

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

Week – 5.3
Lecture – 18
Policing and Online Social Media

Welcome back. Let us continue on the topic of how many Social Media and the Policing. 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.

(Refer Slide Time: 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

 Citizen to Citizen	 Police to Citizen
 Police to Police	 Citizen to Police

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 to Citizen, Police to Police and Citizens to Police, in the frame work of 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 cognitive response process; in simple term it is basically looking at finding out how

positive or negative the interactions are, how frequently the engagement happens, how is the response between the parties involved in the interaction.

(Refer Slide Time: 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?

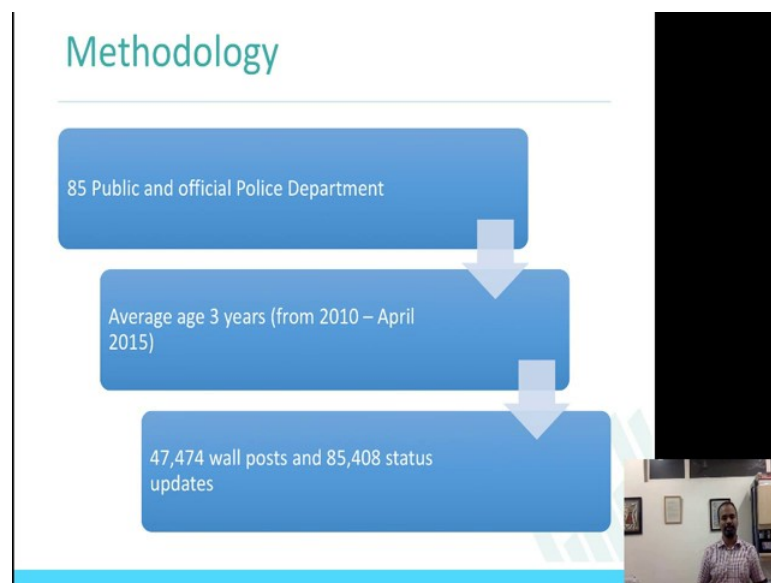
Since, specific questions that **were** look at are; Topical Characteristics, Engagement Characteristics, Emotional Exchanges, and Cognitive and Social **Orientation**. Let me walk you through little bit about what these are, and then show you some data and try to find **some** results with the data 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 are the four the different aspects **that** we will look at in terms of analyzing the data 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 looked at the content how the police and citizen, only Bangalore City Police interactions **were** going on.

(Refer Slide Time: 03:22)

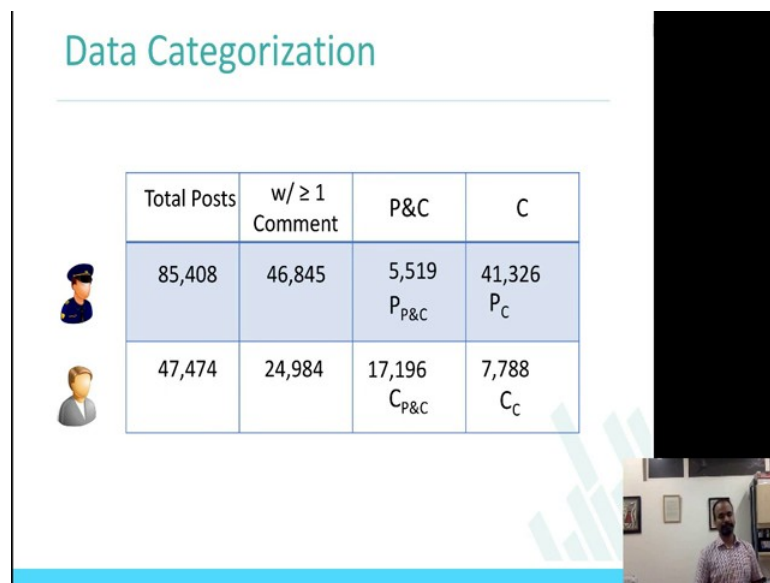


If you remember in the second part of this week I actually showed you a page where we were collecting, **where they were** pages of different Police Organization **that was** listed. **It's** the same list that we took which is 85 Police Organizations list, of there are 85 public and official police 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 average age of 3 years between 2010 and April 2015, where the data collected and there



were 47,474 wall posts and status updates.

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.

(Refer Slide Time: 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, total posts that were done by a police. The icons that will consistently go through this 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 citizens **were** 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 P and C is when the police post, the way to read this table is row two which is the police post and the column number 3 which is P and C means that police is posting

the content and the police and citizens both are interacting which is P, P and C. And the C 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 way to read this is citizens are posting and then police and citizens are commenting on it which is about 17,000 and only citizens commenting is about 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.

(Refer Slide Time: 06:32)



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.

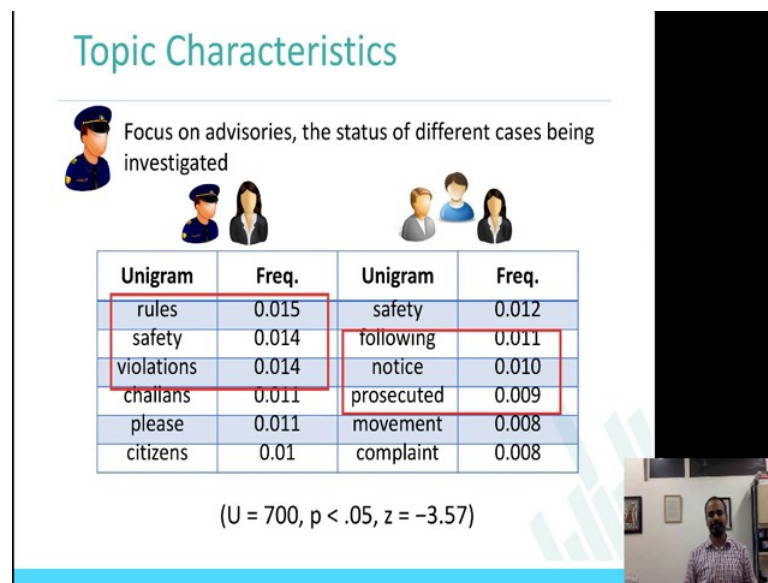
K-means Clusters, clusters are basically way in when we see users together interacting on the same post more number of people, there is a cluster there are. We can actually look in at how much of clusters happen. Clusters for example, the students who are taking this class from anywhere in the country 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 have looked at 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 of saying 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 the course also we will get here introduce in the tutorial, we will get you introduced to NLTK, 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 assume or 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 happy words 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 you are 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.

(Refer Slide Time: 09:27)

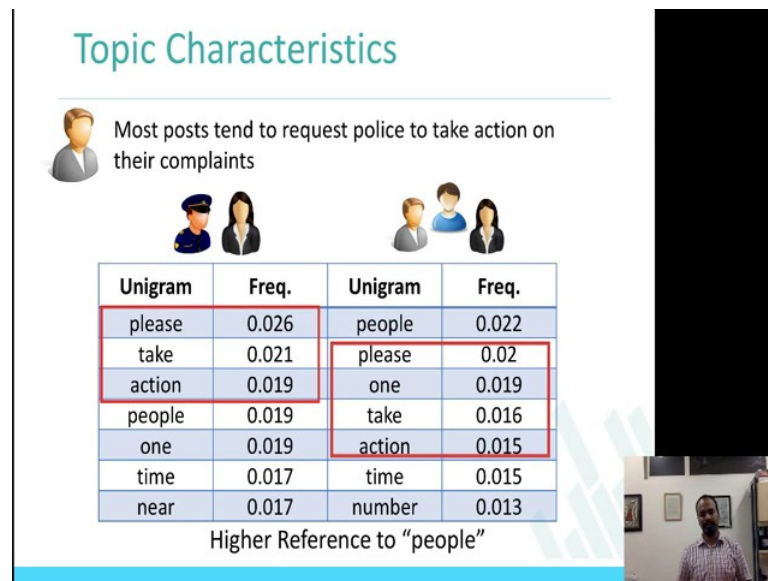


First let us look at the topic characteristics. Four things we said, the first one is the topic that were being discussed. So focus on, so when police post again if you look at icon on the top it is police, when the police post are focusing on advisories and the status of different cases being investigated 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 about **these 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. This just gives you a sense and the highlighted is basically **to tell you** how the differences are, 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.

(Refer Slide Time: 10:50)



Now if you flip the question and see, when citizens do the post what happens. Most posts **request** the police to take action, which is why probably citizens are using these pages, which is citizens post on the page for asking some action. If you look at only the unigram analysis, we see that please take action, **very evidently**, please take the action seems to be 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.

(Refer Slide Time: 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.

(Refer Slide Time: 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

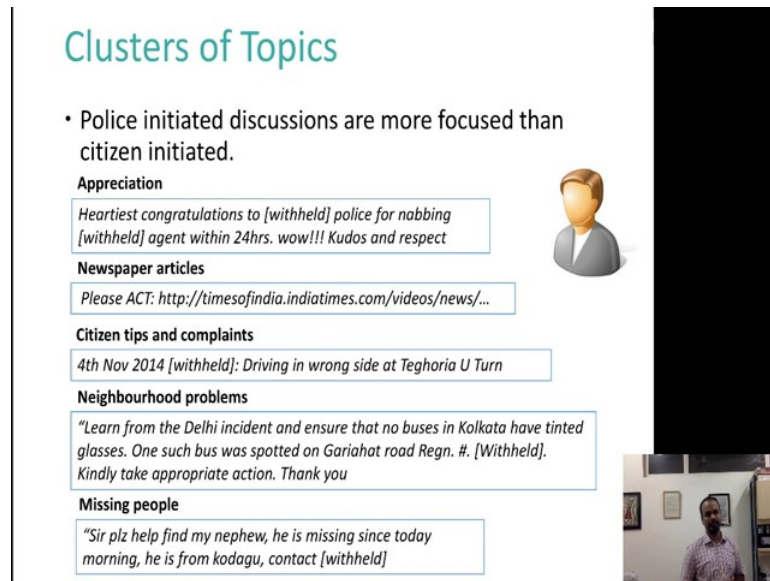
4th Nov 2014 [withheld]: Driving in wrong side at Teghoria U Turn

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.

(Refer Slide Time: 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** looking what the interaction says, can be actually **learn** something specific. Engagement characteristics; how do the citizens and police engage in social media discussion threads?

(Refer Slide Time: 15:15).

Engagement / Comments Characteristics

• Content Generators

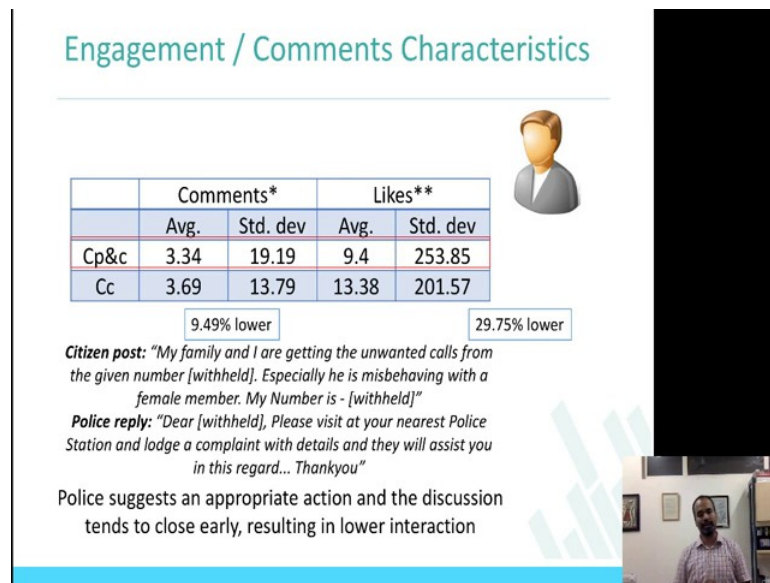
	Police		Citizen		
					
Police + Citizens	55,028	1,79,176	17,124	12,630	26% lower
Citizens Only	54,982	1,79,176	17,081	12,630	

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. If you look at the first set of columns which is the second and the third column, this is basically showing that when police does the post what happens to 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 comments on the Facebook pages. If the post are coming from citizens and the comments 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.

(Refer Slide Time: 16:46)

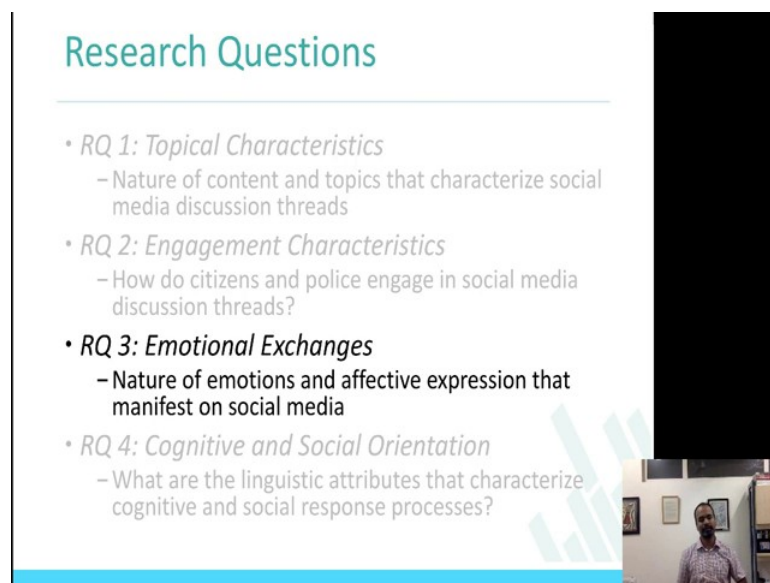


This slide is just an inside view of the data of the last slide. Last slide was this. This is total number whereas this is average number. This is average number of comments for the posts that we saw in the last slide, which is if the police and citizens are interacting there the average 3.3 comments are there for post, whereas likes is about 9 likes, only citizens is about 3.69 and about 13.38 is on the average. So, the total gives you a picture, whereas when you look at an average it is actually the picture is slightly different which is C P C 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. Police replied: dear please visit at your nearest police station and 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 police starts 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 citizens are there is actually lower than only a citizens being there. A simple reason could be police can actually quickly close in the loop of the discussion.

(Refer Slide Time: 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?

The next one we look at is emotional exchange, which is nature of emotions and affective expression that are being discussed on social media.


(Refer Slide Time: 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 higher negative sentiment. Here is a table which actually connects to the topics that I said earlier which is negative effect, anxiety, anger and arousal. In the same **one** you can see that negative affect is higher when police is actually involved in the interaction, which is 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.

(Refer Slide Time: 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

*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.*

200% higher in Cc



If you look at the next one where the anxiety is actually lower when police starts interacting in these **posts** which are initiated by the citizens, you can see the value to be 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 of the citizens to go **low**.

(Refer Slide Time: 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

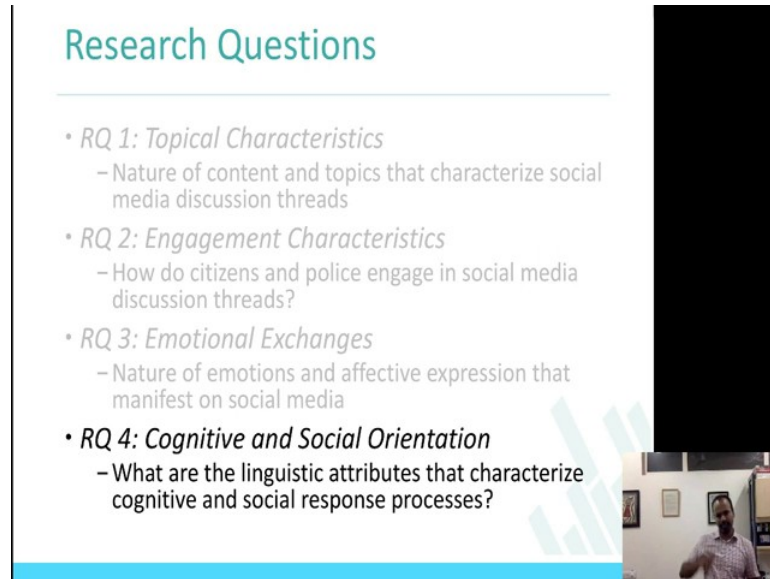
12.82% higher in Cpc

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 and they are approaching the police to actually get things done as quickly as possible. And arousal is of course higher in C P C. **Again** when the citizens initiated threads, the C P C is higher, the arousal because the **intensive** discussion is higher because you want actually get things done from police again.

(Refer Slide Time: 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 that we started with. We started the first question to be topical characteristics, second one to be the engagement characteristics, third one to be emotional exchanges, and the fourth 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.

(Refer Slide Time: 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
shehe	0.002	0.01	0.003	0.003
they	0.005	0.013	0.008	0.008

75% More

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 ?



The one approach that we could it was looking at the proper nouns i's and he's and they'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.

(Refer Slide Time: 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.

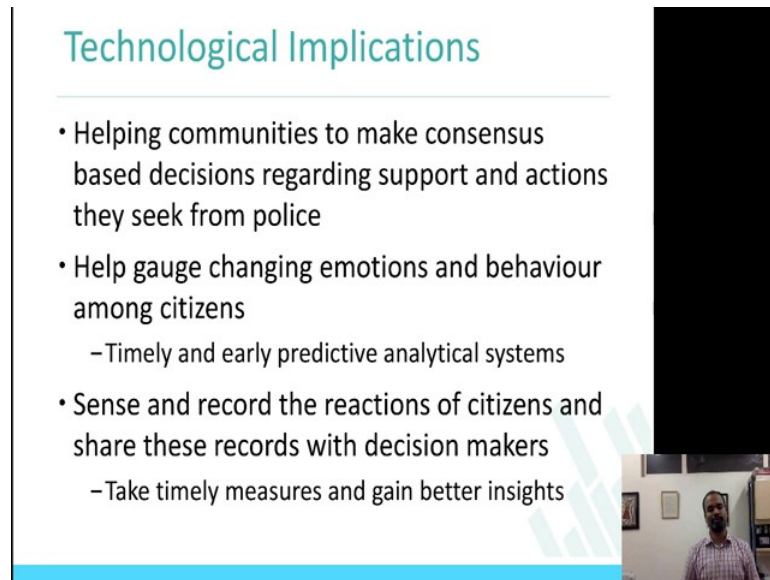
Now let us look at why all this matters. For the entire week for this course, this week I kind of took you around the data that are created on Facebook, that are created on social media for Police Organizations doing some analysis, seeing what kind of interactions **that are** going on between police 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, social support. Which is what we are trying to capture in 5.3. 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 actually lower 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 comes and says, **Oh good**, thank you for letting us know about it and **we will** actually do

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.

(Refer Slide Time: 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. They can actually build 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.

It can also help sense and the record the reactions of citizens and share these records with decision makers. It can actually get, meaning some organizations also can do this.

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 there **we went on** to ask some research questions are on that. And these kind of research is not only done in terms of just **Indian webpages, Indian** Police Organizations, this is done all across the world.

Thank you.