Social Network Analysis and Graph Mining (社群網路分析)

Prof. Shou-de Lin 林守德
CSIE/GINM, NTU
Course Information

• Lecturer: Shou-de Lin (office: CSIE R333)
• Meeting Time: Tue 2-5pm R102
• TA: Cheng-Te Li 李政德 (office: CSIE R302)
  – TA Time:
    17:00~18:00 Monday, 13:00~14:00 Tuesday
What is Social Network

• As a concept
  – It represents people connect to each other in the real world.

• As an encoding schema
  – A kind of data structure that encodes the dependency between objects

• As an application
  – An online program that enables the social mingling of people (e.g. Facebook)
Research Topics about Social Networks

• How to identify, find, or simulate a social network?
  – Natural Data: Information Extraction from Web
  – Artificial Data: Sampling

• How to analysis and model a social network
  – The main focus of this course

• How to present a social network?
  – Visualization
  – Abstraction
  – Explanation
Social Network Analysis: before and now

• Previously: people study networks of hundreds or thousands of nodes/links
  – Sample Q: Which vertex in this network would prove most crucial to the network's connectivity if it were removed
  – The human eye is an analytic tool of remarkable power, and eyeballing pictures of networks is an excellent way to gain an understanding of their structure.

• Now: study networks of millions or billions of nodes
  – Sample Q: What percentage of vertices need to be removed to substantially affect network connectivity in some given way?"
  – How can one tell what this network looks like, when one can't actually look at it?
Snapshot: from “synthesis” to “analysis”

Problems
- **High cost of manual surveys**
- **Survey bias**
  - Perceptions of individuals may be incorrect
- **Logistics**
  - Organizations can spread across several countries.

[from J. Srivastava’s Slides]
Three Primary Focus of SNA

1. Find statistical properties that characterize the structure and behavior of networked systems,
   – E.g. path lengths, degree distributions, centrality
   – suggest appropriate ways to measure these properties

2. Create models of networks that can help us understanding the meaning of these properties
   – how they came to be as they are, and how they interact with one another.

3. Predict the behavior of network systems, or using the network system for prediction. For example
   – How will network structure affect traffic on the Internet?
   – How will disease/information propagate in a social network?
   – How can a social network evolve through time?
About this SNA Course (1/2)

• What this course is about?
  – Conveying the basic ideas and advanced technologies in social network analysis
  – Introducing some graph mining packages/tools
  – Equipping learners with capability of implementing a social network analysis system
  – Discussing state-of-the-art research results in social network analysis
  – This course is CS-oriented rather than sociology-oriented

• Why should I take this course
  – I have interests to know why social network becomes so popular in these days😊
  – I am hoping to do research in this area 😊
  – I heard SD offers generous grades 😊
  – I need a networking course to graduate in GINM 😊
About this SNA Course (2/2)

- What do I need in order to take this class
  - Familiar with probability and basic programming skills
  - Having the passion and time to study hard
- How to do well in this class
  - Try coming to the class and spending time on the assignments
- I have taken some of your previous courses (e.g. SAI), would it be a waste to take this course?
  - No, they are completely different
- Will I guarantee to pass the class?
  - No, last year about 20 students dropped or failed in the SAI class.
Grading

- Three to Four Programming Assignments (55%)
  - You can choose any language you prefer (recommend java)
  - Most of the assignments should be submitted Individually

- Paper Presentation and reports (10%)
  - We will have about 12 papers to be presented
  - People who did not present shall submit their reports

- Final Project (30%)
  - Team project

- Class Participation (5%)
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Homework/Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>9月15日</td>
<td>Intro, basics</td>
<td></td>
</tr>
<tr>
<td>9月22日</td>
<td>basics &amp; small world</td>
<td></td>
</tr>
<tr>
<td>9月29日</td>
<td>power law, random graph</td>
<td>homework 1.1 out</td>
</tr>
<tr>
<td>10月6日</td>
<td>process model</td>
<td></td>
</tr>
<tr>
<td>10月13日</td>
<td>paper presentation 1</td>
<td>homework 1 due</td>
</tr>
<tr>
<td>10月20日</td>
<td>community detection and clustering</td>
<td>homework 2.1 out</td>
</tr>
<tr>
<td>10月27日</td>
<td>social position analysis &amp; cohesive subgroup</td>
<td>homework 2.2 out</td>
</tr>
<tr>
<td>11月3日</td>
<td>outlier detection &amp; centrality</td>
<td>homework 2.3 out</td>
</tr>
<tr>
<td>11月10日</td>
<td>paper presentation 2</td>
<td>homework 2 due</td>
</tr>
<tr>
<td>11月17日</td>
<td>Dynamic social network</td>
<td>homework 3 out</td>
</tr>
<tr>
<td>11月24日</td>
<td>Existing works on heterogeneous social network</td>
<td></td>
</tr>
<tr>
<td>12月1日</td>
<td>Heterogeneous social network Mining</td>
<td></td>
</tr>
<tr>
<td>12月8日</td>
<td>link prediction &amp; learning</td>
<td>homework 3 due, homework 4 out</td>
</tr>
<tr>
<td>12月15日</td>
<td>project proposal presentation, presentation 3</td>
<td></td>
</tr>
<tr>
<td>12月22日</td>
<td>Graphic learning model (HMM, CRF)</td>
<td>homework 4 due</td>
</tr>
<tr>
<td>12月29日</td>
<td>paper presentation 4</td>
<td></td>
</tr>
<tr>
<td>1月5日</td>
<td>final project presentation 1</td>
<td></td>
</tr>
<tr>
<td>1月12日</td>
<td>final project presentation 2</td>
<td></td>
</tr>
</tbody>
</table>

Project proposal due