Introduction to Social Networks

Listed as COMS 4995-1, Fall 2011, Columbia University
Instructor: A. Chaintreau,

This website provides general information and access to teaching material for an introductory graduate course on social networks for computer science students.

- You will find below on this page a very brief statement about the course, more information is available on the ongoing detailed content section of this site.
- Find more about logistics and grading
- For your information, this website contains a list of references pointing to class with overlapping contents, a brief bibliograph, and a list data-sets
- You can browse opportunities to get involved in real project related to social networks, or hear about what the web and newspapers recently tell about them

Objective of this course:

Social network concepts are ubiquitous in the transformation of many commercial and public services today, and they become more and more common in many aspects of computing. The objective of this course is to introduce these concepts to students with a computer science background, with two special focus:

- Interpret real-world phenomenon using elaborate concepts from social network analysis.
- Design algorithms that exploit social properties to optimize system design, with a justification of their convergence and performance.

Said differently, this course will tell you a first answer to WHY and HOW computing methods that manipulate social networking concepts have been shown surprisingly efficient.

Note that this course does NOT cover some important aspect of dealing with social network and computation, in particular:

- Game-theoretic and economic analysis of the role of social network is not covered in these lectures (but it will be occasionally mentioned, and some aspect will be covered in a upcoming advanced class)
- How to build from a system point of view a social networking service (although the concepts you will learn in the class would be important to have in mind to properly design such a system).
- How to analyze massive data sets using Hadoop or Map reduce, and how to conduct inference and machine learning on such data set. This topic is introduced through some examples, but it is too important by itself to be treated well during a general introductory course.
Note that these topics are not completely independent or in contradiction with the objective of this course, but they require complementary skills. If your long lived goal is to train on those skills, attending this introductory class will give more breadth in general concepts connected to these other important questions.

Who should take this class?
A - Is there some programming skills required for this course?
I am not planning to ask any programming homework that require to handle some examples and code (It would in fact be a great thing to do, but it seems to me that, outside a little bit of practice of the concepts, it would take too much time for what you can actually do in this semester).

Bottom line:
So, if you for any reason (different background etc.) you are uncomfortable with coding, this will not be a problem for this course. It's still a good idea to practice that soon somewhere else though!
If you have the opposite problem (you can only breath by coding), I think it's not entirely impossible to do some of that for the course. As an example, answering an interpretation question with a small empirical study (take one of the data set referenced on the website and run some code on it) would be (1) super cool and (2) a way to earn extra credit for some appropriate extra time, (3) a good way to get yourself on the starting block for a project.

B - How comfortable should I be with maths to take the course?
Although no formal prerequisite is enforced, the course is about how algorithms take advantage of important social network concepts with the goal of justifying them by proof, so we will do proof (no later than Thursday). Only simple concepts of Graph theory, Elementary linear algebra and Probability are considered, but they are used frequently and extensively in the proof.

Bottom line:
If you are still wondering how to judge if you can do it, since you need to figure out this week and the next I recommend the following
(1) have a look at a Chapter 2-3 from Easley-Kleinberg's book "Network Crowd and Markets" (available for free online), including the advanced material section. This is a very well written book, so you should be able to read it like a novel, and basically in an hour or perhaps two digest the contents pretty well. If you start encountering some difficulties in doing that quickly, then shoot me an email and we can discuss it further. It's also a great preparation for the class on Tuesday next week.
(2) attend the two first weeks as a test: I organize these 4 first classes on purpose as an increasing slope. By the time we get to the algorithmic analysis of small world (about Sep.15th), you will have been introduced to some of the most subtle concepts. If at this time you feel comfortable, that's a good sign because the class will move on to other topics and different tools, but not much more vertically anymore.

Contents and Organization:
The course is organized in 3 parts:
Part I - STRUCTURES focuses on properties that characterizes graph structure that represent social interaction between individuals. It answers the following question:
How do we connect to each other in a graph (i.e., Why am I only separated from all Facebook users through very small number of hops, and why it matters)?
Why and how are large scale networks skewed (i.e., Should we expect very popular videos on YouTube)?
How are social network spread in groups and places (i.e., How are my community, my Flickr images and Foursquare embedded in my social networks)?

Part II - DYNAMICS focuses on interaction between individual along the edge of a social network
How do epidemics spread (i.e., why are computer worm, and peer-to-peer networks, so efficient)?
How am I influenced by my friend (i.e., how to make a product viral)?
How sharing information with my friends allows to learn collectively (i.e., can 10,000 persons communicate and be wrong)?

Part III - APPLICATIONS deals primarily with important concepts in direct use of computing in social networking
How to measure social network (i.e., are social networking entirely closed or entirely open) ?
How to extract important nodes (i.e., how does google rank pages, and how does Netflix recommend movies)?
How to protect users and their data from malicious usage (i.e., are fake accounts a real threat and should privacy be in history books)?

For more information, please visit Detailed Contents