this work is to apply inference models considering all these location attributes to verify if they are

So, we look at the figures. So, this is figure 1(a), 1(b) and 1(c). So, if you look at here quality of valid and unambiguous geographic information. Foursquare, if you look at city,citygives youtheusers'homecityandvenuecity;itisaboutclosetoeightypercent.

So, that is what is written here this says about vast majority 80 percent of Foursquare users and venues have location information at the city level. Some have at the country level, some have at the state level, some have at the street level, so that is the different level of details that the geographic information is available for the location from Foursquare.

So, if you look at Google plus, the information is maximum available for example, it is education that is available at a city level about 70 percent or 70 plus percentage, so thatis what is its written here.The same user behavior is observed in Google plus figure 1(b) and figure 1(c) where the majority of the users of the system 79.63, 62.54 respectively provide thehome location atthe citylevel. Cityis highestinterms ofplaces lived, cityis highest in terms of user location also, so that basically says that we should be able to actually get the city level of information without any problem, because large amount of data for the information about the users is available at the city level.

(ReferSlideTime:17:36)



have valid locations associated, although a substantial fraction of all venues have non-valid locations (around 26%) or valid but ambiguous location (18%). This large fraction of non-valid or ambiguous venue locations comes as a surprise, particularly considering that, unlike the user home city field, the venue location information is a mandatory attribute.

In comparison with Foursquare, the fraction of valid locations in our Google+ dataset is much lower for all considered attributes. Note that the fraction of users with valid locations in their places lived field is higher (61.85% of 27,556,390 users) than those with valid locations in the education and employment fields, possibly because many users fill those fields with institution or company names, which cannot be recognized by *Yahoo!*. Note also that only a tiny fraction of the users share their addresses (5,162 out of over 27 million users analyzed). Moreover, as Google+ users may opt not to publicly display their attributes, we cannot distinguish between private and empty fields. Thus, we here treat both as empty.

In Twitter, we see that all geographically tagged tweets in our dataset (716,681 in total) contain valid location information associated, which is represented by a valid latitude and longitude pair. Considering the set of users who tag their tweets, 94% provide location information in their profiles, being 82.9% valid and 11.9% not valid geographic information.

Next, we analyze the "quality" of the valid (and unambiguous) geographic information available in the datasets of the social networks addressed. The distributions of the "quality" of the information provided in the analyzed attributes are shown

B. Attribute Characterization

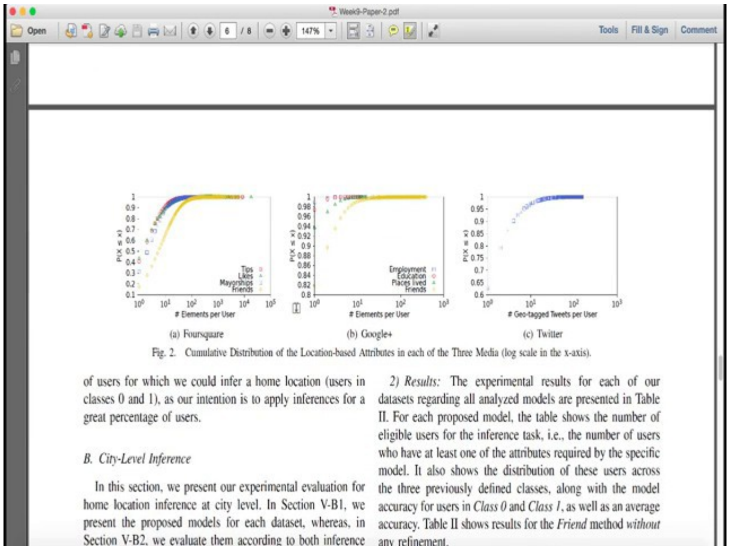
In the previous section, we analyzed the availability of valid and unambiguous geographic information as well as the "quality" of this information across all analyzed attributes. Now, we focus on the usage of these attributes and analyze their distributions across users in each dataset. We aim at assessing the potential of exploiting these attributes for inference purposes in terms of the fraction of users we would cover.

Starting with Foursquare, Figure 2(a) shows the cumulative distributions of the numbers of mayorships owned, tips posted and likes given per user. Clearly, all three distributions are heavy tailed, since most users tend to have few mayorships (tips or likes), whereas a few users are very active considering these attributes. The curves are very similar and show that, for each attribute, 90% of the users considered have up to 10 mayorships (tips or likes). Nevertheless, we find that, out of all users in our dataset, almost 4.2 million (i.e. 30%) have at least one mayorship, tip or like, whereas about 890 thousands users have all these attributes. Moreover, 1 million users have only mayorships and about 670 and 367 thousands have only tips or likes, respectively. Thus, exploiting these attributes to infer home location seems to be promising as the required information is available for a large number of users.

In Google+ dataset, around 10.7 million of the users (39%) have defined at least one location in the places lived field, at least one education institution, one employment location, or provided any address information. In total, 1,878 users have

Now, let us look at more analysis with this data. So, now, figure 2(a) shows the cumulative distributions of the numbers of mayorships owned, tips, posted and likes. If you remember even in the last paper, we saw this kind of graphs, which is to show the cumulative distribution of the number of the mayors, tips and dones right. So, if you rememberthegraphthereweresmallsetofpeoplewhohadalotofmayorships,anda large set of people who had less number of mayorships, so that is the kind of general social media behavior also.

(ReferSlideTime:18:40)



(a) Foursquare (b) Google+ (c) Twitter

Fig. 2. Cumulative Distribution of the Location-based Attributes in each of the Three Media (log scale in the x-axis).

of users for which we could infer a home location (users in classes 0 and 1), as our intention is to apply inferences for a great percentage of users.

B. City-Level Inference

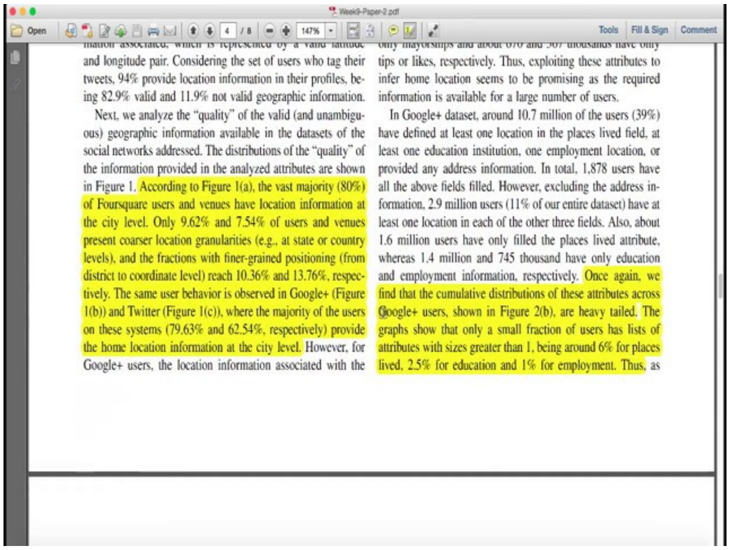
In this section, we present our experimental evaluation for home location inference at city level. In Section V-B1, we present the proposed models for each dataset, whereas, in Section V-B2, we evaluate them according to both inference

2) Results: The experimental results for each of our datasets regarding all analyzed models are presented in Table II. For each proposed model, the table shows the number of eligible users for the inference task, i.e., the number of users who have at least one of the attributes required by the specific model. It also shows the distribution of these users across the three previously defined classes, along with the model accuracy for users in *Class 0* and *Class 1*, as well as an average accuracy. Table II shows results for the *Friend* method *without* any refinement.

So, let us look at figure 2(a), which is giving you the distribution of the number of mayorships, tips and dones. So, if you see here, the figure 2(a) is giving you the cumulative distribution of location based attributes in each of the three media; a is for Foursquare,bisforGoogleplusandcisforTwitter.GiventhatTwitterhasallofthemas geo tagged you can clearly seen there is only one line, whereas in the other one there is, friends, mayorships, likes and tips that is in the Foursquare; in Google plus – employment, education, places lived and friends right. So, this is the graph and you can clearly see the graph is very similar to what we have seen in the past in terms of social media data.

So, clearly all three distributions are heavy-tailed (Refer Time: 19:31) which is what I just now said which is social media looking data, since most users tend to have few mayorshipswhereasafewusershaveveryactiveconsideringtheseattributes.The curves areverysimilarandshowsthateachattributes90percentoftheusersconsideredhaveup to10mayorshipsright.So,itisthesameprinciplePareto principlethatwetalkedabouta power law that wetalked about in the course all of that is playing into this data also.And thisisvery,veryimportanttoshowbecausethereviewersandthereaderscanactually believe that this data is actually representative of other social media research that has been done and analysis that has been done.

(ReferSlideTime:20:25)



mation associated, which is represented by a valid latitude and longitude pair. Considering the set of users who tag their tweets, 94% provide location information in their profiles, being 82.9% valid and 11.9% not valid geographic information.

Next, we analyze the "quality" of the valid (and unambiguous) geographic information available in the datasets of the social networks addressed. The distributions of the "quality" of the information provided in the analyzed attributes are shown in Figure 1. According to Figure 1(a), the vast majority (80%) of Foursquare users and venues have location information at the city level. Only 9.62% and 7.54% of users and venues present coarser location granularities (e.g., at state or country levels), and the fractions with finer-grained positioning (from district to coordinate level) reach 10.36% and 13.76%, respectively. The same user behavior is observed in Google+ (Figure 1(b)) and Twitter (Figure 1(c)), where the majority of the users on these systems (79.63% and 62.54%, respectively) provide the home location information at the city level. However, for Google+ users, the location information associated with the

only mayorships and about 670 and 367 thousands have only tips or likes, respectively. Thus, exploiting these attributes to infer home location seems to be promising as the required information is available for a large number of users.

In Google+ dataset, around 10.7 million of the users (39%) have defined at least one location in the places lived field, at least one education institution, one employment location, or provided any address information. In total, 1,878 users have all the above fields filled. However, excluding the address information, 2.9 million users (11% of our entire dataset) have at least one location in each of the other three fields. Also, about 1.6 million users have only filled the places lived attribute, whereas 1.4 million and 745 thousand have only education and employment information, respectively. Once again, we find that the cumulative distributions of these attributes across Google+ users, shown in Figure 2(b), are heavy tailed. The graphs show that only a small fraction of users has lists of attributes with sizes greater than 1, being around 6% for places lived, 2.5% for education and 1% for employment. Thus, as

Also if you see 2(b), again it's describing all the different social networks; figure 2(b) is showing you for Google plus the graph shows that only a small fraction of users has list of attributes with sizes greater than one being around 6 percent of places lived 2.5 percentage of education and 1 percent of employment.

(ReferSlideTime:20:57)

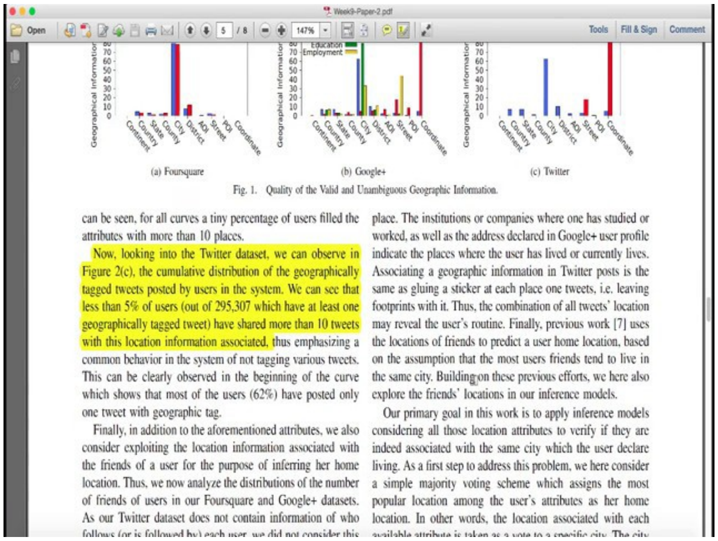


Fig. 1. Quality of the Valid and Unambiguous Geographic Information.

can be seen, for all curves a tiny percentage of users filled the attributes with more than 10 places.

Now, looking into the Twitter dataset, we can observe in Figure 2(c), the cumulative distribution of the geographically tagged tweets posted by users in the system. We can see that less than 5% of users (out of 295,307 which have at least one geographically tagged tweet) have shared more than 10 tweets with this location information associated, thus emphasizing a common behavior in the system of not tagging various tweets. This can be clearly observed in the beginning of the curve which shows that most of the users (62%) have posted only one tweet with geographic tag.

Finally, in addition to the aforementioned attributes, we also consider exploiting the location information associated with the friends of a user for the purpose of inferring her home location. Thus, we now analyze the distributions of the number of friends of users in our Foursquare and Google+ datasets. As our Twitter dataset does not contain information of who follows (or is followed by) each user, we did not consider this

place. The institutions or companies where one has studied or worked, as well as the address declared in Google+ user profile indicate the places where the user has lived or currently lives. Associating a geographic information in Twitter posts is the same as gluing a sticker at each place one tweets, i.e. leaving footprints with it. Thus, the combination of all tweets' location may reveal the user's routine. Finally, previous work [7] uses the locations of friends to predict a user home location, based on the assumption that the most users friends tend to live in the same city. Building on these previous efforts, we here also explore the friends' locations in our inference models.

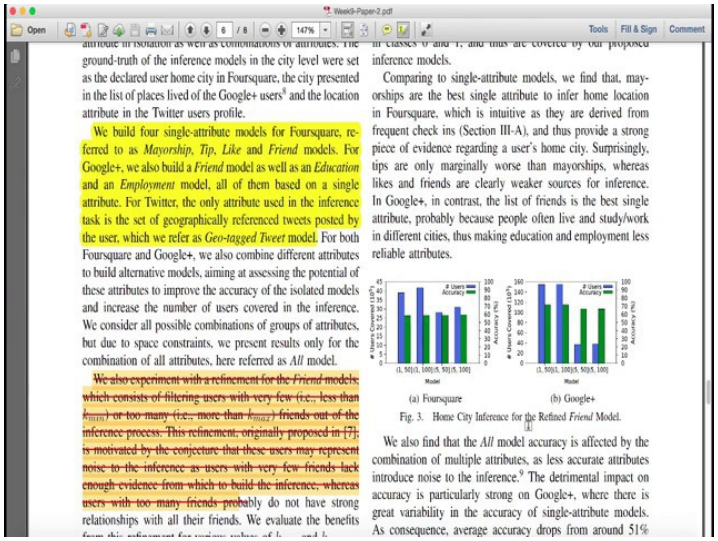
Our primary goal in this work is to apply inference models considering all those location attributes to verify if they are indeed associated with the same city which the user declare living. As a first step to address this problem, we here consider a simple majority voting scheme which assigns the most popular location among the user's attributes as her home location. In other words, the location associated with each available attribute is taken as a vote to a specific city. The city

If you look at the Twitter data, we can observe that figure 2(c), the cumulative distribution of geographically tagged tweets posted by users in the system. We can see that less than 5 percent of the users have shared more than 10 tweets with this location information associated, which is again small percentage of people doing location information sharing, more than 10 tweets with the location associated with them.

Also now let us get into inferring location. The methodology that this paper uses is very same to the last methodology, which is written in this paragraph. We group users into three classes, class 0 consists of users who have only one vote that is only one location information that is predominant, and that is only one. Thus, allowing only a unique option tobeassigned fortheuser'shomecity.Class1containstheuserwhohasmultiple votes with the predominant location across them.

And the class 2 as we have seen in the last paper also consists of users with multiple votes in which there is no single location that stands out. So, three categories of classes three classes that they are made 0, 1 and 2. Wewill see the table with 0, 1 and 2 that lets this locatethehowthedatawas collected, howtheanalysis was done,how thebucketing was done. The results of our experimental evaluation are assessed using two metrics which measure the effectiveness of the proposed model. Accuracy is the fraction of correct inferences of users of class 0, or class 1, right yet again there are (Refer Time: 22:58) the current thing with class 2 will not work.

(ReferSlideTime:23:04)



attribute in isolation as well as combinations of attributes. The ground-truth of the inference models in the city level were set as the declared user home city in Foursquare, the city presented in the list of places lived of the Google+ users[8] and the location attribute in the Twitter users profile.

We build four single-attribute models for Foursquare, referred to as *Mayorship*, *Tip*, *Like* and *Friend* models. For Google+, we also build a *Friend* model as well as an *Education* and an *Employment* model, all of them based on a single attribute. For Twitter, the only attribute used in the inference task is the set of geographically referenced tweets posted by the user, which we refer as *Geo-tagged Tweet* model. For both Foursquare and Google+, we also combine different attributes to build alternative models, aiming at assessing the potential of these attributes to improve the accuracy of the isolated models and increase the number of users covered in the inference. We consider all possible combinations of groups of attributes, but due to space constraints, we present results only for the combination of all attributes, here referred as *All* model.

~~We also experiment with a refinement for the *Friend* models, which consists of filtering users with very few (i.e., less than k_{min}) or too many (i.e., more than k_{max}) friends out of the inference process. This refinement, originally proposed in [7], is motivated by the conjecture that these users may represent noise to the inference as users with very few friends lack enough evidence from which to build the inference, whereas users with too many friends prob~~ably do not have strong relationships with all their friends. We evaluate the benefits from this refinement for various values of k_{min} and k_{max}

in classes 0 and 1, and thus are covered by our proposed inference models.

Comparing to single-attribute models, we find that, mayorships are the best single attribute to infer home location in Foursquare, which is intuitive as they are derived from frequent check ins (Section III-A), and thus provide a strong piece of evidence regarding a user's home city. Surprisingly, tips are only marginally worse than mayorships, whereas likes and friends are clearly weaker sources for inference. In Google+, in contrast, the list of friends is the best single attribute, probably because people often live and study/work in different cities, thus making education and employment less reliable attributes.

(a) Foursquare

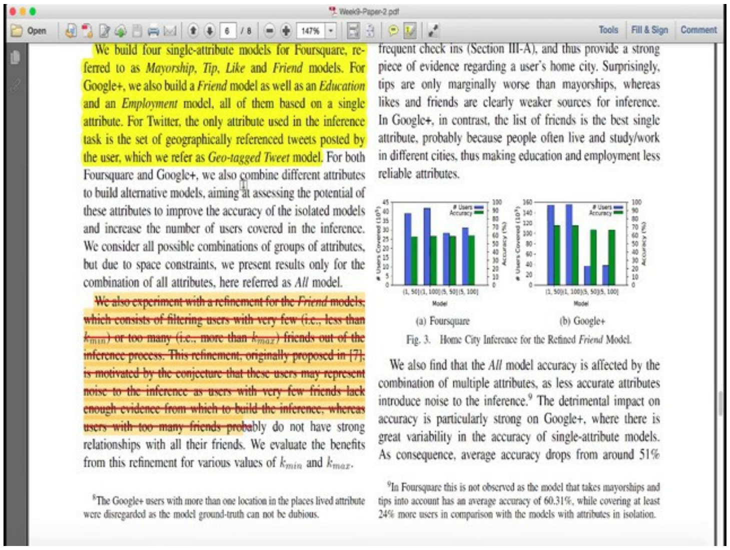
(b) Google+

Fig. 3. Home City Inference for the Refined *Friend* Model.

We also find that the *All* model accuracy is affected by the combination of multiple attributes, as less accurate attributes introduce noise to the inference.[9] The detrimental impact on accuracy is particularly strong on Google+, where there is great variability in the accuracy of single-attribute models. As consequence, average accuracy drops from around 51%

So, the model that was built was four single-attribute models for Foursquare, referred to as mayorship, tip, like and friends. For Google plus, friend model and education and employment model all of them based on single-attribute. For Twitter, the only attribute used in the inferences task is the geo tagged location right. So, basically this explains what details were used in collect and making the inference about the location.

(ReferSlideTime:23:44)



We build four single-attribute models for Foursquare, referred to as *Mayorship*, *Tip*, *Like* and *Friend* models. For Google+, we also build a *Friend* model as well as an *Education* and an *Employment* model, all of them based on a single attribute. For Twitter, the only attribute used in the inference task is the set of geographically referenced tweets posted by the user, which we refer as *Geo-tagged Tweet* model. For both Foursquare and Google+, we also combine different attributes to build alternative models, aiming at assessing the potential of these attributes to improve the accuracy of the isolated models and increase the number of users covered in the inference. We consider all possible combinations of groups of attributes, but due to space constraints, we present results only for the combination of all attributes, here referred as *All* model.

~~We also experiment with a refinement for the *Friend* models, which consists of filtering users with very few (i.e., less than k_{min}) or too many (i.e., more than k_{max}) friends out of the inference process. This refinement, originally proposed in [7], is motivated by the conjecture that these users may represent noise to the inference as users with very few friends lack enough evidence from which to build the inference, whereas users with too many friends proba~~bly do not have strong relationships with all their friends. We evaluate the benefits from this refinement for various values of k_{min} and k_{max}.

[8] The Google+ users with more than one location in the places lived attribute were disregarded as the model ground-truth can not be dubious.

frequent check ins (Section III-A), and thus provide a strong piece of evidence regarding a user's home city. Surprisingly, tips are only marginally worse than mayorships, whereas likes and friends are clearly weaker sources for inference. In Google+, in contrast, the list of friends is the best single attribute, probably because people often live and study/work in different cities, thus making education and employment less reliable attributes.

(a) Foursquare

(b) Google+

Fig. 3. Home City Inference for the Refined *Friend* Model.

We also find that the *All* model accuracy is affected by the combination of multiple attributes, as less accurate attributes introduce noise to the inference.[9] The detrimental impact on accuracy is particularly strong on Google+, where there is great variability in the accuracy of single-attribute models. As consequence, average accuracy drops from around 51%

[9] In Foursquare this is not observed as the model that takes mayorships and tips into account has an average accuracy of 60.31%, while covering at least 24% more users in comparison with the models with attributes in isolation.

Also this is also information we also experiment with the refinement of the friends' model which consists of filtering users. So, another wayjust think about the another way of looking at the friends model is the use the friends from that location to make the decision, so that is filtering with a very few, less than k kilometers, or too many that is more than k max friends out of the inference process.

The refinement originally proposed is motivated by the conjecture that these users may representnoisetotheinferenceasuserswithfewfriendslackenoughevidenceforwhich to build the inference, whereas users with too many friends probably do not have strong relationships with all their friends. It is basically saying that we build a model where we take the users with very few, less than few kilometers; because they are not going to be connected.Alot of friends who may be connected from that location also will not lead a lot of information.

(ReferSlideTime:25:03)

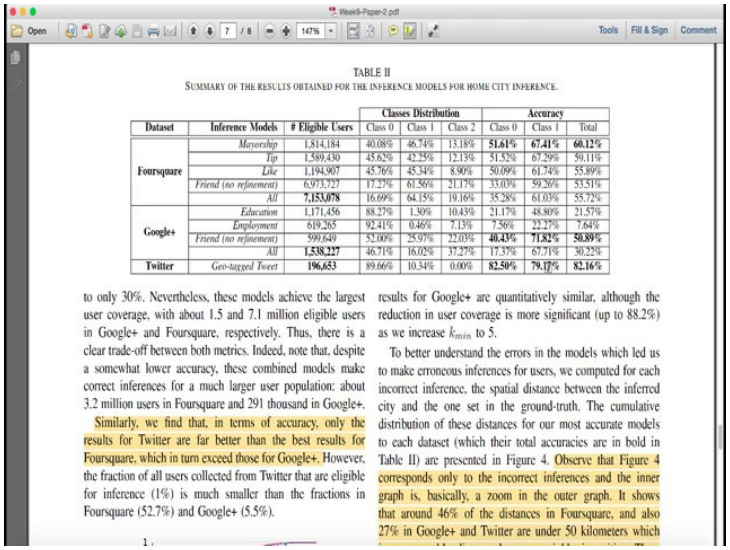


TABLE II
SUMMARY OF THE RESULTS OBTAINED FOR THE INFERENCE MODELS FOR HOME CITY INFERENCE.

			Classes Distribution			Accuracy		
Dataset	**Inference Models**	**# Eligible Users**	Class 0	Class 1	Class 2	Class 0	Class 1	Total
Foursquare	*Mayorship*	1,814,184	40.08%	46.74%	13.18%	**51.61%**	**67.41%**	**60.12%**
	Tip	1,589,430	45.62%	42.25%	12.13%	51.52%	67.29%	59.11%
	Like	1,194,907	45.76%	45.34%	8.90%	50.09%	61.74%	55.89%
	Friend (no refinement)	6,973,727	17.27%	61.56%	21.17%	33.03%	59.26%	53.51%
	All	**7,153,078**	16.69%	64.15%	19.16%	35.28%	61.03%	55.72%
Google+	*Education*	1,171,456	88.27%	1.30%	10.43%	21.17%	48.80%	21.57%
	Employment	619,265	92.41%	0.46%	7.13%	7.56%	22.27%	7.64%
	Friend (no refinement)	599,649	52.00%	25.97%	22.03%	**40.43%**	**71.82%**	**50.89%**
	All	**1,538,227**	46.71%	16.02%	37.27%	17.37%	67.71%	30.22%
Twitter	*Geo-tagged Tweet*	**196,653**	89.66%	10.34%	0.00%	**82.50%**	**79.17%**	**82.16%**

to only 30%. Nevertheless, these models achieve the largest user coverage, with about 1.5 and 7.1 million eligible users in Google+ and Foursquare, respectively. Thus, there is a clear trade-off between both metrics. Indeed, note that, despite a somewhat lower accuracy, these combined models make correct inferences for a much larger user population: about 3.2 million users in Foursquare and 291 thousand in Google+.

Similarly, we find that, in terms of accuracy, only the results for Twitter are far better than the best results for Foursquare, which in turn exceed those for Google+. However, the fraction of all users collected from Twitter that are eligible for inference (1%) is much smaller than the fractions in Foursquare (52.7%) and Google+ (5.5%).

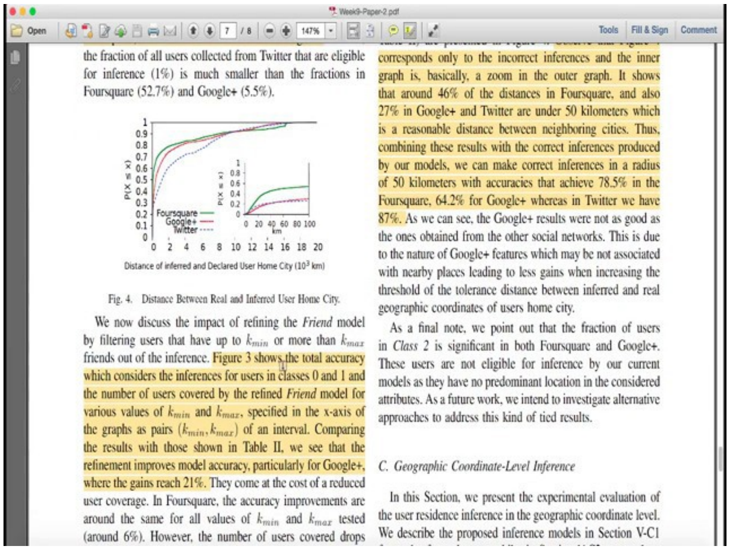
results for Google+ are quantitatively similar, although the reduction in user coverage is more significant (up to 88.2%) as we increase k_{min} to 5.

To better understand the errors in the models which led us to make erroneous inferences for users, we computed for each incorrect inference, the spatial distance between the inferred city and the one set in the ground-truth. The cumulative distribution of these distances for our most accurate models to each dataset (which their total accuracies are in bold in Table II) are presented in Figure 4. Observe that Figure 4 corresponds only to the incorrect inferences and the inner graph is, basically, a zoom in the outer graph. It shows that around 46% of the distances in Foursquare, and also 27% in Google+ and Twitter are under 50 kilometers which

So, this table is the most important analysis orinference from this paper which is to see the summary of results obtained for inference models for a home and home city inference. Remember, we did for three networks - Foursquare, Google plus and Twitter, the inference models that was used for mayorships, tip, like, friend all, education, employment, friend all geo tagged tweets.

Classes distributed 0, 1 and 2; classes 0 and 1 are the only two things that can be done withthisdata(ReferTime:25:45),soitisdone51.So,thewaytoreaditisthatjustusing mayorship, in the class 0, 51.61 percent you can identifythe home cityfor that particular user in the category who has Foursquare account and mayorship data. Class 1, 67 percent; class 1 is basically there are multiple locations, one being predominant. Google plus, the highest seems to be with friend, no refinement; and in tweets, it is since the geo located the accuracy is also being more than anything else.

(ReferSlideTime:26:43)



the fraction of all users collected from Twitter that are eligible for inference (1%) is much smaller than the fractions in Foursquare (52.7%) and Google+ (5.5%).

Fig. 4. Distance Between Real and Inferred User Home City.

We now discuss the impact of refining the *Friend* model by filtering users that have up to k_{min} or more than k_{max} friends out of the inference. Figure 3 shows the total accuracy which considers the inferences for users in classes 0 and 1 and the number of users covered by the refined *Friend* model for various values of k_{min} and k_{max}, specified in the x-axis of the graphs as pairs (k_{min}, k_{max}) of an interval. Comparing the results with those shown in Table II, we see that the refinement improves model accuracy, particularly for Google+, where the gains reach 21%. They come at the cost of a reduced user coverage. In Foursquare, the accuracy improvements are around the same for all values of k_{min} and k_{max} tested (around 6%). However, the number of users covered drops

corresponds only to the incorrect inferences and the inner graph is, basically, a zoom in the outer graph. It shows that around 46% of the distances in Foursquare, and also 27% in Google+ and Twitter are under 50 kilometers which is a reasonable distance between neighboring cities. Thus, combining these results with the correct inferences produced by our models, we can make correct inferences in a radius of 50 kilometers with accuracies that achieve 78.5% in the Foursquare, 64.2% for Google+ whereas in Twitter we have 87%. As we can see, the Google+ results were not as good as the ones obtained from the other social networks. This is due to the nature of Google+ features which may be not associated with nearby places leading to less gains when increasing the threshold of the tolerance distance between inferred and real geographic coordinates of users home city.

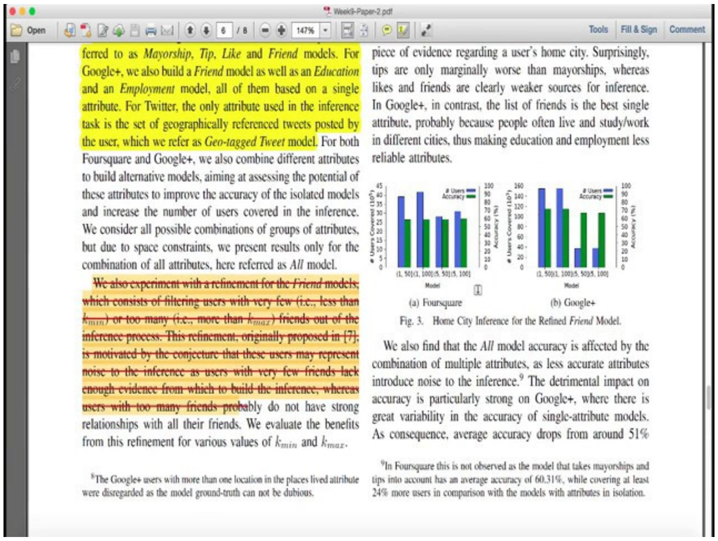
As a final note, we point out that the fraction of users in *Class 2* is significant in both Foursquare and Google+. These users are not eligible for inference by our current models as they have no predominant location in the considered attributes. As a future work, we intend to investigate alternative approaches to address this kind of tied results.

C. Geographic Coordinate-Level Inference

In this Section, we present the experimental evaluation of the user residence inference in the geographic coordinate level. We describe the proposed inference models in Section V-C1

So, you can that is what I have said the accuracy for the Twitter seems to be higher than the restofit. Let us look atfigure 3,andthen wewillgoback tothe description ofit. So, if you look at figure 3, figure 3 shows the total accuracy which considers the inferences for users in class 0, and 1.And the number of users covered bythe refined friend model, for the various values of k min and k max specified in the x-axis of the graph. Often comparing the results with those in table 2, we see that the refinement improves model accuracy particularly for Google plus where the gain is about 21 percent.

(ReferSlideTime:27:30)



ferred to as *Mayorship*, *Tip*, *Like* and *Friend* models. For Google+, we also build a *Friend* model as well as an *Education* and an *Employment* model, all of them based on a single attribute. For Twitter, the only attribute used in the inference task is the set of geographically referenced tweets posted by the user, which we refer as *Geo-tagged Tweet* model. For both Foursquare and Google+, we also combine different attributes to build alternative models, aiming at assessing the potential of these attributes to improve the accuracy of the isolated models and increase the number of users covered in the inference. We consider all possible combinations of groups of attributes, but due to space constraints, we present results only for the combination of all attributes, here referred as *All* model.

~~We also experiment with a refinement for the *Friend* models, which consists of filtering users with very few (i.e., less than k_{min}) or too many (i.e., more than k_{max}) friends out of the inference process. This refinement, originally proposed in [7], is motivated by the conjecture that these users may represent noise to the inference as users with very few friends lack enough evidence from which to build the inference, whereas users with too many friends prob~~ably do not have strong relationships with all their friends. We evaluate the benefits from this refinement for various values of k_{min} and k_{max}.

piece of evidence regarding a user's home city. Surprisingly, tips are only marginally worse than mayorships, whereas likes and friends are clearly weaker sources for inference. In Google+, in contrast, the list of friends is the best single attribute, probably because people often live and study/work in different cities, thus making education and employment less reliable attributes.

(a) Foursquare

(b) Google+

Fig. 3. Home City Inference for the Refined *Friend* Model.

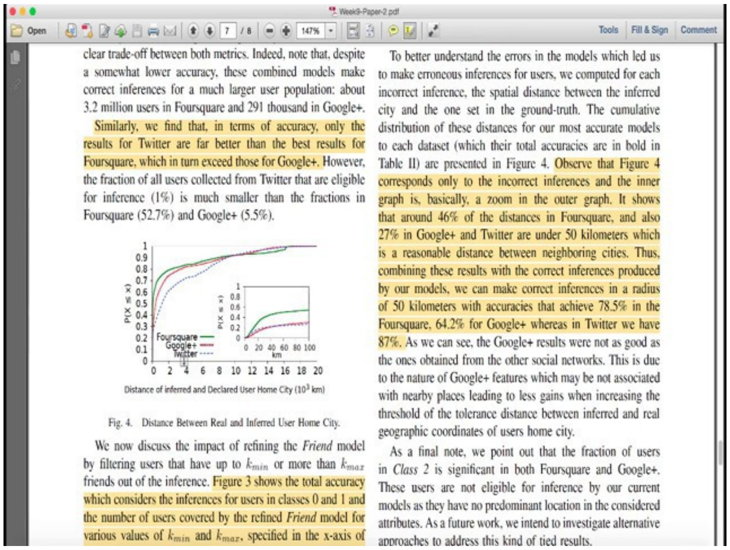
We also find that the *All* model accuracy is affected by the combination of multiple attributes, as less accurate attributes introduce noise to the inference.[9] The detrimental impact on accuracy is particularly strong on Google+, where there is great variability in the accuracy of single-attribute models. As consequence, average accuracy drops from around 51%

[8] The Google+ users with more than one location in the places lived attribute were disregarded as the model ground-truth can not be dubious.

[9] In Foursquare this is not observed as the model that takes mayorships and tips into account has an average accuracy of 60.31%, while covering at least 24% more users in comparison with the models with attributes in isolation.

So, essentially this is x-axis is the model that was used within the kilometers and y-axisis the percentage of users that were covered. So, if you see Google plus there is if you will infer the home city for a refined friend model, which is what we said where the min and the max were removed Google plus seems to be doing much better than the added advantage of removing these friends is higher for Google plus compare to Foursquare. That is the inference there.

(ReferSlideTime:28:19)



clear trade-off between both metrics. Indeed, note that, despite a somewhat lower accuracy, these combined models make correct inferences for a much larger user population: about 3.2 million users in Foursquare and 291 thousand in Google+.

Similarly, we find that, in terms of accuracy, only the results for Twitter are far better than the best results for Foursquare, which in turn exceed those for Google+. However, the fraction of all users collected from Twitter that are eligible for inference (1%) is much smaller than the fractions in Foursquare (52.7%) and Google+ (5.5%).

Fig. 4. Distance Between Real and Inferred User Home City.

We now discuss the impact of refining the *Friend* model by filtering users that have up to k_{min} or more than k_{max} friends out of the inference. Figure 3 shows the total accuracy which considers the inferences for users in classes 0 and 1 and the number of users covered by the refined *Friend* model for various values of k_{min} and k_{max}, specified in the x-axis of

To better understand the errors in the models which led us to make erroneous inferences for users, we computed for each incorrect inference, the spatial distance between the inferred city and the one set in the ground-truth. The cumulative distribution of these distances for our most accurate models to each dataset (which their total accuracies are in bold in Table II) are presented in Figure 4. Observe that Figure 4 corresponds only to the incorrect inferences and the inner graph is, basically, a zoom in the outer graph. It shows that around 46% of the distances in Foursquare, and also 27% in Google+ and Twitter are under 50 kilometers which is a reasonable distance between neighboring cities. Thus, combining these results with the correct inferences produced by our models, we can make correct inferences in a radius of 50 kilometers with accuracies that achieve 78.5% in the Foursquare, 64.2% for Google+ whereas in Twitter we have 87%. As we can see, the Google+ results were not as good as the ones obtained from the other social networks. This is due to the nature of Google+ features which may be not associated with nearby places leading to less gains when increasing the threshold of the tolerance distance between inferred and real geographic coordinates of users home city.

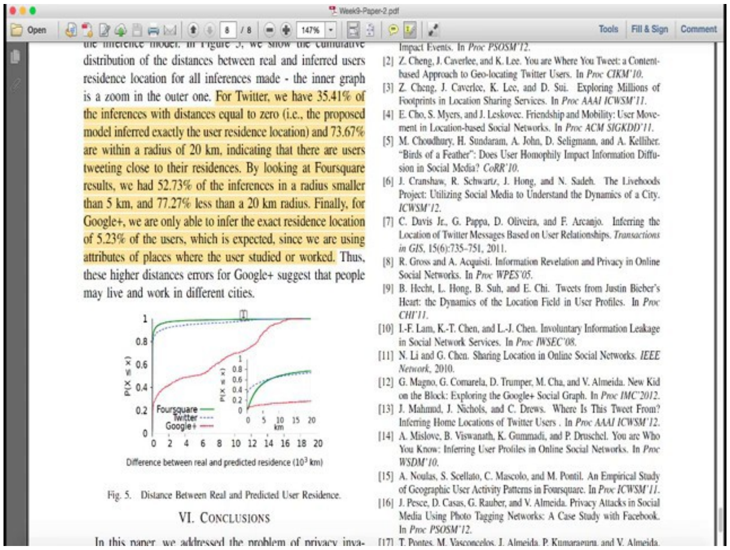
As a final note, we point out that the fraction of users in *Class 2* is significant in both Foursquare and Google+. These users are not eligible for inference by our current models as they have no predominant location in the considered attributes. As a future work, we intend to investigate alternative approaches to address this kind of tied results.

See there is another interesting inference, similar graph we saw on the last paper also. TherewesawonlyforFoursquare;herewe areseeingitforFoursquare,Googleplusand Twitter. Figure 4, corresponds only to the incorrect inferences and the inner graph is basically zoomed into the outer graph. It shows that 46 percent of the distances in Foursquare, and also 27 percent in Google plus, and Twitter are under 50 kilometers. So, that is what is this here 50 kilometers is this part of the inside graph. 50 kilometers is reasonable distance between neighboring cities.

Thus combining these results with the correct inference produced we can make correct inferences in a radius of 50 kilometers with the accuracies that achieves 78.5 percent in Foursquare, 64 in Google plus and 87 in Twitter,which are the things that we saw in the table earlier. This is the representation in the graph which is x-axis is the distance of inferred and declared home city which is something that I declared. And something we wereabletomyaccountPKponguruaccounthasalocationandIinferredthroughthe process the location what is the distance between these two, the lower the difference the better.

The inside graph is just showing you assumed immersion which shows that about 50 kilometers we were able to get about 46 percent, where if you see here, 50 percent and if this is about 46 percent. So, they just shows you that we are able to actually identify 46 percentofthedistanceinFoursquareisactuallylessthanerrorof50kilometers,which isjustneighboringcities,neighboringplaces, orsometimesitcouldbejustinthesame city.

(ReferSlideTime:30:46)



the inference model. In Figure 5, we show the cumulative distribution of the distances between real and inferred users residence location for all inferences made - the inner graph is a zoom in the outer one. For Twitter, we have 35.41% of the inferences with distances equal to zero (i.e., the proposed model inferred exactly the user residence location) and 73.67% are within a radius of 20 km, indicating that there are users tweeting close to their residences. By looking at Foursquare results, we had 52.73% of the inferences in a radius smaller than 5 km, and 77.27% less than a 20 km radius. Finally, for Google+, we are only able to infer the exact residence location of 5.23% of the users, which is expected, since we are using attributes of places where the user studied or worked. Thus, these higher distances errors for Google+ suggest that people may live and work in different cities.

Fig. 5. Distance Between Real and Predicted User Residence.

VI. CONCLUSIONS

In this paper, we addressed the problem of privacy inva-

Impact Events. In *Proc PSOSM'12*.

[2] Z. Cheng, J. Caverlee, and K. Lee. You are Where You Tweet: a Content-based Approach to Geo-locating Twitter Users. In *Proc CIKM'10*.

[3] Z. Cheng, J. Caverlee, K. Lee, and D. Sui. Exploring Millions of Footprints in Location Sharing Services. In *Proc AAAI ICWSM'11*.

[4] E. Cho, S. Myers, and J. Leskovec. Friendship and Mobility: User Movement in Location-based Social Networks. In *Proc ACM SIGKDD'11*.

[5] M. Choudhury, H. Sundaram, A. John, D. Seligmann, and A. Kelliher. "Birds of a Feather": Does User Homophily Impact Information Diffusion in Social Media? *CoRR'10*.

[6] J. Cranshaw, R. Schwartz, J. Hong, and N. Sadeh. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. *ICWSM'12*.

[7] C. Davis Jr., G. Pappa, D. Oliveira, and F. Arcanjo. Inferring the Location of Twitter Messages Based on User Relationships. *Transactions in GIS*, 15(6):735–751, 2011.

[8] R. Gross and A. Acquisti. Information Revelation and Privacy in Online Social Networks. In *Proc WPES'05*.

[9] B. Hecht, L. Hong, B. Suh, and E. Chi. Tweets from Justin Bieber's Heart: the Dynamics of the Location Field in User Profiles. In *Proc CHI'11*.

[10] I.-F. Lam, K.-T. Chen, and L.-J. Chen. Involuntary Information Leakage in Social Network Services. In *Proc IWSEC'08*.

[11] N. Li and G. Chen. Sharing Location in Online Social Networks. *IEEE Network*, 2010.

[12] G. Magno, G. Comarela, D. Trumper, M. Cha, and V. Almeida. New Kid on the Block: Exploring the Google+ Social Graph. In *Proc IMC'2012*.

[13] J. Mahmud, J. Nichols, and C. Drews. Where Is This Tweet From? Inferring Home Locations of Twitter Users . In *Proc AAAI ICWSM'12*.

[14] A. Mislove, B. Viswanath, K. Gummadi, and P. Druschel. You are Who You Know: Inferring User Profiles in Online Social Networks. In *Proc WSDM'10*.

[15] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An Empirical Study of Geographic User Activity Patterns in Foursquare. In *Proc ICWSM'11*.

[16] J. Pesce, D. Casas, G. Rauber, and V. Almeida. Privacy Attacks in Social Media Using Photo Tagging Networks: A Case Study with Facebook. In *Proc PSOSM'12*.

[17] T. Pontes, M. Vasconcelos, J. Almeida, P. Kumaraguru, and V. Almeida.

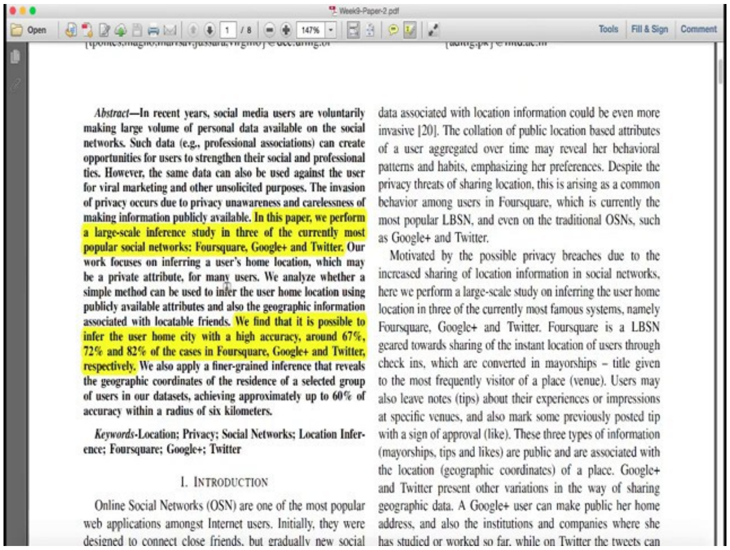
So, now the same thing you can actually do it for the residence right. Now what we didin this graph is basically showing you only the home city, whereas this graph is actually showing you the home residence. So, here is the graph for residence; red is Google plus; blueisTwitter;andgreenisFoursquare.Youcanseetheinsidethegraphalsoherewhich is from 0 to 20 kilometers, whereas this is 0 to 20, but they are all ten to the power of 1000 kilometers is the distance here.

So, you can see that for Twitter we have 35 percent, where is Twitter, so Twitter is blue line,blue line is here,35percentofthe inferences withdistanceequalto0.So,thatis the starting point here if you see, that is the proposed model inferred exactly the user residence. And the reason why this is so high and this is so accurate is because, we are, tweetswerecollectedwhichwereactuallygeotaggedright.And73.67percentarewithin the 20 kilometers radius.

So, if we see here this is 20 kilometers and if we see the blue line it is here that is about 76 percent, 73 percent, which is within the 20 kilometer difference, we were able to find out where the home is which is pretty good. Indicating that that there are users tweeting close to their residences, because I could be living in Okhla, I could be tweeting from somewhere near Jawaharlal Nehru Stadium in Delhi which is less than 20 kilometers, I went to watch a match and I actuallyposted tweet which is also geo tagged,sothat is the kind of 20 kilometers that we can get.

By looking at Foursquare results, we find that Foursquare is green. We find that 52 percent of the inferences in the radius smaller than 5 kilometers. So, if you see heregreen one, if you go at that point it is about 52 percent, 52.73 percent less than 5 kilometers; 77 percent less than 20 kilometers, that is here, 77 percent. Finally, for Google plus, we are only able to infer the exact residence of 5.23 percent of the users, which is expected since we are using attributes of places where the user studied or worked, because here we are only using their employment and education details right.

(ReferSlideTime:33:55)



Abstract—In recent years, social media users are voluntarily making large volume of personal data available on the social networks. Such data (e.g., professional associations) can create opportunities for users to strengthen their social and professional ties. However, the same data can also be used against the user for viral marketing and other unsolicited purposes. The invasion of privacy occurs due to privacy unawareness and carelessness of making information publicly available. In this paper, we perform a large-scale inference study in three of the currently most popular social networks: Foursquare, Google+ and Twitter. Our work focuses on inferring a user's home location, which may be a private attribute, for many users. We analyze whether a simple method can be used to infer the user home location using publicly available attributes and also the geographic information associated with locatable friends. We find that it is possible to infer the user home city with a high accuracy, around 67%, 72% and 82% of the cases in Foursquare, Google+ and Twitter, respectively. We also apply a finer-grained inference that reveals the geographic coordinates of the residence of a selected group of users in our datasets, achieving approximately up to 60% of accuracy within a radius of six kilometers.

Keywords-Location; Privacy; Social Networks; Location Inference; Foursquare; Google+; Twitter

I. Introduction

Online Social Networks (OSN) are one of the most popular web applications amongst Internet users. Initially, they were designed to connect close friends, but gradually new social

data associated with location information could be even more invasive [20]. The collation of public location based attributes of a user aggregated over time may reveal her behavioral patterns and habits, emphasizing her preferences. Despite the privacy threats of sharing location, this is arising as a common behavior among users in Foursquare, which is currently the most popular LBSN, and even on the traditional OSNs, such as Google+ and Twitter.

Motivated by the possible privacy breaches due to the increased sharing of location information in social networks, here we perform a large-scale study on inferring the user home location in three of the currently most famous systems, namely Foursquare, Google+ and Twitter. Foursquare is a LBSN geared towards sharing of the instant location of users through check ins, which are converted in mayorships – title given to the most frequently visitor of a place (venue). Users may also leave notes (tips) about their experiences or impressions at specific venues, and also mark some previously posted tip with a sign of approval (like). These three types of information (mayorships, tips and likes) are public and are associated with the location (geographic coordinates) of a place. Google+ and Twitter present other variations in the way of sharing geographic data. A Google+ user can make public her home address, and also the institutions and companies where she has studied or worked so far, while on Twitter the tweets can

So, that is how this paper ends, which is to show that let us go to the abstract again, which is to show that they used they used data from Foursquare, Google plus and Twitter. They used this data to infer the home location. This is an extension or the next step for the last paper that we saw which was done only on Foursquare. And they were abletoactuallyshowthatabout67,72and82percentwiththataccuracytheywereable tofindoutthehomecity,andhomelocation,for,withahighaccuracyintermsofTwitter and then Foursquare, but with less data in a Google plus.

Withthat,Iwillstopthisparticularpaper.Iwillseeyou soon.