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

Week – 8.1
Lecture – 26
**Profile Linking on Online Social Media**

(Refer Slide Time: 00:19)

Welcome to the course of Privacy and Security in Online Social Media. This is week 8, the first part of week 8. So, just look at the profiles on the screen, it has first one Facebook handle called ponnurangam dot kumaraguru (ponnurangam.kumaraguru), the second one which is Twitter profile called ponguru, and the third one which is LinkedIn, which is ponguru again.

So, the question is, can you actually match all these 3 URLs or all these 3 profiles and say that it is the same profile. That is the question that, we are going to try and answer in this, this part of, this week of the course. Which is I have handles of ponnurangam.kumaraguru from Facebook, ponguru from Twitter, and ponguru from LinkedIn. Can I actually use this? What do I need to do to make sure that these 3 profiles are same or to understand that whether these 3 profiles are same. There are multiple actually test cases, scenarios for it; I will actually discuss a little bit later in the lecture.

(Refer Slide Time: 01:36)
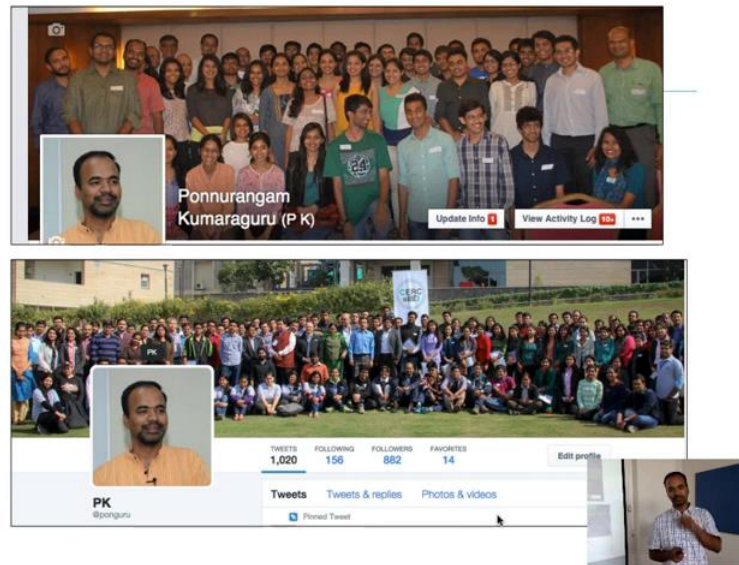
Here is, the top one is my Facebook profile, the one at the bottom is my Twitter profile the one.

(Refer Slide Time: 01:44)

The next slide, this is my LinkedIn profile, public LinkedIn profile. So, if you look at these three images, you can actually see or you can actually think about some features

that you can use for deciding whether these 3 profiles are mine. For example, you can look at my profile picture in both, they seem to be the same thing, you can look at probably some friends that I have on Facebook and people who are following me or then the accounts that I am following on Twitter, you can look at some of these features to make the decision. Unfortunately, in the public profile that I have on LinkedIn, there is no profile picture, but there are details like associate professor at IIIT Delhi, Data Security Council of India, Carnegie Mellon University and connections like that.

For example, my personal website, the personal website from here may be actually linked to my website at IIIT Delhi. So, you can actually make all these connections to find out whether this is actually the same details is both in Facebook and the Twitter. I am sure many of you are listening to this lecture also have multiple accounts. So, the question that you can ask yourself is, how do I put, how do you put your own accounts together to find out whether they are same or not. So, that is the problem that we look at.

(Refer Slide Time: 03:21)

This lecture

- Tracking social footprint / identities across different social network

So, tracking social footprint identities across different social networks, which is finding out whether they are the same.

(Refer Slide Time: 03:33)

Other Times, Other Values:
Leveraging Attribute History to Link User Profiles across
Online Social Networks

Paridhi Jain
Indraprastha Institute of
Information Technology (IIIT-D),
India
paridhij@iiitd.ac.in

Ponnurangam Kumaraguru
Indraprastha Institute of
Information Technology (IIIT-D),
India
pk@iiitd.ac.in

Anupam Joshi
University of Maryland, Baltimore
County (UMBC),
USA
joshi@cs.umbc.edu

ABSTRACT

Profile linking is the ability to connect profiles of a user on different social networks. Linked profiles can help companies like Disney to build psychographics of potential customers and segment them for targeted marketing in a cost-effective way. Existing methods link profiles by observing high similarity between most recent (current) values of the attributes like name and username. However, for a section of users observed to evolve their attributes over time and choose dissimilar values across their profiles, these current values have low similarity. Existing methods then falsely conclude

mation, lists her friends and later creates content to share with her friends. The quality, quantity and veracity of the information created and shared by her vary with the OSN, thereby resulting in dissimilar profiles of the same user, scattered on the world wide web, with no explicit links directing to one another. These disparate profiles liberate her from any privacy concerns that could emerge if the profiles were implicitly collated. However, linking these disparate unlinked profiles can benefit various stakeholders.

Companies like Disney and PepsiCo carry out psychographic segmentation based upon customers' activities, interests, opinions

Jain, P., Kumaraguru, P., and Joshi, A. Other Times, Other Values:
Leveraging Attribute History to Link User Profiles across Online Social Networks

I have said many of these topics that I am discussing in the class it is all connected to some research done. So, here is a paper that I am going to be talking in detail today, which is 'Other times, Other values: Leveraging Attribute History to Link User Profiles across Online Social Networks'.

(Refer Slide Time: 04:02)

Knowing this can be useful!
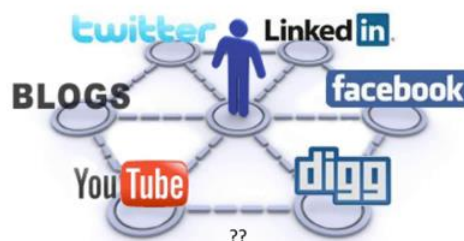
So, big advantage of actually <mark>knowing</mark> these connections, whether they are same is actually very, very useful.

Let us look at this slide, please look at this slide, which says about duplicating audience. Where if I were, so, in this case there are 437,000 likes on a Facebook page and about 153,000 followers that the account has and about 800,000 followers that the handle has on LinkedIn. So, the question is, if I were to send an advertisement, if I where to actually send some information to these users, will it be same, <mark>will it</mark> be a sum of all of them or will it be something smaller, because that could be some of these 470,000 profiles, the same users are actually 153,000 in Twitter and the same users were actually on LinkedIn.

For example, I am sure some of you in listening to this lecture will have accounts on Facebook, Twitter and LinkedIn. If you have account on all the 3 and if PK wants to send you about information on PSOSM on NPTEL, it is actually useless to send the information to the same handle, which is, let's take, Sonu Gupta in Facebook, Sonu Gupta in Twitter and sonu dot gupta 24 in LinkedIn, because it is the same person. We're actually wasting our resources in sending this information about (Refer Time: 05:47) PSOSM on NPTEL to the same person in all 3 accounts.

So, that is the problem to actually look at. So, the question is, people have multiple accounts on social media and sending information to all of them, you want to send information to the people only once. So, that is the goal then, but there are many test cases for this problem. At the end of this lecture I actually talk about some other test cases in <mark>law enforcement</mark> (Refer Time: 06:17) agencies and in other situations.

(Refer Slide Time: 06:22)
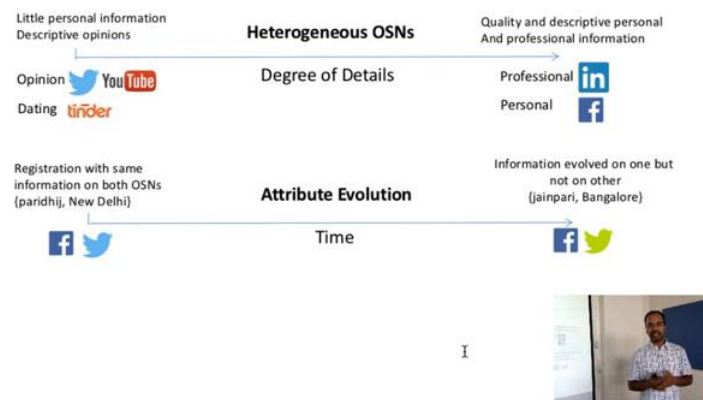
A technical challenge for actually putting them together is also harder <mark>because</mark> if you look at some networks, you get actually details, which are something more personal, in some networks, they are not actually that personal.

For example, in the in the top part, I am showing you here about YouTube, being a video sharing service, you can <mark>get</mark> actually opinions, you can get what they like, what kind of videos they actually saw, in Tinder, which is the dating side, little bit of <mark>personal</mark> information is available, connecting in to LinkedIn, which is professional and Facebook, which is also personal details, right. So, the question is, what information can you actually collect from these different social networks, which have different types of information, how do you put them together and create answer the question that we <mark>started</mark> off with, finding out whether multiple handles are same or different, right. I hope that is clear.

So, the question about profile linking, what are the approaches that we can take? The approaches that we can take is list out common attributes, which is Facebook has my gender, my age, my university that I work at, places that I got my degrees from. Twitter has my followers, my profile again, the website that I am connected to, the place that I work, all that information. We can actually list on all common attributes, compare the attributes, which I think in the example that I showed you, I showed you profile picture being same, profile picture being same on Twitter and on Facebook, we can actually compare that.

Compare attribute values using syntactic, semantic or graph based, which is what I am typing in, on, the social networks, what content are am I posting and what will, what is the details in my profiles and the graph is basically my networks - my friends in Facebook, my followers, followings in Twitter..

And then high similarity, if there is, in my case in the example that I showed you it is the exact the same picture profile picture on both the places. If things are like that, it mostly likely the same person. And then the question is also, you can, so, one thing that I will talk about few slides later is not just that you want to look at these details only that is now,

but you can actually look at details that are past also, which is, you do not have to look at only the post that I did now or the profile picture that I had now or the handle that I have now. You can actually go back in time and look at the post that I have done and you can derive some information even from that. For example, one thing I will show you also is people actually change their user handles sometimes. So, can you actually use that information to actually derive whether it is the same profile?

(Refer Slide Time: 09:47)

But the values change!
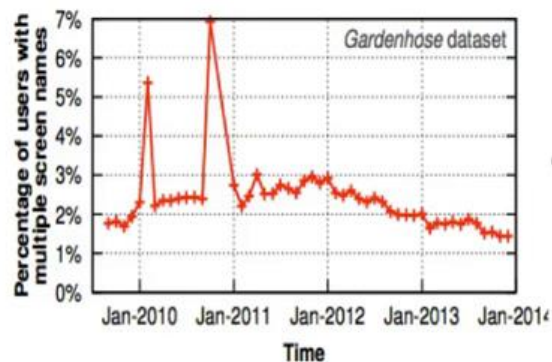
Attribute: Username

# of users tracked: 376 million [random]

Tracking period: 4 years

Now, if you look at this graph, this graph is actually showing you the changes that has happened in terms of just username, the point which I just now said, which is, some details of the profiles can actually change over time. It is not that you have to look at the details that are now, but you can actually look at the past - that is the problem, that is the question that we are trying to answer there. So, here 376 million users where tracked and the graph is showing you x axis to be the time and y axis to be the percentage of users with multiple screen names, which is names that they have changed.

For example, in my case, currently I have ponguru, whereas in the past, let's take if I would have had ponnuguru123 or professor@iiitdelhi, all those things are actually

getting captured here; which means some 7 percent of the people had different, change their usernames sometimes, in the data that we collected in around January 2011. That is the way to infer. And then there is another peak around January 2000 or February 2010, the two peaks in this graph are basically showing you, that, what the first peak is showing you, about between 5 and 6 percent of users that we were tracking, the handles where changed, and about 7 percent of the account user handles where changed around January 2011 or December 2010.

So, this basically shows you that people change accounts, people change their handles. (Refer Time: 11:39) Again for people listening to this lecture, think about yourself, how many people have actually changed the handles that you use. In Facebook I think you can change it with only once, but in Twitter you can change it as many number of times you want, which means it is actually possible to keep changing your account every now and then.

(Refer Slide Time: 12:04)
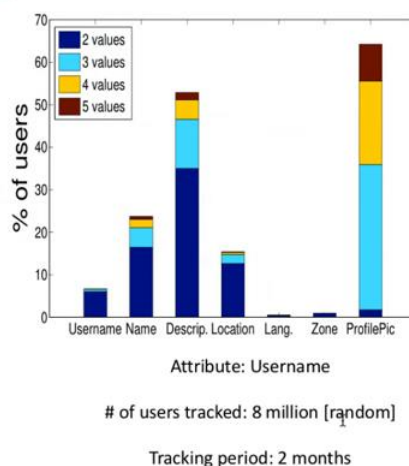
So, continuing on the same thread, which is about the details changing for the users. So, in this case if you see, the x axis is the detail of the user, which is username, name, description, location, language, zone and profile picture, just basically showing you, and the y axis is showing you the percentage of users, which is in this case 8 million users,

were actually seen for period of 2 months. What does it mean to for the username, when about 6 to 7 percent of the users change their username at least once, which is, there were two values for these users, that is the way to read the graph.

Let us go to the good one, or the or the one that is higher, in terms of profile picture, if you see, of the 8 million people that were tracked about 40 percent, 35 percent, of the people change the profile picture 3 times at least, right. So, that is the one that is there in the blue. Just on top of it, which is yellow, which is about 40 to 20 percent of the people changed profile pictures for 4 times, and about 10 percent of the people changed it 5 times.

Which means in the period of two months, 40 percent, 10 percent of the people changed their profile pictures at least 5 times. I am sure you can relate it to the behavior that you have, which is just how many times that you change, in my case, probably I change my profile picture once a year or once in a year and half or so. But profile picture changing, I have seen many people change their profile picture pretty often. So, that is what is reflecting on this, the right most thing. And left most thing, where username, similarly for name people have changed the names. And if you look at, about 35 percent of the people are changing their description, which is say, professor at IIIT Delhi, at least two times in the data that was collected.

Nobody changes language, nobody is changing, very few people are changing the zone, time zone, that they are in, right that. So, basically this graph and this graph, the slide 11 and slide 12, is basically showing you the change in information in the account.
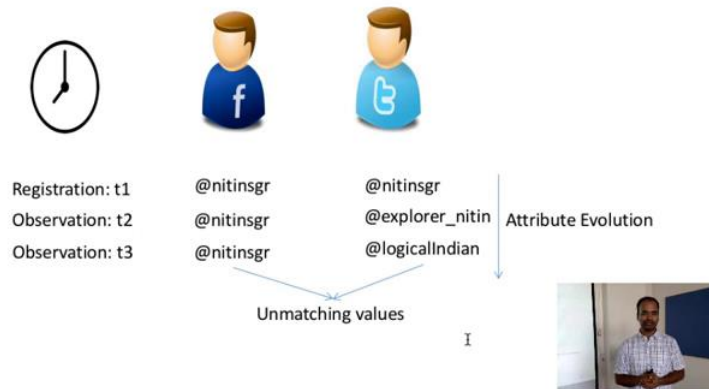
So his is just an example to show you how people change their handles. For example, in my case, I have Twitter now, that is how registered, which is t1, whereas later I could change my user handles as ponguru to become explorer underscore p k (explorer_pk) and at time three, I could change my account as logical Tamilian, for that matter. In that case first one, it was actually identifiable, ponguru, we can probably derive it from my name, second one when I had explorer PK explorer, professor, something like that, slightly getting anonymized (Refer Time: 15:43) and same thing as logical Tamilian also, it is getting anonymized. (Refer Time: 15:44). And it is also unmatching, the point that is expressed in the slide is also to show that the handles ponguru and ponguru at t1 versus ponguru and logical Tamilian at t3 is actually not possible to put together and find the answer. So, there is difficulty in putting this handles together.

Problem Statement

*Given two user profiles and the respective **username** sets, each composed of past and current **usernames**, find if profiles refer to a single individual?*

Given two users - this is more scientific way of asking the question - given two user profiles and the respective usernames sets, each composed of past and current usernames, find if profiles refer to the same individual. That is the question that we are trying to ask, which is, I give you ponguru, and I give you ponguru's current user handle and the past user handles, can you put them together and say that whether it is the same ponguru, which is a (Refer Time: 16:48) professor at IIIT Delhi and ponurangam dot kumaraguru (ponurangam.kumaraguru) in Facebook, ponguru in Twitter, and ponguru in LinkedIn.

So, in this slide the point that is described is that why only usernames, why should we look at only the usernames as the change, as the history, use the information from the history to actually study this profile linking. Because it is unique attribute for a user universally and publicly available attribute, because it is not, you cannot make your user handle private. And sometimes the lines of the handles are also restricted. So, it is not infinite space that I have to actually look for. And of course, in terms data collection, in terms of details that we can actually collect from social media, it is easy for collecting user handles.

So, that is the reason why studying usernames is the way that we looked at.

(Refer Slide Time: 18:00)

Methodology
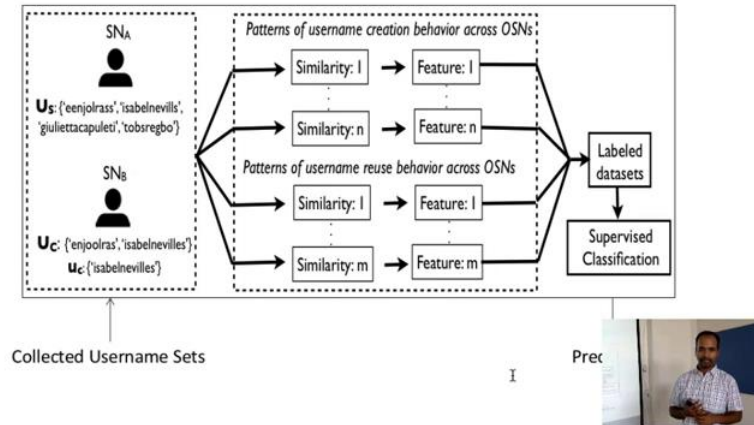
So, this is slightly a dense slide. Let us see how we can actually get this slide across. So, what is given?

(Refer Slide Time: 18:15)


Problem Statement

Given two user profiles and the respective **username** sets,
each composed of past and current **usernames**,
find if profiles refer to a single individual?

So, we would let us go back to the problem statement, the problem statement is, given two user profiles and the respective username sets, which is, I am giving you the handle of ponguru and can l actually find out.

(Refer Slide Time: 18:25)

Methodology
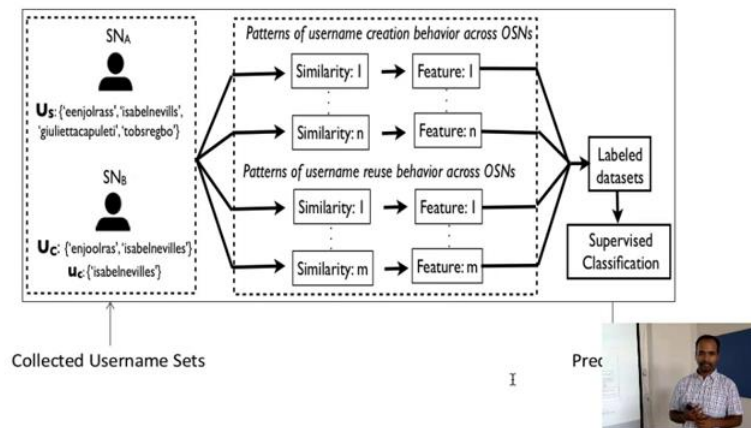
So, that is what is actually explained in this slide which is SN A. SN A, which is in our case let us take it as Twitter, SN B which is something like, let's take we keep it as Facebook. We are going to look at handles in these two networks and find out whether the handles that we are looking at are same, right. So, we look at features, for example, that we said earlier, profile picture, location of the account, use this details as the features and find out whether they are actually the same user.

So, here is a one very, very interesting way and very easy way actually to find out whether it is the same image and there is no probability that this feature may not be useful. But majority of the time this feature is actually very very helpful, what is this? This is Twitter handle, which is, which says, in this case, l u z y, and the user is actually connecting her own Tumblr account in this page, right. So, this basically allows you to say that if I were to find out this handle's, luzy's, Tumblr account, I should just look at the profile. Same way, in my case, if you go, I think, my LinkedIn or my Facebook, has my Twitter account also there. Which means I have explicitly specified, that self identification, which is I am identifying myself, that I am this in Twitter, I am also this in Tumblr, which I'm sure some of you may have done, in my case, I definitely have my precog.iiitd.edu.in URL in my Twitter account.

There is also another way this self identification happens, which again I do it very often, I post pictures on Facebook, I take the link of the album, and then I go post the link to the album in my Twitter account. Which now, if you see, you can actually connect that ponguru, @ponguru in Twitter account, is the same account which is actually posting the pictures on Facebook, which is this album. And therefore, they should be actually the same people. Even without the profile picture, even if my profile picture is different, you can use this to make the decision that it is actually the same user. I hope that is making

sense. As I said, there is a small probability, that this may not be true, but majority of the times this is actually predictable.

Student: Sir (Refer Time: 21:23) say if a person is not very much active on Twitter, but active on Facebook, then how can we link?

Ponnurangam Kumaraguru: Not active, I do not think so, activity frequency actually matters here, right, because let us take -

Student: No, activity, I mean to say if a person is updating something on Twitter (Refer Time: 21:46) he need not update the same thing -

Ponnurangam Kumaraguru: Oh, sure, sure, if the person does not update the same thing on it then it is okay, there is there is going to be always a problem. But if the person does it, it does not have to be the same like what I am saying, linking of the pictures, it could be the same post at the same time and for both pages, both accounts. Even that is useful right? If the person is not posting then I cannot help you, but if the person has the same profile picture, same description and things like that, I can put them together.

But if the person doing the same, like for example, if you see my post, right, I'll do 10-30 on Facebook, 10-30 on Twitter and 10-30 on LinkedIn, all at the same time and it will all be the same content. So, now, it is easy to find out, right, even though my, let's take if I change my Twitter account to instead of ponguru I change it as professor at IIIT, still you can actually make it because it's the same content at the same time.

Student: (Refer Time: 22:55) If I have a Facebook account and I am using it very often, but I never update anything on Twitter, then there is any possibility that we can link the accounts? If there is nothing on the account, the Twitter account, to link it?

Ponnurangam Kumaraguru: Sure. So, if there is no content on Twitter then I think it is slightly hard, but if the account has some details about the account, let us take description, like user profile picture, then it's easy. But in the let me stretch the (Refer

Time: 23:27) even slightly further away, which is that if the person does not even have a Twitter account, what do you do? It's also a problem, right. So, you just went to the extent of saying of the person does not have activity, but if the person does not even have a handle, then it's even harder, right.

Student: It may happen that have my actual Twitter account or some else (Refer Time: 23:54).

Ponnurangam Kumaraguru: Yeah yeah. So, here right Ponguru's my account as in my in in Twitter and I have my Facebook account of Sonu Gupta, what do you do? You just cannot put them together, right. At least from the handles and this, you cannot put them together, but that is why we need to use all these features to put them together. In terms of profile picture, posts that you do, but if you are conscious enough to keep this account two independent, then I think it is impossible to do it, and, but companies like Facebook can do which is beyond what we are taking about in this lectures, because they can actually look at it from the IP address, they can actually look at it from the time of access and still make it (Refer Time: 24:39).

(Refer Slide Time: 24:42)
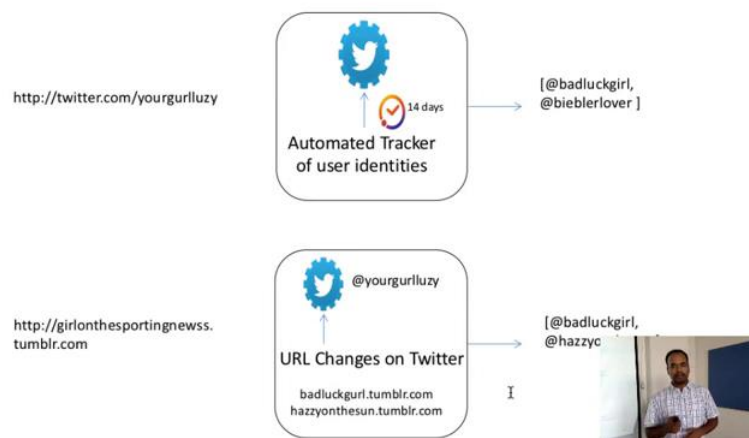
So, if you look at, here, this is showing you ways by which you can actually collect this data, so here the one on the top is showing you that, keep track of a handle. For example, you keep track of ponguru at now, what my handle is, what post am I doing, every day you come and look at this handle, and then you say, oh, suddenly he changes to professor IIIT, professor at IIIT, you can say that the change has happened, right. Because what have, how is the data that is getting stored in Twitter? The basic idea on Twitter is that they give you a unique id - that does not change, right - that id is associated with the handle, you can change that handle.

So, now, you keep track of this id and you know that this id 24 - just making it simple – 24, is actually ponguru. Now you can actually keep track of this 24 always and then ponguru changes to professor at IIIT. Then you add, update, it to your into your database saying that, oh, this handle actually changed. And another way of looking at this is the URL change, which is, the person actually changes the URL in terms of connecting to the other accounts. Like the Tumblr one that I said, somebody is actually going it is in, I am keeping track of Tumblr accounts also and I am keeping track of other accounts. There, the profile is actually changing their description to say that my Tumblr account changed, or in Tumblr they are saying my Twitter account changed. So, this track again you can actually set (Refer Time: 26:38).

(Refer Slide Time: 26:39)

Sample

- **User ID:** 595929421
- **Past usernames on Twitter:**
  - ["**bigeasye_**", "reezy11_", "epiceric_", "soulanola", "**swampson_**", "**hebetheeeric**", "swampkidd_"]
- **Past Usernames on Instagram:**
  - ["bigeasye_", "epiceric17", "swampson", "hebetheeeric"]}

So, here is an example of users whose accounts have changed, whose handles have changed. User ids, as I said before, 24, in this case is 595929421, that is the handle. That is the user id. We keep like a track of that, and if you look at the names that this handle has changed it, when it from bigeasye underscore, to reezy11, to epiceric underscore, to something else, to swampson x y, swamkidd, right. It changed 1 2 3 4 5 6 7 times, it has actually changed the handles.

And if you look at the same user account in Instagram, this seems to also changed four times, but there is a connection between the users handles that the person had in Twitter and then Instagram also. Right, you can actually look at this to also make sense, oh, that there is this swampson in Instagram, then swampson underscore in Twitter, is it the same person? Like, for example, I am sure many of you were, who have common names, for example, Ponnuragan Kumaraguru is not so common, so, if you want to create an account probably you are the only one, you can get the handle, but let's take Shristi Gupta it is so common.

That if you want to create an account now any of the social networks, you are not going to get (Refer Time: 28:21) Shristi Gupta. So, you are probably going to get Shristi Gupta 1 2 3, Sonu Gupta 2 4 6, 19 o 7, things like that. This was, therefore; you can actually use this information also, that some parts of the handle is very similar, so, are they the same people? You could use this Jacqard's distance, and there are many other - Edit distance – there are many other measures by which you can find out whether the, how far is the handle from each other can be also used to say whether it is the same person.

(Refer Slide Time: 28:59)

Methodology

Assumption: Consistent user behavior within and across networks over time
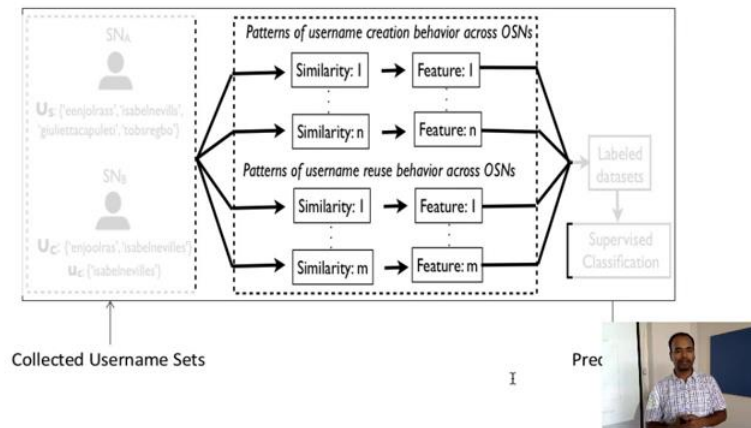
So, here is one version of the same slide that I showed you earlier, which is, usernames are collected, which is what we discussed now. Now we look at some features that you can actually use to put them together and then we'll find out what the predictions are. This is the same slide that I had about 5 or 8 slides before on the whole process of actually identifying whether the handles are same.

(Refer Slide Time: 29:33)

Features
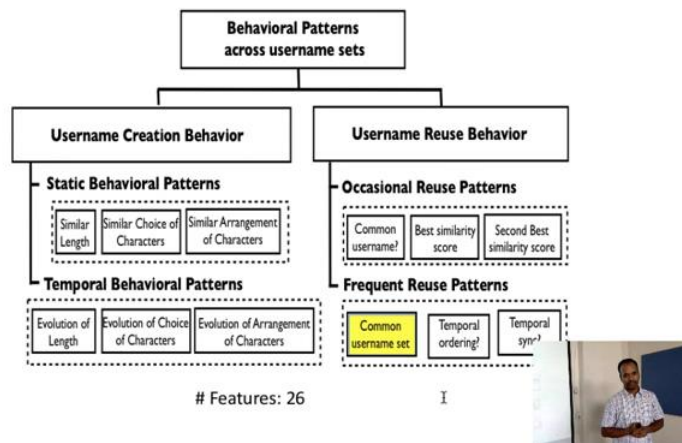
So, here these sets of features that probably actually used in our work. Username creation behavior, she, she just (Refer Time: 29:46) categorized the features into different buckets. Similar length in terms of username, similar choice of characters, similar arrangement of characters ponguru in Twitter and t o n g o r e a 24 in LinkedIn and temporal behavioral feature also, evolution of length, I started with 6, now at 7, now it is 8, what kind of characters are changing, evolution of choice of characters.

(Refer Slide Time: 30:22)

Sample

- **User ID**: 595929421
- **Past usernames on Twitter**:
  - ["**bigeasye_**", "reezy11_", "epiceric_", "soulanola", "**swampson_**", "**hebetheeeric**", "swampkidd_"]
- **Past Usernames on Instagram**:
 - ["bigeasye_", "epiceric17", "swampson", "hebetheeeric"]}

If you see here, all these features, some of these features can be discussed here. If you look at the account details then probably this user started, both Twitter and Instagram is just the same, and then after some point in time the person had epiceric (Refer Time: 30:39) the third in Twitter is the same as the second one in Instagram. Fifth in Twitter is very similar to the third in Instagram. So, you can find this evolution and make something out of this also. Occasional reuse patterns.
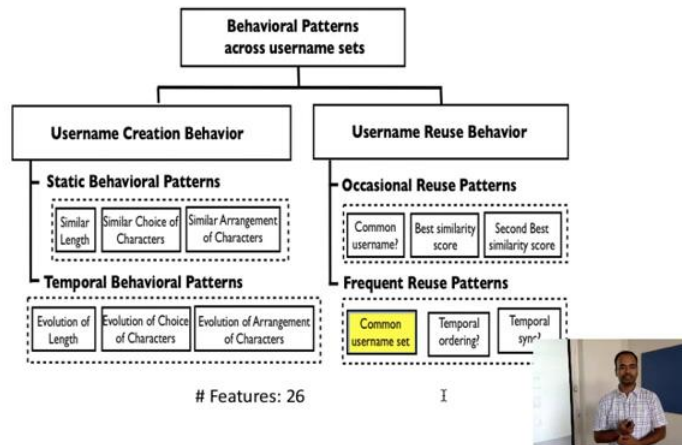
So, common username, the same username being used and features, temporal ordering again, characters, how they are placed; you can use all of these features, which is, Paridhi is calling it more as the behavioral patterns across usernames. You can use these features to say whether the handles are same and the number of features sets that we had was about 56. And if you remember the account, if you remember the trust and credibility section, which is I think week 1 or week 2 that we saw, then we actually saw 45 features in TweetCred and in trust content in Twitter, we saw about 45 features that Aditi used in terms of actually finding out whether this particular content that is posted on Twitter is credible or not (Refer Time: 31:59).

Methodology
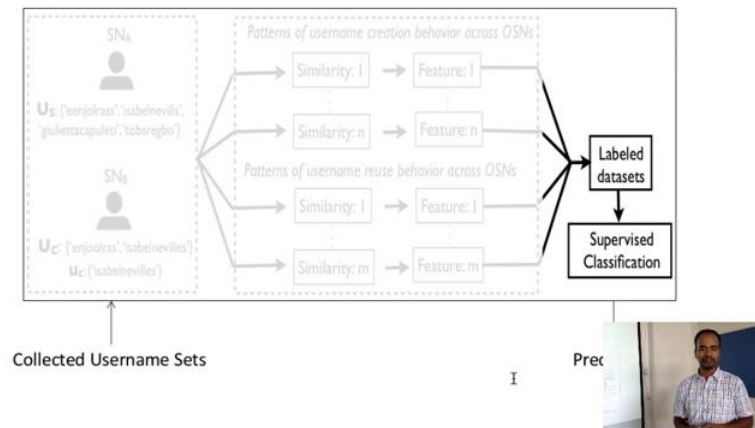
So, now, we have the details from the users, details from the handles, what are the changes in the features, and we can all, we can put them all together to create a set of users, candidates sets, so to call, and then actually make a judgment, give the, give the output as, here is a probability of Sonu Gupta and Sonu Gupta 1 2 3 being the same is 0.9, versus Sonu Gupta and Sonu Gupta 0917 being the probability is about 0.4. So, you can actually make that output, that is the last part.

Now, details. So until now it is more theoretical about how this could be done. Now let just look at specific things what Parishi did in terms of actually getting the data from multiple social networks and finding out how much we can actually do well. So, let us look at some specific examples. In this case we, are looking at data collected between Twitter and Instagram, Twitter and Tumblr, Twitter and Facebook. Past usernames were collected. 21,000 positive pairs, which is, details that collected from these social networks, and about past usernames available only on Twitter, but current usernames available on other profiles is about 140,000. So, essentially the idea is that the data was collected between multiples social networks of the current and the past user handles and how this was put together, and what kind of mechanisms was used to find out, whether these handles are same.
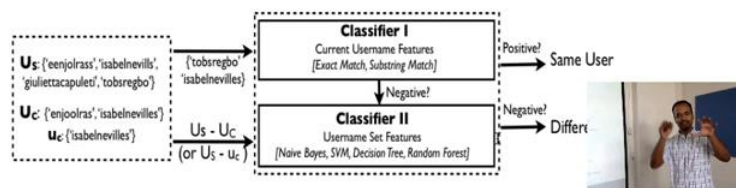
(Refer Slide Time: 33:54)
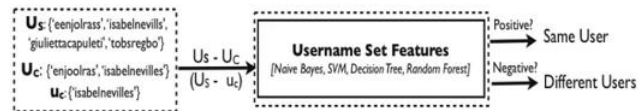
And it is the same diagram that I showed you before, take the handles, understand some features, put the features together, and create the score. And that is what is done here - two methods are done, one is you just do only the features, and the other method that we did was, do a classifier and then apply it and to find out whether it is the same user.

(Refer Slide Time: 34:22)

Prediction

Framework Config.	Accuracy	FNR	FPR
Exact Match (b1)	55.38	89.34	0.00
Substring Match (b2)	60.99	78.46	0.00
Independent [Naive Bayes]	72.19	55.86	0.13
Cascaded [b1→Naive Bayes]	72.48	55.27	0.14
Cascaded [b1 → SVM [Linear]	76.74	45.16	1.65
Cascaded [b2 → Naive Bayes]	72.51	54.97	0.1
Cascaded [b2 → SVM [Linear]	76.84	45.16	1.2

Here are the different methods that was used, which is, exact match, right, exact match of the handles, substring match, and then classifiers, different classifiers were applied and if you look at the first one and the fifth one, are the ones which had the maximum accuracy, which basically says that if you look at the handles, the way that the handles look similar,

, and if they are exact match, is very high probability that they are the same users, that is, very less probability, that they would actually be different users. And if we use the SVM classifier and then apply it on to find out whether there is same users using all the, using all the 26 features that we talked about, there is high probability that will be able to, about 76 percent is the accuracy to find out whether they are actually the same handles.

(Refer Slide Time: 35:29)

Prediction

A comparison of cascaded framework accuracy with and without Twitter-Tumblr instances

Framework Config. [History on Both or One]	Accuracy	FNR	FPR
Exact Match (b1)	55.38	89.34	0.00
Cascaded [all network]	76.74	45.16	1.65
Exact Match without Tumblr (b1)	66.17	67.51	0.00
Cascaded [without Tumblr]	91.20	16.60	C

(Refer Slide Time: 35:31)

So, let me show you some, why, so, if you remember the motivation that I started off with is this sending this advertisement to people, I do not want to waste my money. Just let us go back and connect to the motivation also. There are, if PK or NPTEL wants to send out advertisement to all the students who are on Facebook and Twitter, we have to send this information to only one user only once, you do not want to duplicate and waste money, that is the motivation I started, but the motivation can be many other things. There are some examples also.

So, in this case if you see, the sentiments of the user for a, let us take any topic that you take, the one on the left is on Facebook, the one on the right is on Twitter, you really want to understand whether sentiment, for example, you just look at this, you want to understand whether the sentiments of these people are expressed on Facebook and Twitter, are they same, and if they are same or if they are different, are they are same user.

So that, you can actually measure that the negative sentiment of any topic is not a sum of all negative sentiments in all social networks, but only the unique people that you want to take a note of, (Refer Time: 36:52) right. Because if I say something positive in Facebook, and I am same person who is saying positive in Twitter, it is not, if you, you

cannot measure the positivity as twice, but it is only once because it is only one persons sentiment, right. So, that is another motivation. The other motivation also that, the other reason why this ==identity== resolution is an interesting problem is because you can actually look at, even law enforcement can actually use this. Which is, somebody uploaded a malicious video on YouTube and in there is another handle, which uploaded the same video on Twitter.

Now I want to find out whether ==it is the same person who is posting it.== Somebody is actually speaking against some people or some organization on Twitter. And there is another handle which is speaking against some persons or some organization in a different network. I want to know whether it is the same person, because they do not want to be wasting the time in assuming that it's two people and wasting time in finding two people, but it is only one person that they have to chase and catch, right. So, that is the motivation, that is also another motivation to actually find out whether these two handles are same or not, right. So, there is very interesting motivation for doing this work, and there is a lot of interesting things one could actually try out.

(Refer Slide Time: 38:29)

Conclusion

- Profile linking may be necessary for many organizations / needs
- Better profile linking is possible with past history of user handles

So, here the last take away from this part, which is profile linking may be necessary for many organizations as the questions that we said, I do not want to waste my money, I

want to actually understand, whether, how many people have posted, what is the volume of actually positivity or negativity, or I want to find out who is actually speaking online and to link users.

And the conclusions from this work are that essentially you do not have to only bank on the current handles that people have, current information that people have; even using the ones that from the past can actually prove the efficiency, accuracy, of the – and that is Paridhi, who is in the picture, who just graduated with a PhD from the work that she did on this topic.

(Refer Slide Time: 39:29)

I am going leave you for this week just to try this out. We can probably connect this to the quiz that we have or homework that we have for the course also, but here is what I want to try, you to try. Take two of your accounts on your two different social networks, which is Facebook and Twitter. Let's just stick to only Facebook and Twitter. Just take these two handles, list down all the things that you can actually do, find out various ways in which you can actually link these two accounts. List the features, features just we talked about, right? These ones or there could be many others also. List down these, and list down things that you will change in the profile to make them look two different account networks also. You could do both ways.

You could do list down things that you will change to make it two different account, or list down things that you will do to make it the same account. Right? Share it in the forum, let's see what you people actually come up with. I hope the activity is clear. It is that take your Facebook account, take your, take your Twitter account, list down the features that are available, that you think you can actually connect with the accounts. And list - that is the first output - second output is list down all the things that you will do to make it look same - that is the second account, second output. The third output is, list down all the things that you will do which will make that it is two different accounts.
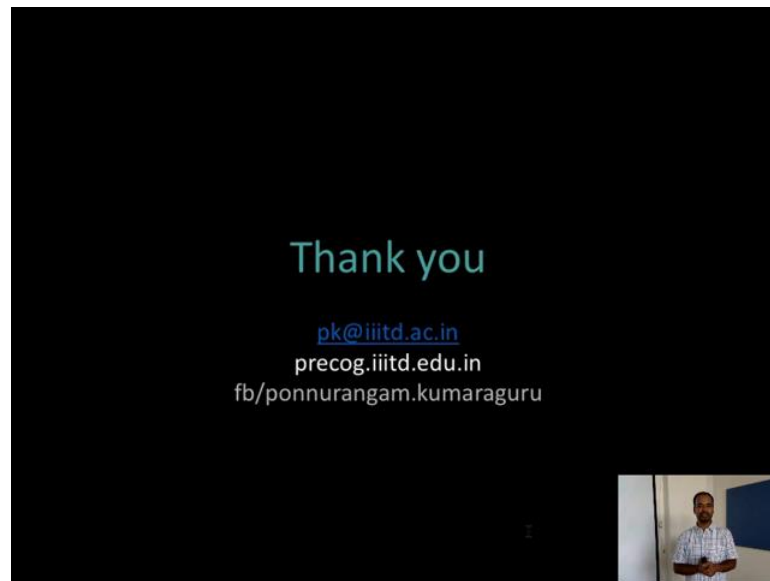
(Refer Slide Time: 41:16)

References

- Paridhi Jain's Ph.D. thesis work

(Refer Slide Time: 41:19)

So, that is all I had for this week that this 8.1. I will actually continue in a different topic when I start off with 8.2.