

Privacy and Security in Online Social Networks
Prof. Ponnuram Kumaraguru (“PK”)
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Week – 8.2
Lecture – 27
Anonymous Networks

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.

(Refer Slide Time: 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 what anonymous networks are with some examples and I will also show you some

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.

(Refer Slide Time: 02:10)



The slide features a title 'Why use Anonymous Networks?' in teal text. Below the title is a bulleted list of five points: 'Increasing awareness of privacy', 'Snowden disclosures', 'PRISM Surveillance program', 'Bal Thackeray incident', and 'Many other incidents around the world'. To the right of the text is a vertical black bar. At the bottom right of the slide, there is a small video inset showing a man in a light-colored shirt speaking.

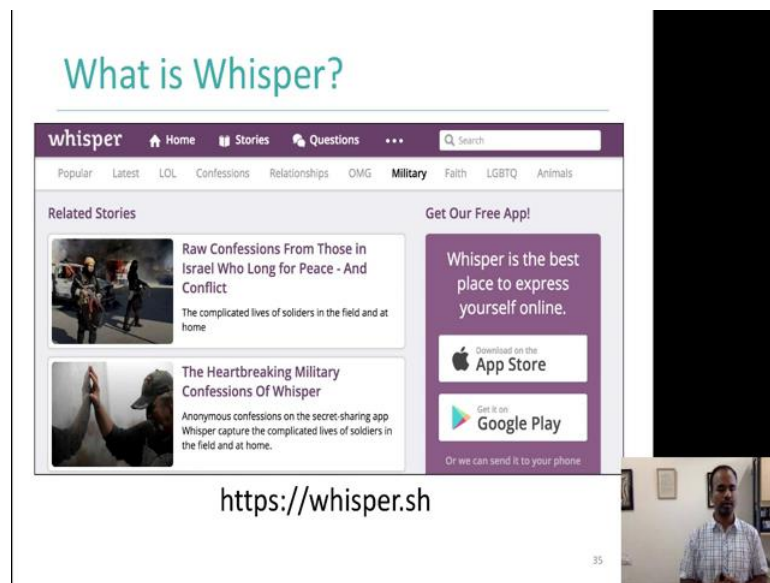
- Increasing awareness of privacy
- Snowden disclosures
- PRISM Surveillance program
- Bal Thackeray incident
- Many other incidents around the world

Why do you need anonymous social network? So, we already have Facebook, we already 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 were also incidences like Snowden; projects like PRISM were the information that is publicly available or the information that is available to these organizations can be 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. Because for example, If I do a post on facebook, if I do a post on twitter it is

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.

(Refer Slide Time: 04:13)



So, here is that slide screen shot of the website whisper dot sh in the content is organized as in the top order in this image **popular, latest, Lol,** confessions, relationship, Oh my god and they create **these** categories **so that the content** that is uploaded on whisper gets into 1 of these categories and the URL is 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.

(Refer Slide Time: 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 look at video now and then I will describe some details which is from the video and other features of whisper.

(Refer Slide Time: 05:22)

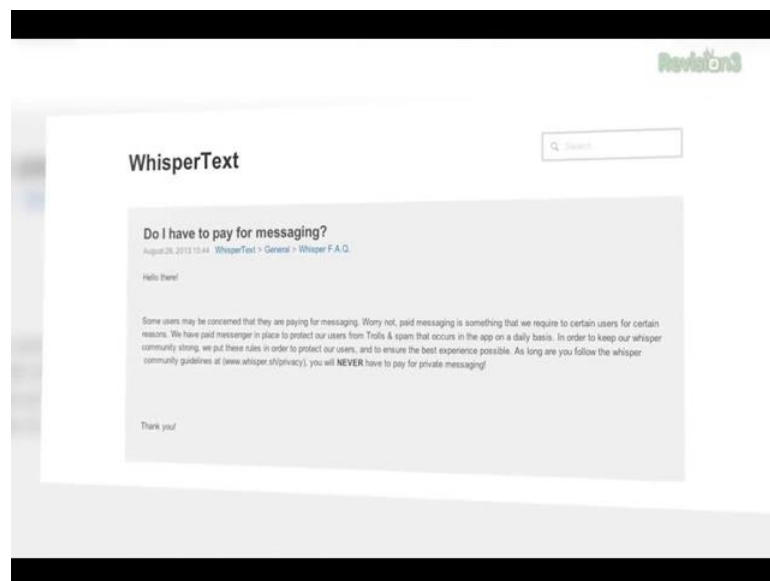


Secret,s lies and plenty of spam on this super popular mobile app - whisper. Whisper is a confessional app that encourages you to post your secrets behind a screen of **anonymity**. You know **kinda** like that other website - **Post Secret**, where in users can submit those deep dark secrets they would not even tell their best friend, like I secretly took nude

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, **taboo** or sometimes disturbing confession or sometimes none of 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.

(Refer Slide Time: 06:18)

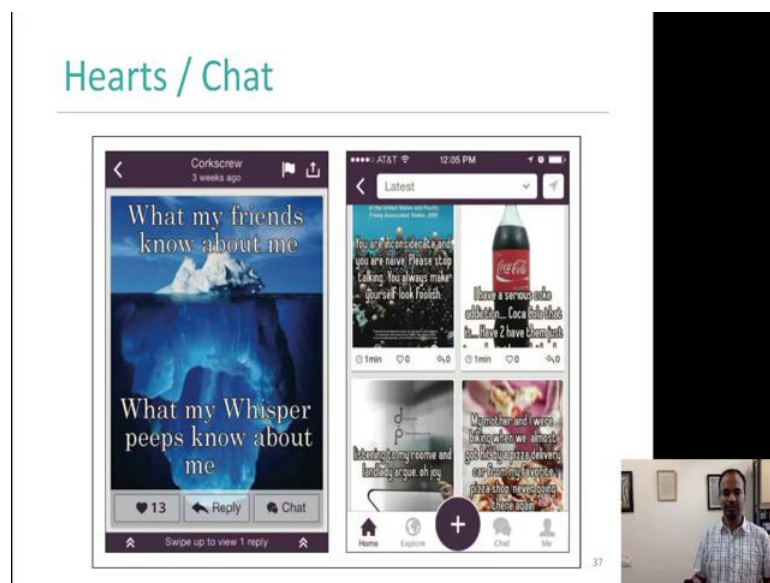


And you can private message, there is premium messaging for certain users, trolls and spammers according to the F.A.Q, but for everyone else it is free **as long as** you play by the **rules**. For that reason you have all seen posts like these, a lot of them people trying to hookup or these - message your favorite singer and **if you are dumb enough to fall for that, you are probably under the age of 10 and god help you**, but wherever there is secrets there are also bad apples and sometimes lies.

In September for example, someone posted a supposed murder confession on Post Secret, prompting a **frenzied reddit search for the self professed criminal**. And **till date** no

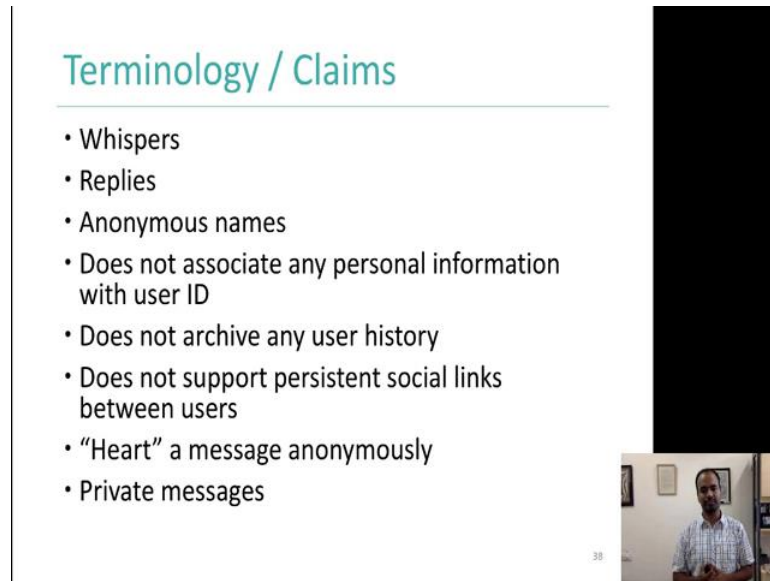
crime has ever been found associated with that post. And this week in Arizona, a cop was arrested for having sex with a minor he met on whisper app after she posted that she wanted to get pregnant. But considering the allure of posting secrets and knowing other people's, it's unlikely to slow down whisper for now. As always you can let me know what you think I am on Twitter, Facebook, Google or VK on anniegaus and you can get a free netflix trial with a signup on Netflix dot com slash wtbd, thanks for watching.

(Refer Slide Time: 07:16)



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 onto images and you 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.

(Refer Slide Time: 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

The slide features a title 'Terminology / Claims' in teal text. Below the title is a horizontal line. To the right of the list is a large black rectangular area. At the bottom right of the slide is a small video inset showing a man in a light-colored shirt speaking.

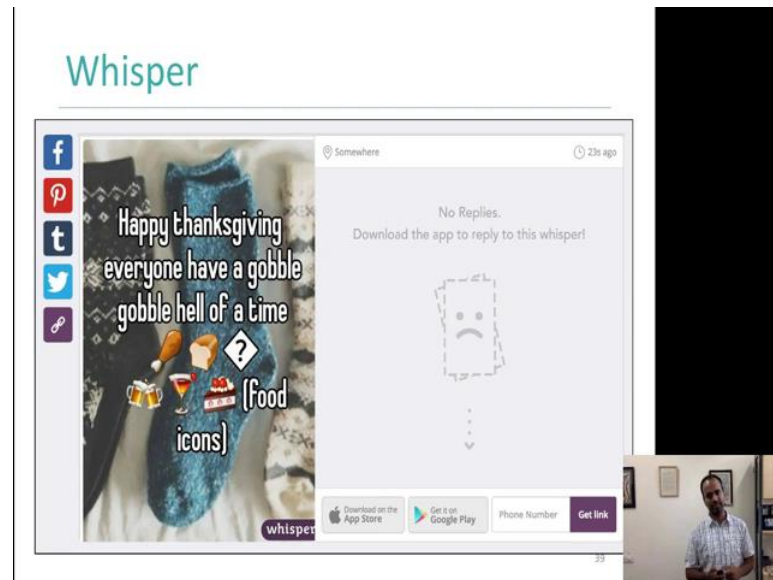
So, terminologies that we'll see to understand the rest of the lecture, we need to understand some terminologies, whispers or **the posts, replies** or I do a post and you are actually replying like a comment in Facebook or a reply in Twitter. And the posts are anonymous you really do not get to see 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 and look 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's what 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 the questions that you have seen in the past where in 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 (Refer Time: 10:02). A heart is basically the one that I showed you in the slide, like **the**

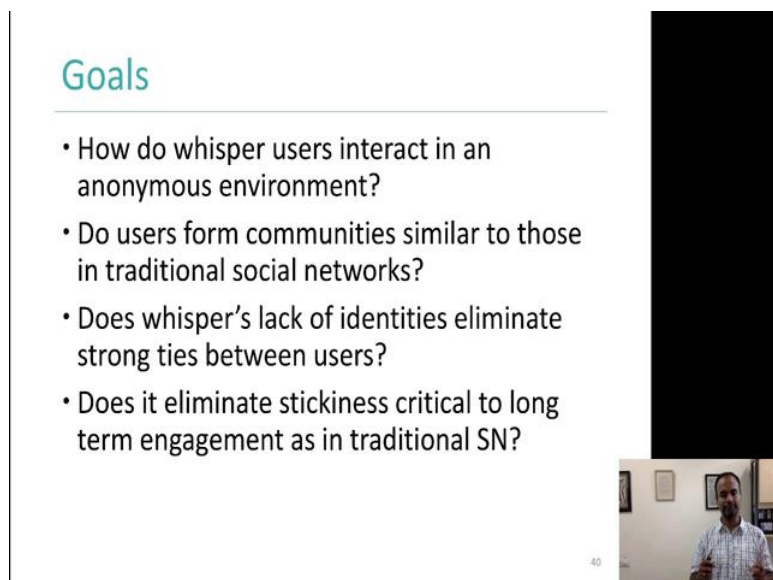
like in facebook. If in the private messages against or this in the video that I had a few minutes before, which showed private messages also you can actually post private messages between the users.

(Refer Slide Time: 10:28)



That is the screenshot from whisper.

(Refer Slide Time: 10:36)



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? Do users form communities similar to those in traditional networks, like for

example people interaction on facebook, how is this different from people interactions on 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 to you, does **it** eliminate the strong relationship that you and I would have. **Let's take** both of **are on** whisper, you do your post and I come react to it, I do a reply to it, if you do not know **it is** PK, who is the faculty at IIT Delhi or some profile that I have, if you do not know that it is **me**, will you continue talking to me? Is there is a stronger relationship that happens.

For example, you could also see in twitter or in facebook that some people are very strongly connected. For example, if I do any post that are some sets of people who would always like it, **who** would always accurately make a comment or reply or retweet. So, those that are basically called strong ties and that does it exists on **whisper** is the question 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 by which you are actually **glued on** 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 in this case whisper, we will actually try an answer for these four questions using some data, using some inferences that we draw.

(Refer Slide Time: 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

41

So, data that was collected for doing this analysis is from 2014. And of course, that whisper does not have an api so, 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 - the text that was on the picture, author's nickname which is the your handle, names so to say in the traditional sense, a location 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.

(Refer Slide Time: 14:26)



The slide is titled "Data collection" in teal text. It features a bulleted list of statistics and actions. A small video inset of a man in a checkered shirt is visible in the bottom right corner of the slide area.

- 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

42

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, the users get one unique id which is **what was** collected. 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 **that using** this id you could actually go back and find out which **user did what**. So, the user GUID concept was **removed** on June 2014.

(Refer Slide Time: 15:27)



The slide is titled "Data collection" in teal text. It features a bulleted list of dates, collection methods, and data types. A small video inset of a man in a checkered shirt is visible in the bottom right corner of the slide area.

- 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

41

(Refer Slide Time: 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

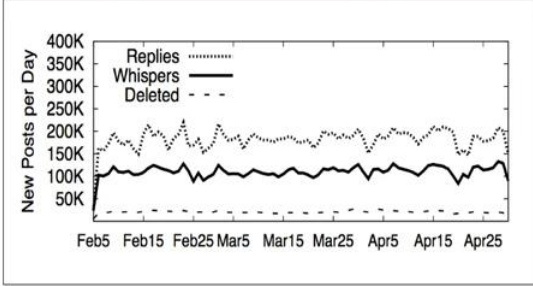


42


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 **wrote about**.

(Refer Slide Time: 15:40)

Data



- 55% of whispers receives no replies
- 25% have a chain of at least 2 replies



43

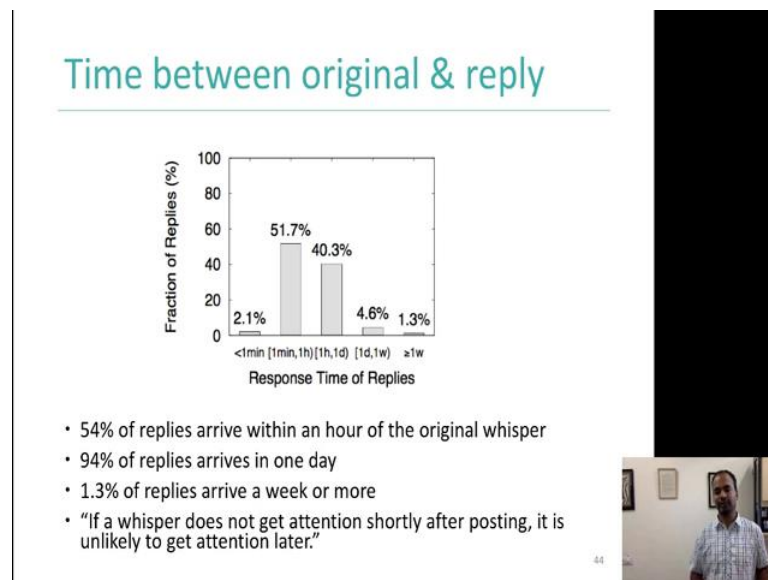
So, now look at analysis again, we **have** done this in the past **so I'm going to** go slowly in terms of what analysis, first time **using** this data, what kind of inferences to be true. And all connected to **these four questions that** we have. In the x axis such as the time, it is the

time, in this case between February and May, that is where they collected the data. And y 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 is how many replies are 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 whatever reasons 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. Only 25 percent so to say actually we should read it that way, the 25 percent have a chain of at least 2 replies. 55 percent of the replies, 55 percent of the whispers don't get a reply.

(Refer Slide Time: 17:27)

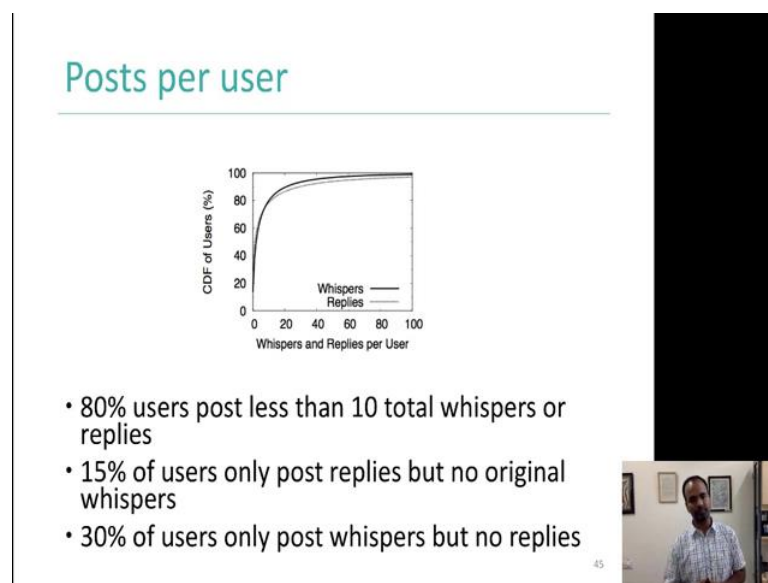


Time between original replies this is also interesting thing how quickly are the responses to whisper that is posted. How quickly do the people actually look at the post that is done and how they reply and what they reply. In this case, we are only looking at the time we are not looking at the content. So, if your x axis is time again less than one minute, one minute to one hour, one hour to one day, one day to one week and greater than 1 week.

That is the x axis, y axis is fraction of replies, fraction of replies it shows you what is 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, this will show you 54 percent, 54 percent of the replies arrive within one hour, 94 percent within one day. Basically, shows that if they do not get a response in 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 a whisper does not get attention shortly after posting, it is unlikely to 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.

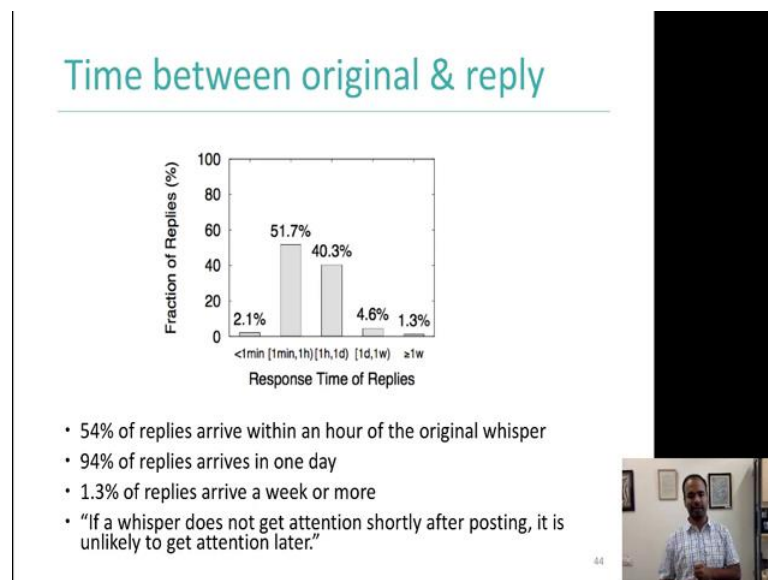
(Refer Slide Time: 07:16)



So, a post per user, which is just the x axis is whispers and replies per user. Whispers and 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, the y 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 are actually 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 the network, which we have already seen in other networks also. 15 percent of the users only 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.

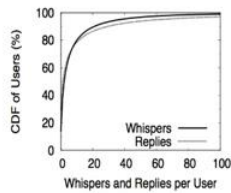
(Refer Slide Time: 21:32)



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.

(Refer Slide Time: 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

45

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.

(Refer Slide Time: 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”

46

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, the first column is graph whisper, facebook, twitter, second column is

number 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 pretty large.

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, they have 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 actually the lowest and there is average path length. Basically what does it mean it means 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.

(Refer Slide Time: 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**.

(Refer Slide Time: 27:27)



Content moderation

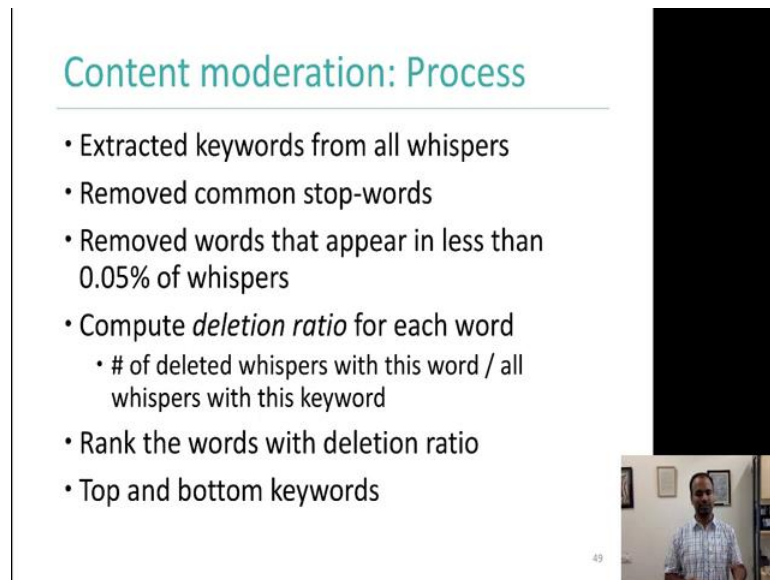
- 1.7 million whispers have been deleted in 3 months
- 18% of content deleted compared to 4% in Twitter

48

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.

And this **begs** the question which is that for **whisper** why is this percentage 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 it deleted**. So, there is higher proportion of content generated on whisper which is getting deleted.

(Refer Slide Time: 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.

(Refer Slide Time: 30:30)

Content moderation

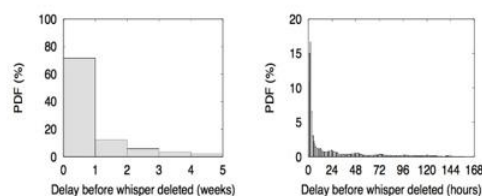
Topic	Top 50 Keywords Most Related to Deleted Whispers
Sexting (36)	sext, wood, naughty, kinky, sexting, bj, threesome, dirty, role, fwb, panties, vibrator, bi, inches, lesbians, hookup, hairy, nipples, freaky, boobs, fantasy, fantasies, dare, trade, oral, takers, sugar, strings, experiment, curious, 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.

(Refer Slide Time: 31:34)

Deletion delay



- 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

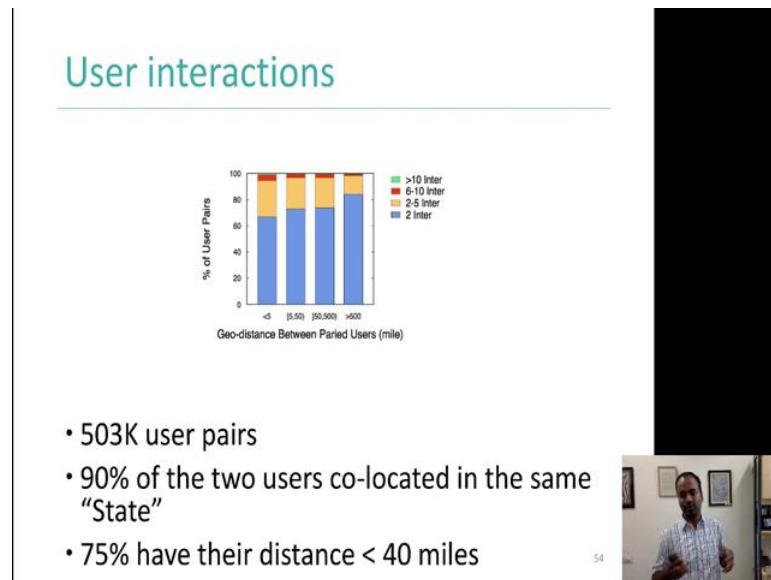
This graph is actually showing you, I mean let us look at the way in which the deletion is happening, how much time and relationship for actually deletion. 70 percent of the deleted whispers are deleted within one week after posting. So, that is the first graph from the left, which is 70 percent of the 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.

(Refer Slide Time: 32:44)



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 you **zoom in** to the data for this one week, majority **of them are** actually getting deleted within the first 24 hours.

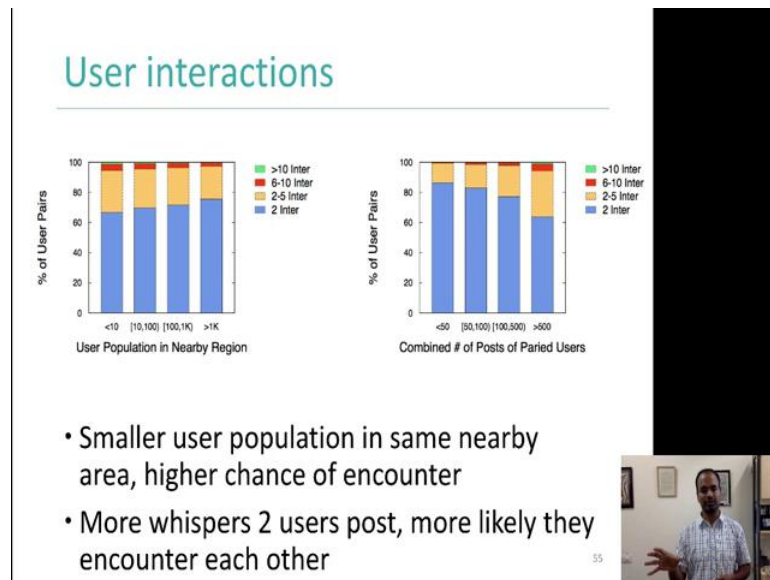
(Refer Slide Time: 33:40)



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 **paired users**, 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.

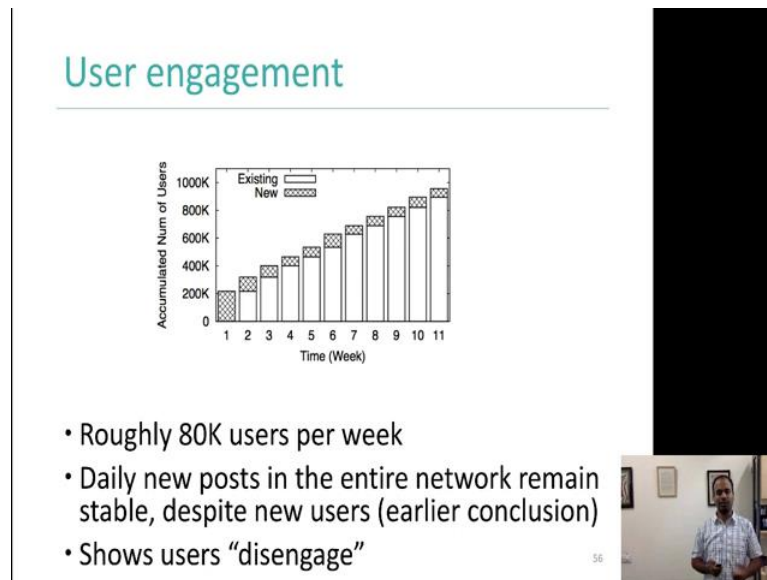
The colors are blues **is two** interactions, yellows **are two to five** interactions, **red** is 6 to 10 interactions, anything that was **above 10 interactions**, which was actually given **green**. You could already guess, that the number of interactions which is higher, is actually very low in the whisper, which again using the network analysis, using the things that we have already seen, you could actually make the inference. That is why green is very low on the 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.

(Refer Slide Time: 35:21)



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's actually 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.

(Refer Slide Time: 36:37)



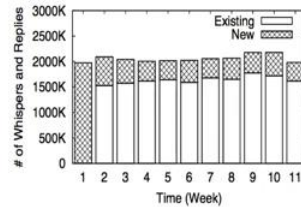
Now, look at how users **engage** in these in this network on whisper. So, here x axis is the time of week, y axis is the accumulated number of users and the two data points that are drawn in this graph is **the** existing and the new content that is getting generated. Roughly 80,000 user per week are interacting daily, new posts in the entire network remain stable that we will actually see in the next graph also. How this renames same is actually, if you 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 like the new content that is getting generated is actually increasing. That is the interesting 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, this basically shows that even though the **users** are increasing, which means the posts 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 to it.

(Refer Slide Time: 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

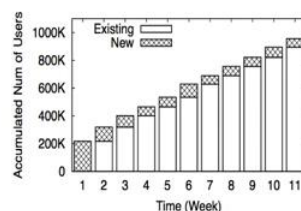
57



User engagement, so, here is the graph that I will actually show you. Number of whispers 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.

(Refer Slide Time: 39:17)

User engagement



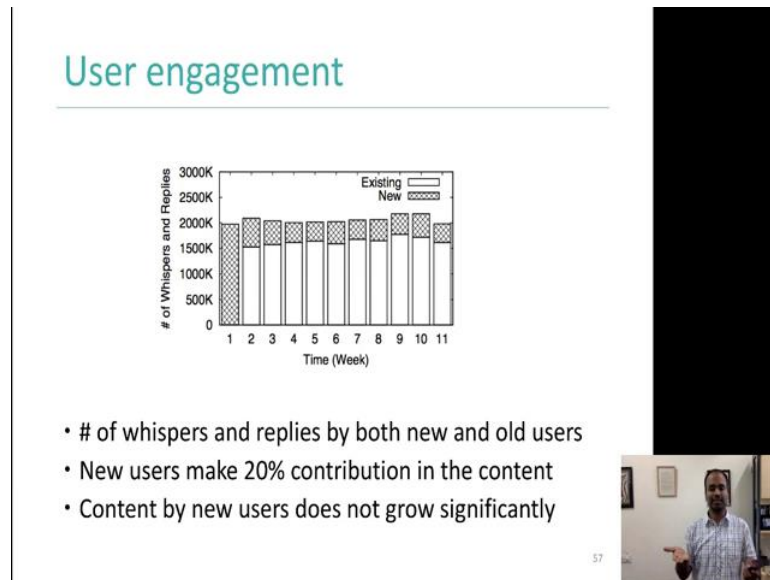
- Roughly 80K users per week
- Daily new posts in the entire network remain stable, despite new users (earlier conclusion)
- Shows users “disengage”

56



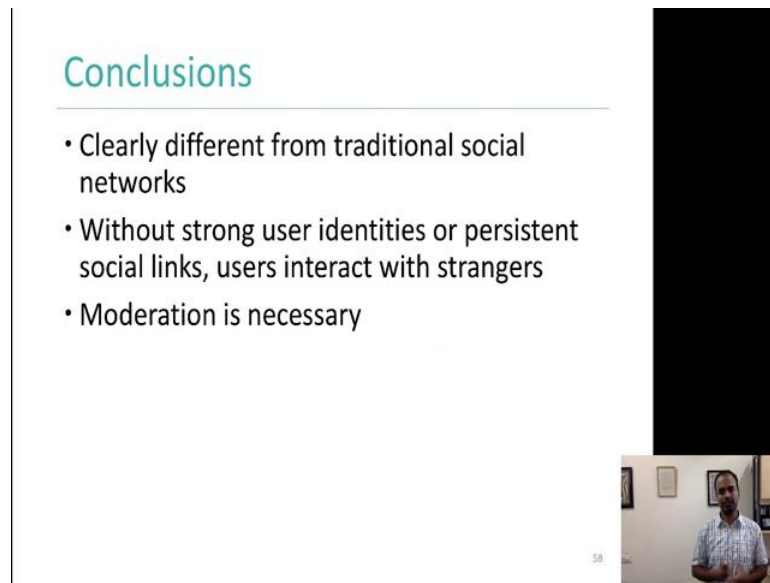
The earlier graph was the cumulative number of users right, accumulated number of users. So, the existing user which is the light without the check that the graph the bar is actually, which is risen which is showing you that the number of users are actually risen.

(Refer Slide Time: 39:38)



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.

(Refer Slide Time: 40:06)



Conclusions

- Clearly different from traditional social networks
- Without strong user identities or persistent social links, users interact with strangers
- Moderation is necessary

58

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, which are networks where you can post content where you can maintain a high 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.