

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

Week 1.1
Lecture – 01
Intro to Course

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

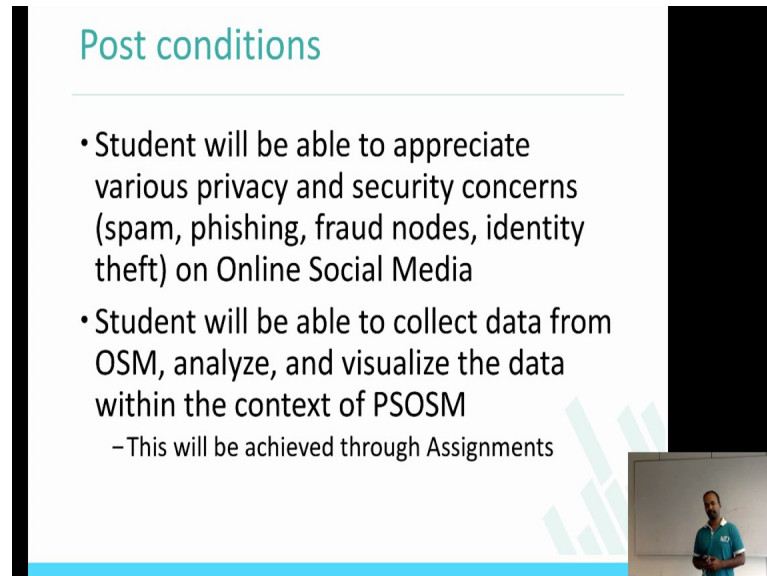
(Refer Slide Time: 00:52)



This is the Facebook post, I did about a year and half back which is regarding the course feedback that students give us at IIIT, Delhi. I have been teaching this course PSOSM Privacy and Security in Online Social Media at IIIT, Delhi for a couple of times. So, this is the feedback and we are there with the course. We actually have a post session where

students actually present their work, what they have done over the semester in the form of a poster, in the form of a demo. So, this is the picture with all the students from the class, wearing **the same T shirt** and with **actually the** external evaluators who to came to evaluate **these** projects.

(Refer Slide Time: 01:33)



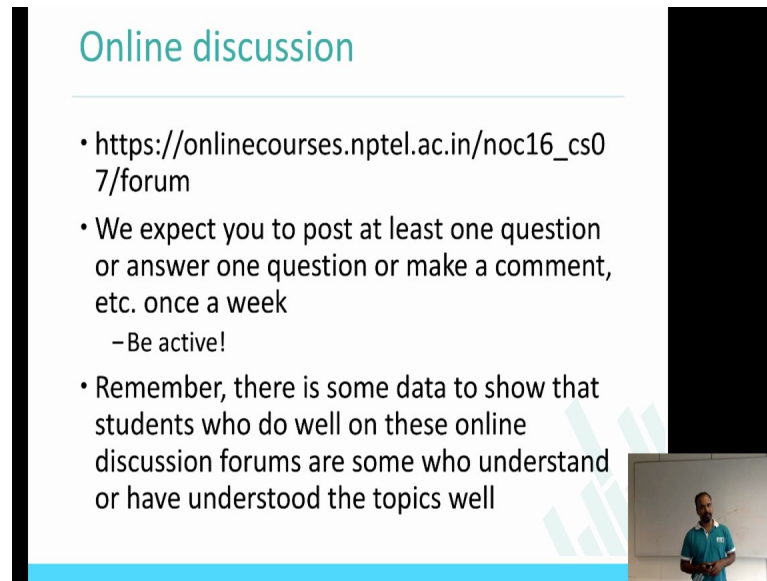
Post conditions

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

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

Throughout the course, you **will also be** exposed to actually collecting data from online social networks like Facebook, Twitter **analyzing these** content and visualizing this data in terms of the question that you are trying to ask, for example, 1 percent could be I want to understand whether the followers that I have on Twitter are actually legitimate or fake. We can do actually have achieve this goal, this post condition in terms of able to collect data, analyze data and visualize data through the assignments that you could be getting across these course.

(Refer Slide Time: 02:36)



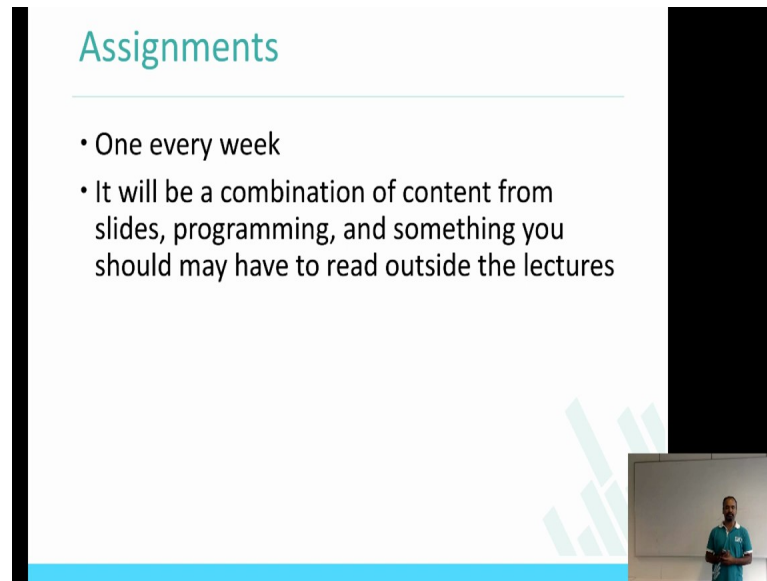
The slide is titled "Online discussion" in a teal font. It contains a bulleted list of expectations for students. A small video inset in the bottom right corner shows a man in a blue shirt standing in front of a whiteboard.

Online discussion

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

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

(Refer Slide Time: 03:11)

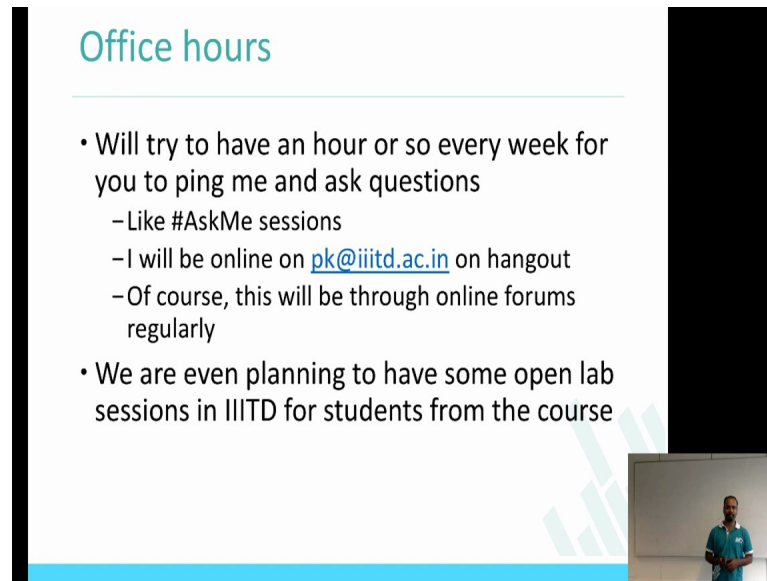


Assignments

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

So, what we plan to do in some of assignments is, we plan to have one homework assignment per week capturing the topics that we have covered in the class, in the lecture and we will actually have homework questions around that every week and we hope to actually get you a sense of these kind of different topics same as collecting data from online social network, what kind of analysis can be done with this data? How to visualize a data and things like that? Mostly these questions will be from slides, some programming; something that should actually be able to answer if you actually read content outside the lectures in the pointers that we discuss in the forum also.

(Refer Slide Time: 04:00)



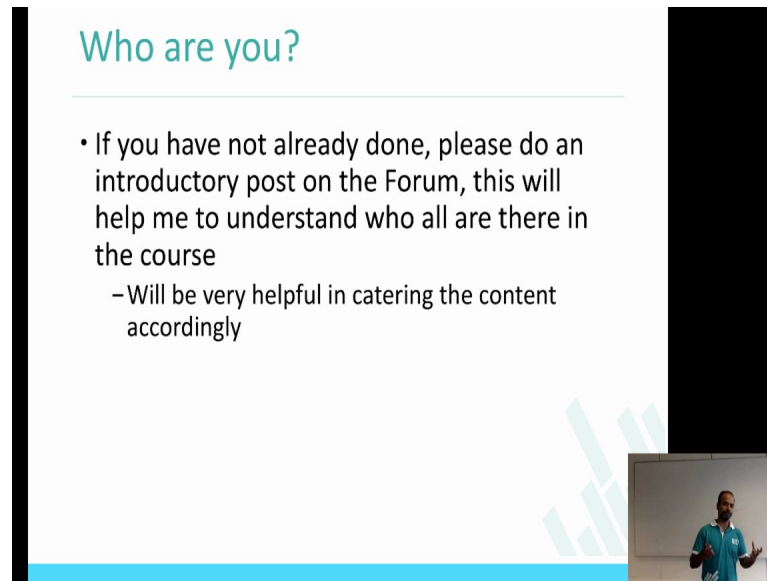
Office hours

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

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

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

(Refer Slide Time: 05:14)



The slide has a light blue background with a dark blue vertical bar on the left and a dark blue vertical bar on the right. The title 'Who are you?' is in a teal color. Below the title is a horizontal line. The main content is a bulleted list. In the bottom right corner, there is a small video inset showing a man in a blue shirt speaking.

Who are you?


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

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

(Refer Slide Time: 05:52)

Why Should I Teach this course?

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



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

(Refer Slide Time: 06:39)



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

That is **Srishti** Gupta, whose primary interest is studying the online social media and with the phone numbers and OTT kind of technologies that are available. That is **Prateek**, whose interest is on studying malicious content on Facebook. That is Niharika Sachdeva, whose interest is primarily on visible security and studying how technology and social networks have been used by police organizations around the world and particularly in India. So, these 4 TA's will help us in doing **lab** sessions, setting up questions answers for the homework and helping us in general making the course more **exciting** and interesting for us.

(Refer Slide Time: 08:11)



Topics that we will cover

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

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

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

Then we will also look at privacy issues on **online** social network. Privacy is becoming such a big topic because of the **proliferation of** online social networks, what information is leaked? What information can be actually collated? What information can be stitched together to create a profile user or which can be actually misused against. Then we also look at social network analysis, text analytics that can be done using the content for social network. NLTK is one of the platforms **which we** will also expose you to analyze

the content from social networks. This again would be a hands-on session where you will get experience on using these tools, techniques from the content in the course. E-crime which is also much related and very important topic in the **context** of online social networks is something we had also covered.

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

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

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

At the end of the course, we will actually review with some very broad questions and very broad topics which are connected to social networks which we will not be able to cover in this course like deep learning, machine learning, natural language processing, image analysis. These are the topics that are becoming very, very popular in terms of using social network data because of the proliferation of the social network and understanding what data is available and how we can use this data is becoming a very important topic.

Mostly if you all see, image analysis is also becoming an important one because mostly these days the posts are coming with images. It is not just text only; it is actually text and images or sometimes only images. So, this will be a very broad tour of these new topics that are popping up around the online social network topic.