ivacy and Security in Online Social Networks
Prof. Ponnurangam Kumaraguru ("PK")
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Week - 9.1
Lecture - 29
Privacy in Location Based Social Networks Part 1

Welcome back to Privacy and Security in Online Social Media course on NPTEL. This is 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 they write a paper what all things fits into the paper, what all analysis that they have done essentially, it is looking at the content that you have seen in the past, but in terms of the structure of the paper itself we will go through some of them.

(Refer Slide Time: 01:22)

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 Airport in Delhi traveling to XYZ place. So, that is location based service, you open your phone, you check into this location or post this status update with the location updated in it. Foursquare, Yelp, these are very, very popular services which actually do location based services, these are called location based social networks.

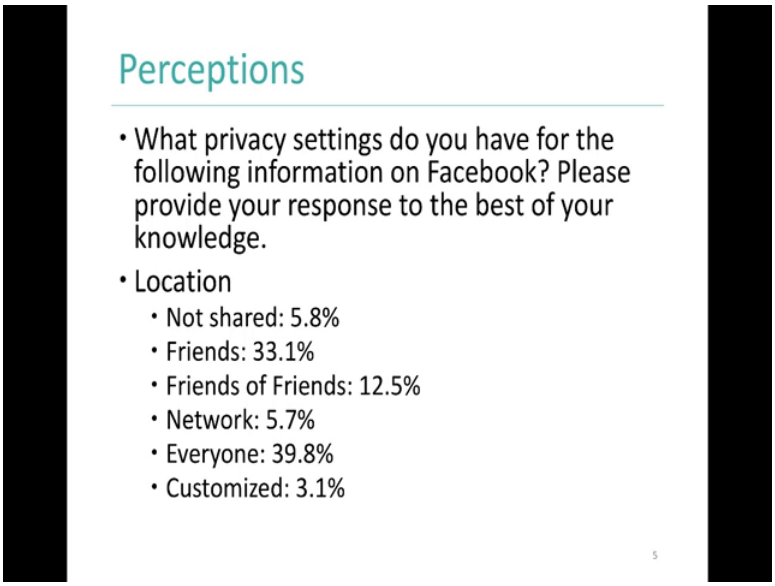
(Refer Slide Time: 03:18)

So, of course, all of them have some sort of privacy concerns too. I am, I am sure by now you already realized the privacy concerns in location based services which are, where my

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.

(Refer Slide Time: 04:08)



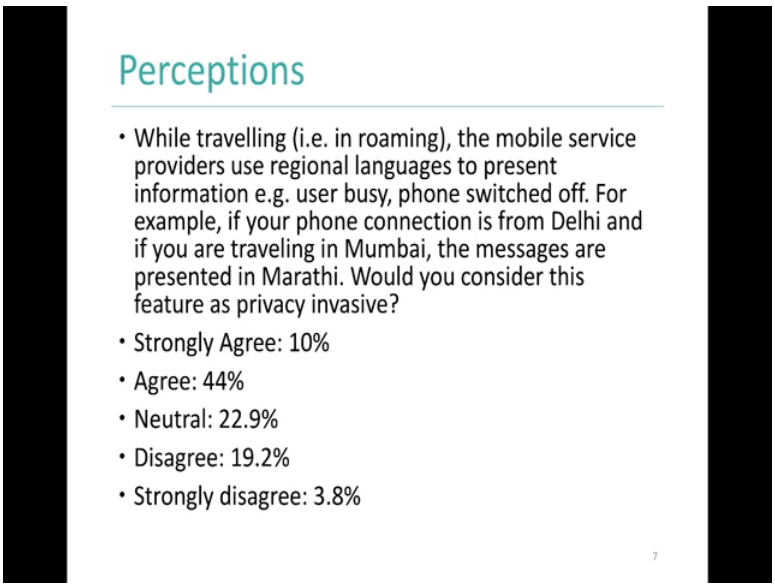
So, just to give you a sense of what are the perceptions in terms of actually these location based services. If you remember, there was a study that I referred earlier in the lectures also, where 10427 people were asked some questions about different aspects of privacy. One section in that was online 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 different options that was given for the question. I do not know where each of you will fit 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.

(Refer Slide Time: 06:26)



Perceptions, again. Another question in the same survey, same data collection, which was 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.

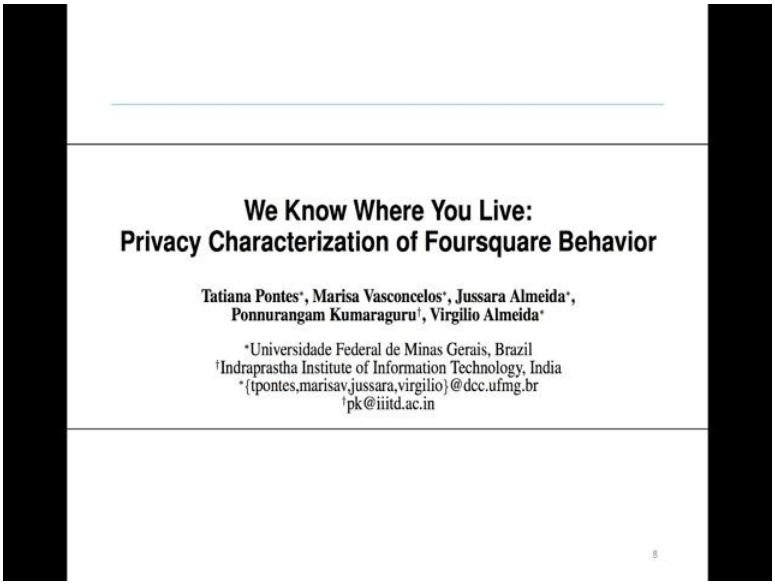
For example if your phone connection is from Delhi and if you are traveling in Mumbai the messages are presented in Marathi. Would you consider this feature as privacy invasive, I am sure you got the question, the question is simply that I have a number which I bought it in Delhi and I am traveling in Mumbai and my phone is switched off.

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 message is going to be in Malayalam, which is basically revealing the information 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 percent of them say, that it is privacy invasive, 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 all of this data was collected only in India. The perceptions here we are talking about person living in India who is thinking about these problems.

(Refer Slide Time: 08:34)

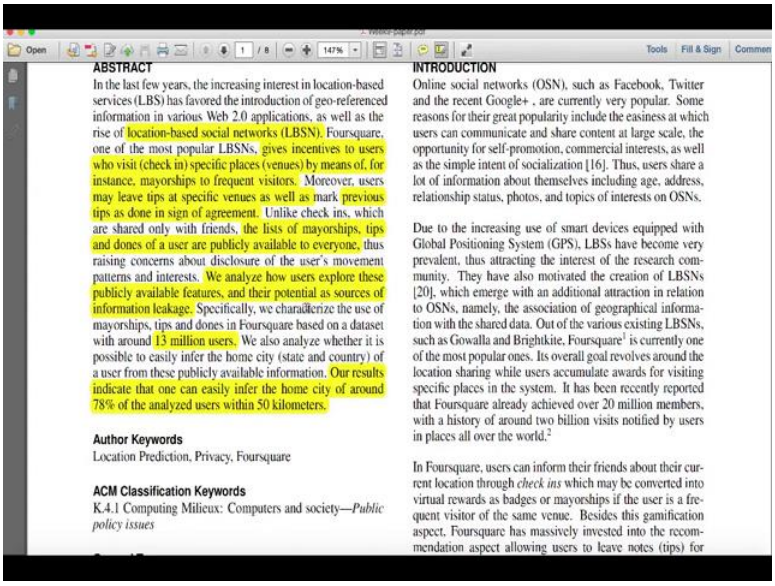


So, that is basically a sense of what location based privacy is. Location based privacy services are what kind of issues you can have, but what we will do is as I said we will actually look at some specific research that have done in terms of information. In terms

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 actually making some inferences. So, now, here is a paper that we will look at and spend some time on the paper. The title of the paper is 'We Know Where You Live: Privacy Characterization of Foursquare behavior'. We Know Where You Live: Privacy Characterization of Foursquare behavior

(Refer Slide Time: 09:52)



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, IIIT Delhi is a venue, you come to IIIT Delhi, which is a venue, and then you do a check-in, which is you are visiting the

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, Saravana Bhavan, 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, analyzing them and making some interesting inferences, particularly looking 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 talk about ,our results indicate that, one easily, one can easily infer the home city of

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 website called please rob me dot com (pleaserobme.com), which they took it down after some time. This website did something interesting - which is p l e a s e r o b dot robbed and me 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.

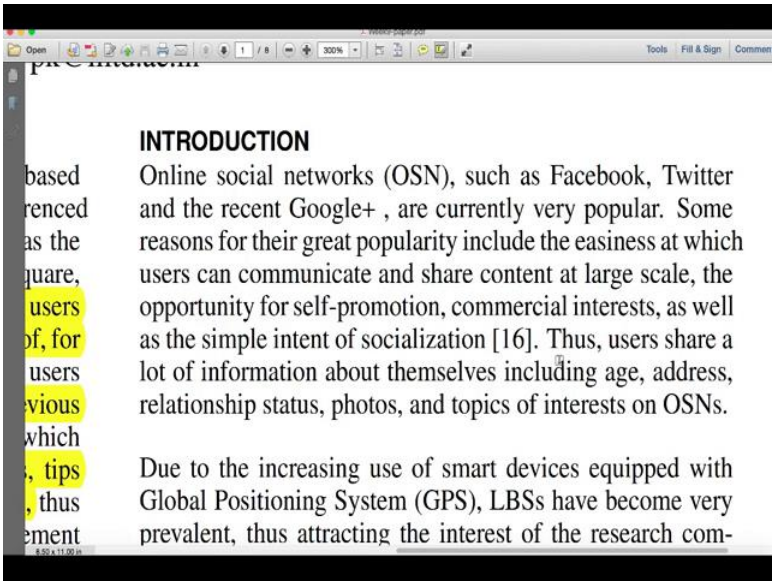
Or if you have checked in, if you have done a post in tweet with your geolocation on from Delhi and in another hour or so, you are talking about actually weather and couple of hours you are talking about weather in Chennai again, that is a probability that you are not at home, you moved from Delhi to Chennai. So, they were using this information and 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 for burglars to go and rob the 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 the benefit of time constraint and the lectures also I am going to look at only the ones

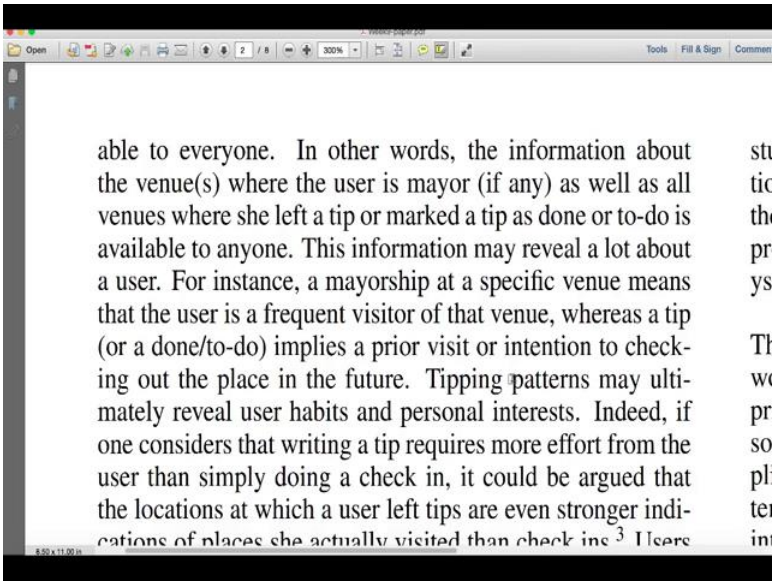
that I have highlighted here.

(Refer Slide Time: 17:54)



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.

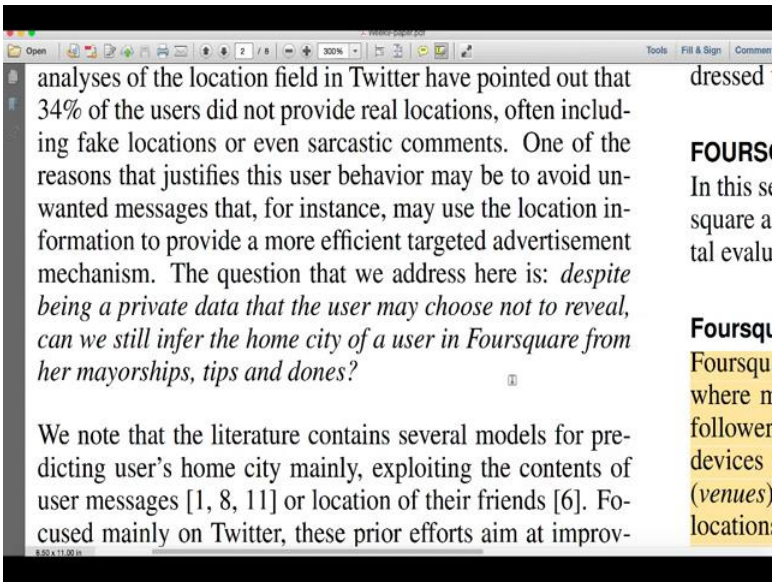
(Refer Slide Time: 18:20)



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

And then you quickly mention about the methodology that you did.

(Refer Slide Time: 18:45)



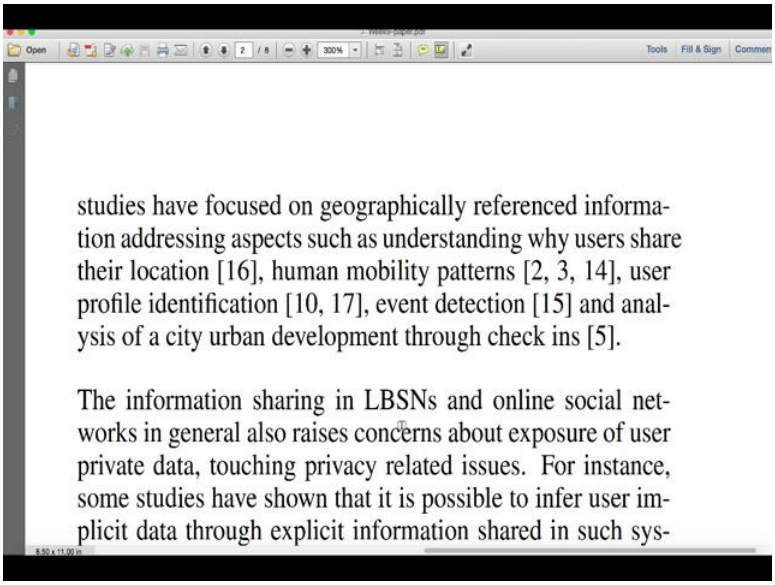
analyses of the location field in Twitter have pointed out that 34% of the users did not provide real locations, often including fake locations or even sarcastic comments. One of the reasons that justifies this user behavior may be to avoid unwanted messages that, for instance, may use the location information 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 generally motivate from 30,000 feet height about the problem saying, this is the problem, 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.

(Refer Slide Time: 19:42)

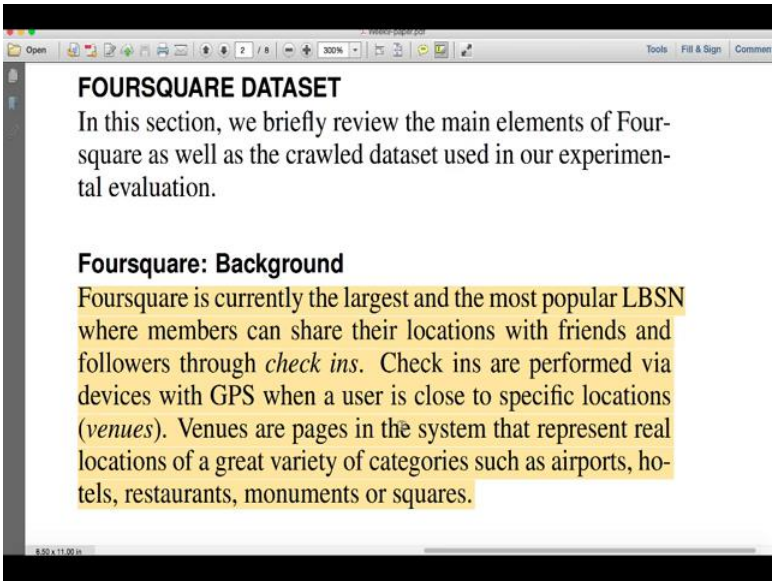


So, in this case, authors are talking about how people have used human mobility patterns, even direction analysis of a city urban development, through check-ins. There is a lot of work actually in terms of using Foursquare data to find out the nature of the city and how people have actually used the 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.

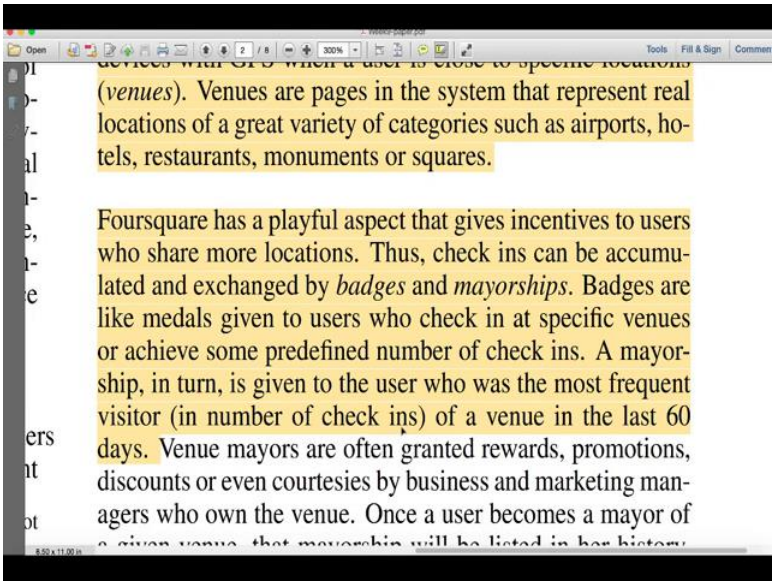
Now, we look at something more closely in terms of just Foursquare Dataset itself.

(Refer Slide Time: 21:13)



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, venues can be airport, restaurant and monuments, you come and check in that particular location.

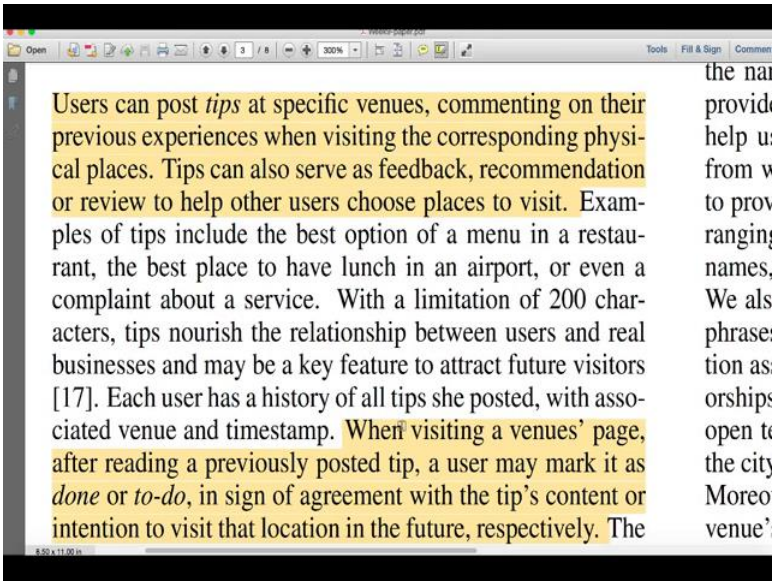
(Refer Slide Time: 21:55)



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 users who check in at a specific venues or achieve some predefined number of check-ins.

(Refer Slide Time: 23:07)



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 to visit. 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 that is 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 essentially saying that I saw this tip the tip is actually very 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.