E6885 Network Science Lecture 1:

*Overview of Network Science*

Ching-Yung Lin, Dept. of Electrical Engineering, Columbia University
September 9th, 2013
## Course Structure

<table>
<thead>
<tr>
<th>Class Date</th>
<th>Number</th>
<th>Topics Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/09/13</td>
<td>1</td>
<td>Overview of Network Science</td>
</tr>
<tr>
<td>09/16/13</td>
<td>2</td>
<td>Network Representations and Characteristics</td>
</tr>
<tr>
<td>09/23/13</td>
<td>3</td>
<td>Network Partitioning, Clustering and Visualization</td>
</tr>
<tr>
<td>09/30/13</td>
<td>4</td>
<td>Network Estimation and Modeling</td>
</tr>
<tr>
<td>10/07/13</td>
<td>5</td>
<td>Network Analysis Use Case</td>
</tr>
<tr>
<td>10/14/13</td>
<td>6</td>
<td>Network Topology Inference</td>
</tr>
<tr>
<td>10/21/13</td>
<td>7</td>
<td>Graph Database</td>
</tr>
<tr>
<td>10/28/13</td>
<td>8</td>
<td>Dynamic and Probabilistic Network</td>
</tr>
<tr>
<td>11/11/13</td>
<td>9</td>
<td>Final Project Proposals</td>
</tr>
<tr>
<td>11/18/13</td>
<td>10</td>
<td>Information Diffusion in Networks</td>
</tr>
<tr>
<td>11/25/13</td>
<td>11</td>
<td>Graphical Models and Bayesian Networks</td>
</tr>
<tr>
<td>12/02/13</td>
<td>12</td>
<td>Impact of Network Analysis</td>
</tr>
<tr>
<td>12/09/13</td>
<td>13</td>
<td>Large-Scale Network Processing System</td>
</tr>
<tr>
<td>12/16/13</td>
<td>14</td>
<td>Final Project Presentation</td>
</tr>
</tbody>
</table>
Course Grading

- 3 Homeworks: 50%
- Final Project: 50%
Other Issues

- Professor Lin: Office Hours: Monday 9:30pm – 10:00pm (Mudd 227) or by appointment

- Contact: cylin{at} ee{dot} columbia{dot} edu

- Telephone: 914-945-1897

- TA: Xiao-Ming Wu <xw2223>; Office Hours: Friday 2-4pm (7LE3 Schapiro Building)
Course Textbook

Networks Everywhere

Gene Coexpression Network

Scientific Workflow

Chemical Compound

Mesh
A new science, named psychological geography, which aims to chart the emotional currents, cross-currents and under-currents of human relationships in a community, was introduced here yesterday at the scientific exhibit of the Medical Society of the State of New York, which opens its 127th annual meeting here today at the Waldorf-Astoria.

The first series of maps of the new human geography were shown by Dr. Jacob L. Moreno of New York, consulting psychiatrist of the National Committee of Prisons and Prison Labor and director of research, New York State Training School for Girls, Hudson, N. Y. The maps represent studies of the forces of attraction and repulsion of individuals within a group toward one another and toward the group, as well as the attitude of the group as a whole toward its individual members, and of one group toward another group.

Emotions are represented on these psychological maps by various colored lines. Red stands for liking, black for disliking. If individual A likes B a red line with an arrow points from A to B. If B reciprocates a similar red line points from him. If he dislikes A this is indicated by a black line with an arrow pointing toward A. If B is merely indifferent the feeling is shown by a blue line.

Group of 500 Girls Studied.
Network Science <=> Graph Technology
Types of Graph

**RDF / Property Graph**

- Attributes
  - Charles Flint
    - born: "1850"
    - died: "1934"
  - IBM
    - HQ: "Armonk"
    - employees: 433,362
    - industry: Software, Hardware, Services

**Activity Graph**

- Micro
  - Personal Event
  - Personality
  - Job Event
  - Unusual Activities
  - Planning
  - Stress
  - Workplace Conflict
  - Attack

- Macro
  - Network of individuals and connections

**Collective Graph**

- Network of collective entities
Characteristics of Network Data

- High-Dimensional
- Dependent
- Massive
The Emergence of Network Science

- Science <=> Observable systematic empirical data
- Facility of large-scale data collection, storage and management.
The Emergence of Network Science

- Science <=> Observable systematic empirical data
- Facility of large-scale data collection, storage and management.

- Statistical Methodologies to combine behavior understanding, link analysis, multi-variant modeling, machine learning, graph theory, and non-parametric statistics for complex network analysis

Contributions made by Physicists and Computer Scientists, have greatly expanded the discipline over the past 15 years.
Multi-disciplinary Research Issues

- Formation of Network
  - Communications
  - Information
  - People
  - Companies / Organizations
  - Nations

- Network Data Collection

- Network Science Infrastructure

- Network Applications

- Network Visualization

- Network Sampling, Indexing and Compression

- Network Flow

- Network Evolution and Dynamics

- Network Impact

- Cognitive Networks
Multi-disciplinary Research Issues

- Formation of Network
  - Communications
  - Information
  - People
  - Companies / Organizations
  - Nations

- Network Data Collection

- Network Science Infrastructure

- Network Applications

- Network Visualization

- Network Sampling, Indexing and Compression

- Network Flow

- Network Evolution and Dynamics

- Network Impact

- Cognitive Networks
Example Use Case Applications of Network Science

1. Network Analytics for Expertise Location
2. Network Analytics for Recommendation
3. Network Analytics for Commerce
4. Network Analytics for Financial Analysis
5. Network Analytics for Social Media Monitoring
6. Network Analytics for Telco Customer Analysis
7. Network Analytics for Deep Q&A
8. Network Analytics for Data Exploration and Visualization
9. Network Analytics for Personalized Search
10. Network Analytics for Anomaly Detection (Espionage, Sabotage, etc.)
11. Network Analytics for Fraud Detection
12. Network Analytics for Cybersecurity
13. Network Analytics for Sensor Monitoring (Smarter another Planet)
14. Network Analytics for Cellular Network Monitoring
15. Network Analytics for Cloud Monitoring
17. Network Analytics for Traffic Navigation
18. Network Analytics for Image and Video Semantic Understanding
19. Network Analytics for Genomic Medicine
20. Network Analytics for BRAIN
Example: Network Science Consortium (2009 – 2014 or 2019)

- U.S. Army is investing $172 million in 10 years to fund 4 Network Science academic research centers:
  - Communication Network Academic Research Center (CNARC)
  - Social and Cognitive Network Academic Research Center (SCNARC)
  - Information Network Academic Research Center (INARC)
  - Interdisciplinary Research Center (IRC)

~ 100 Professor/Researcher Principle Investigators + 250 RAs, Postdocs

- Objectives:
  - Improve Decision Making
  - Develop measures, metrics and models that describe and predict human-network interaction and exchange within & across network layers
  - Develop and validate theory of human-system interaction in network-centric environments
  - Explore techniques for dynamic, flexible, adaptive, and adaptable interaction

1. Charactering and Measuring Networks
2. Understanding Networks for Analysis
3. Controlling and Managing Networks
4. Using Networks
Dynamics in Graphs

**Heterogeneous Synchronicity Networks Predict Performance**

One-class HCRF to detect temporal anomalies

Detected as top 1 anomaly in Sandy Tweets
Dynamics of Information Graphs in Social Media

• Motivation:
  – Info morph: new links keep emerging to give new meaning to existing phrases

• Approach:
  – Compare characteristics of meta-paths between nodes in heterogeneous networks

Entity morph resolution accuracy (ACL 2013)

- Peace West King from Chongqing fell from power, still need to sing red songs?

- Bo Xilai led Chongqing city leaders and 40 district and county party and government leaders to sing red songs.

Bo Guagua

Bo Xilai

Chongqing

CCP

China

TG

Best Actor

Wen Jiabao

Sum

\[ \sum_{i=1}^{N} p_m(x_i) \log \frac{p_m(x_i)}{p_e(x_i)} + p_e(x_i) \log \frac{p_e(x_i)}{p_m(x_i)} \]
SmallBlue asset: Large-Scale People Modeling and Social Network Analysis

- Emails
- Chats
- Meetings
- Web Page Clicks
- Server Logs

Social sensors

Click streams capturer

Feed subscription

Network analysis

People analysis

Content analysis

Applications

Live Data, Live Production System

- 20,000,000 emails & SameTime messages
- 1,000,000 Learning click data
- 14,000,000 KnowledgeView, SalesOne, …, access data
- 1,000,000 Lotus Connections (blogs, file sharing, bookmark) data
- 200,000 people’s consulting financial databases
- 400,000 organization/demographic data
- 100,000 intranet w3 searches per day
SmallBlue Applications (I): Find knowledgeable colleagues

- E.g.: Search for the most knowledgeable colleagues within my 3-degree network for who knows 'healthcare'. (or within a country, a division, a job role, or any group/community)

As a user, you can only see their public information. Private info is used internally to rank expertise but private data can never be exposed.

Click a name to see their profile (SmallBlue Reach).

My shortest path to Susan

As a user, you can only see their public information. Private info is used internally to rank expertise but private data can never be exposed.

Click a name to see their profile (SmallBlue Reach).
SmallBlue Applications (II): Reach – social dashboards

- Is Tom a right person to me?

![Image of SmallBlue Suite dashboard]

- **His official job role, title, contact info**
  - Email: tom.a.coccoza@us.ibm.com
  - Telephone: 1-703-533-4731

- **His public communities**
  - Communities
  - SmallBlue Net
  - Alternative Paths
  - Recommended Path

- **His self-described expertise**
  - Expertise:
    - Federal government financial management, Healthcare financial management
    - Business Strategy, Accounting, Processes, Accounting Standards & Certification, Auditing-Business Intelligence, Executive Communications

- **The public interest groups he is in**
  - The public interest groups he is in
  - SmallBlue Net
  - Alternative Paths
  - Recommended Path

- **His blogs, forum, postings..**
  - Public postings
  - BlogCentral
  - No information

- **My various paths to Tom. SmallBlue can show the paths to any colleagues up to 6-degree away**
SmallBlue Applications (III): Net – corporate social network analysis

- How are company’s top healthcare experts link with each other? Who are the key bridges? Who have the most connections? How do these experts cluster?
SmallBlue Applications (III): Net (cont’d)

Healthcare experts in the world

Connections between different divisions

Healthcare experts in the U.S.

Key social bridges
SmallBlue Applications (IV): Ego – personal social network capital management [an application only visible to the user himself]

- What is a friend’s social capital to me?

My personal social network automatically found by SmallBlue with social distance

What types of unique colleagues my friend Chris can help me connect to?

How many people in my personal networks?

It can also show the evolution of my social network..

SmallBlue finds dynamic existing social networks of every employee That makes it possible to find the shortest path to any colleague.
Personalized Content Recommendation and Search

- What your friends know become what you know. Your friends are your window to the world – Confucius ~600 B.C.
- Utilizing the unique large-scale weighted social network inferred by SmallBlue, personalized ranking becomes possible.
- Fusion of Recommenders: Social Filtering, Collaborative Filtering, Latent Semantic Filtering, Popularity & Freshness Filtering, etc.
- Deploy on IBM KnowledgeView, IBM Learning, and IBM TAP

SmallBlue Whisper

- Whisper recommends...
  - [Image: SmallBlue Whisper and Synergy usages]
  - On Nov. 19, 2008: 11,108
  - On Nov. 20, 2008: 9,567

SmallBlue Synergy

- Get understanding of the users interests based on:
  - Keywords from SmallBlue communication analysis
  - Link analysis

Synergy – Personalized Content Search
**SmallBlue Mobile**

- SmallBlue Applications on Mobile Phones

  Show Expertise of 'SNA' inside:
  
  1. IBM
  2. My 2-degree network
  3. Research division
  4. Global Business Services
  5. Any group – e.g., Distinguished Engineers

SmallBlue Find Widget in Mobile

SmallBlue Whisper Widget in Mobile

Recommend Contents from Friends within 3-degrees

Android

BlackBerry
Structural Diverse networks with abundance of structural holes are associated with higher performance.

Having diverse friends helps.

Betweenness is negatively correlated to people but highly positive correlated to projects.

Being a bridge between a lot of people is bottleneck.

Being a bridge of a lot of projects is good.

Network reach are highly corrected.

The number of people reachable in 3 steps is positively correlated with higher performance.

Having too many strong links — the same set of people one communicates frequently is negatively correlated with performance.

Perhaps frequent communication to the same person may imply redundant information exchange.

Productivity effect from network variables

- An additional person in network size ~ $986 revenue per year
- Each person that can be reached in 3 steps ~ $0.163 in revenue per month
- A link to manager ~ $1074 in revenue per month
- 1 standard deviation of network diversity (1 - constraint) ~ $758
- 1 standard deviation of btw ~ -$300K
- 1 strong link ~ $-7.9 per month
Graph Analytics for Financial Analysis

**Goal:** Injecting Network Graph Effects for Financial Analysis. Estimating company performance considering correlated companies, network properties and evolutions, causal parameter analysis, etc.

- **IBM 2003**

- **IBM 2009**

- **Data Source:**
  - Relationships among 7594 companies, data mining from NYT 1981 ~ 2009

**Targets:** 20 Fortune companies’ normalized Profits

**Goal:** Learn from previous 5 years, and predict next year

**Model:** Support Vector Regression (RBF kernel)

**Network feature:**
- s (current year network feature),
- t (temporal network feature),
- d (delta value of network feature)

**Financial feature:**
- p (historical profits and revenues)

Profit prediction by joint network and financial analysis outperforms network-only by 130% and financial-only by 33%.
Graph Analysis for Image and Video Analysis
Contributing to the Brain Activity Mapping project, later renamed to BRAIN initiative

Signal capturing up to the neuron-level resolution

Demo: neuron detection
Cognitive Networks

- Cognitive Network:
  - 30,000 nodes of dynamic brain MRI functional networks
Composite Social-Cognitive-Info Networks

- Status Monitoring
- Visualization
- Personalized Information Recommendation
- Routing, etc.

EEG / Audio signal detection

GPS / location detection, Information Display

3G

Status Monitoring, Visualization, Personalized Information Recommendation, Routing, etc.
Graph Definitions and Concepts

- A graph: 
  \[ G = (V, E) \]

- \( V = \text{Vertices or Nodes} \)

- \( E = \text{Edges or Links} \)

- The number of vertices: “Order” 
  \[ N_v = |V| \]

- The number of edges: “Size” 
  \[ N_e = |E| \]
A graph $H$ is a subgraph of another graph $G$, if:

$$V_H \subseteq V_G \quad \text{and} \quad E_H \subseteq E_G$$
Multi-Graph vs. Simple Graph

- Loops:

- Multi-Edges:
Directed Graph vs. Undirected Graph

- Directed Edges = Arcs:

\[ \{u, v\} \]

- Mutual arcs:
Adjacency

- $u$ and $v$ are adjacent if joined by an edge in $E$:

- Two edges are adjacent if joined by a common endpoint in $V$: 
Incident and Degree

- A vertex \( v \in V \) is incident on an edge \( e \in E \) if \( v \) is an endpoint of \( e \).

- The degree of a vertex \( v \), say \( d_v \), is defined as the number of edges incident on \( v \).
Degree Sequence

- The degree sequence of a graph $G$ is the sequence formed by arranging the vertex degrees $d_v$ in non-decreasing order.

\[ \{2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4\} \]

- The sum of the elements degree sequence equals to twice the number of edges in the graph (i.e. twice the size of the graph).
In-degrees and out-degrees

- For Directed graphs:

  ![Diagram 1](image1.png)
  ![Diagram 2](image2.png)

  In-degree = 8  Out-degree = 8
Walk

- A walk on a graph $G$, from $v_0$ to $v_l$, is an alternating sequence:

$$\{v_0, e_1, v_1, e_2, \ldots, v_{l-1}, e_l, v_l\}$$

- The length of this walk is $l$.

- A walk may be:
  - **Trail** --- no repeated edges
  - **Path** --- trails without repeated vertices.
Circuit, Cycle, and Acyclic

- **Circuit**: A trail for which the beginning and ending vertices are the same.

- **Cycle**: a walk of length at least three, the beginning node = ending node, all other nodes are distinct

- **Acycle**: graph contains no cycle
Reachable, Connected, Component

- **Reachable**: A vertex \( v \) in a graph \( G \) is said to be reachable from another vertex \( u \) if there exists a walk from \( u \) to \( v \).

- **Connected**: A graph is said to be connected if every vertex is reachable from every other.

- **Component**: A component of a graph is a maximally connected subgraph.
Connection in a digraph

- **Weakly connected**: If its underlying graph is connected after stripping away the direction.

- **Strongly connected**: every vertex is reachable from every other vertex by a directed walk.
Distance

- **Distance of two vertices:** The length of the shortest path between the vertices.

- **Geodesic:** another name for shortest path.

- **Diameter:** the value of the longest distance in a graph
Decorated Graph

- Weighted Edges
Families of Graphs

- **Complete Graph**: every vertex is linked to every other vertex.

- **Clique**: a complete subgraph.
Regular Graph

- **Regular Graph**: a graph in which every vertex has the same degree.

![3-regular graph]

a 3-regular graph
Tree and Forest

- **Tree**: a connected graph with no cycle.

- **Forest**: a disjoint union of trees is called a forest.
Labels in a directed tree

- **Root**
- **Ancestor**
- **Descendant**
- **Parent**
- **Children**
- **Leaf**: a vertex without children
Rooted Tree vs Directed Acyclic Graph (DAG)

- **DAG**: Directed Acyclical Graph. Underlining undirected graph has cycle.
Bipartite Graph

- **Bipartite Graph**: Vertices are partitioned into two sets. Edges link only between these two sets.

\[ \begin{align*}
&v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \\
&v_6 \quad v_7 \quad v_8
\end{align*} \]

- **Induced Graph (Collaborative Filtering)**:
Recommendation Technique – Collaborative Filtering

Customers who bought this item also bought

*Theoretical Neuroscience: Computational and Mathematical* by Dayan

*Biophysics of Computation: Information Processing in Sir Neuroscience Series* by Christof Koch

Customers who bought this item also bought

*Bayesian Statistics: An Introduction (A Hodder Arnold Pl)*

*Markov Chain Monte Carlo in Practice* by W.R. Gilks

*Monte Carlo Statistical Methods (Springer Texts in Statist*  

*Bayes and Empirical Bayes Methods for Data Analysis, See*  

*The Elements of Statistical Learning* by T. Hastie

\[
\text{item} \quad \Rightarrow \quad \text{user} \quad \Rightarrow \quad \text{Recommendation}
\]
Graphs and Matrix Algebra

- The fundamental connectivity of a graph $G$ may be captured in an $N_v \times N_v$ binary symmetric matrix $A$ with entries:

$$A_{ij} = \begin{cases} 
1, & \text{if } \{i, j\} \in E \\
0, & \text{otherwise}
\end{cases}$$

$A$ is called the Adjacency Matrix of $G$
Questions?