Workshop on Social Network Modeling in Education Research

Introduction to Social Networks
Tracy Sweet, Brian Junker, Sam Adhikari, Beau Dabbs, Andrew Thomas

http://hnm.stat.cmu.edu

This work is supported in part by IES (US Dept of ED) Award #R305D120004 and NSF Award #SES-1229271.
Our Goals for the Workshop

- Introduce you to descriptive analysis of social network data.
- Show you how statistical models can help organize and focus your analysis of social network data.
- Get you excited about studies on social networks in your research.
- Get you excited about applying CIDnetworks to your data.
- Bring you into our social network, so you can help us make CIDnetworks and the other tools we are developing better over time, with your suggestions & ideas.
Examples of Social Networks Online

- Facebook
  - Friending is a “symmetric” relationship between people
  - Posting on someone’s wall is asymmetric
  - Following is asymmetric

- Twitter
  - Following and retweeting are both asymmetric

- Classroom 2.0
Examples of Social Networks in Education Research

- **Bully Prevention**: Which interventions are effective in reducing bullying behavior? How do they affect the friendship networks themselves? (Dorothy Espelage, Univ. of Illinois, multiple school-level networks)

- **Organizational Structure**: How do org structure and individual factors shape professional advice-seeking networks? (Jim Spillane, Northwestern, 30 school networks)

- **Evolving Friendship Ties**: Friendship data was collected on 5th graders several times over a school year. What factors affect friendships and changes in friendship over time? (Rebecca Madill, Penn State, 25 networks).
The Pitts & Spillane (2009) Data

- School staffing survey given to teachers in 15 schools
  - Pre-K through 8, private and public schools
  - Does teacher i seek advice from teacher j?
  - Demographics, beliefs, and professional experience were also collected:

  For teachers:
  - Years teaching
  - Sense of trust

  For pairs of teachers (dyads):
  - Similar # of years in school?
  - Same innovative attitudes?
  - Teach same grade?

  For schools:
  - Catholic?
  - School size

An Advice Network

- Teachers are **nodes** or **vertices** in the network
- Teacher i seeks advice from teacher j iff there is an **edge** i -> j
  (edges=links=ties)
  - A **dyad** is a pair of nodes; may have an edge or not
  - Advice-seeking is **asymmetric, directed**
- The graph is a **sociogram**

**Egos vs alters**
- An “**ego**” is the teacher you are looking at right now
- The “**alters**” are his/her neighbors in the graph
An Advice Network

- The social network can also be represented as a **sociomatrix** (adjacency matrix, weight matrix)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>J</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Some basic notation

- $G = \text{a graph or network}$;
  - $V(G) = \text{its vertices (nodes)}$,
  - $E(G) = \text{its edges (ties)}$,
  - $N(G) = \#V(G)$, $K(G) = \#E(G)$.

- For $i, j \in V(G)$, let $y_{ij}$ be the indicator

$$y_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E(G) \\ 0 & \text{else} \end{cases}$$

- The adjacency matrix is $y = A(G)$.

- If the edges have weights, then $y_{ij}$ will have weights as values instead.

Descriptive analysis often emphasizes topological features, e.g.:

- **Graph Density** (fraction of total possible edges in G)
- **Node Centrality Measures:**
  - **Node degree** (how many edges go into or out of this node)
  - **Closeness**
    - \(1/(\text{sum of shortest path} \times \text{lengths to every other node})\)
  - **Betweenness**
    - Sum of fraction of shortest paths between every pair i, j passing through this node
- **Edge betweenness** – Similar to node betweenness
- **Block or community structure**
- **Other topological features** (triads/transitivity, stars, cliques...)
  - (we will mostly omit these)

* If there is no shortest path, use \#V(G)*

4/19/2015
For our little network...

- Tie density is
  \[ \frac{K}{N(N-1)} = \frac{32}{90} = 0.36 \]
- Node centrality measures:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>In deg</td>
<td>3.00</td>
<td>2.00</td>
<td>5.00</td>
<td>4.00</td>
<td>2.00</td>
<td>3.00</td>
<td>2.00</td>
<td>7.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Out deg</td>
<td>3.00</td>
<td>3.00</td>
<td>4.00</td>
<td>1.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>3.00</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Closeness</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0.10</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Betweenness</td>
<td>2.75</td>
<td>3.08</td>
<td>8.75</td>
<td>3.50</td>
<td>4.00</td>
<td>3.50</td>
<td>4.33</td>
<td>12.50</td>
<td>0.50</td>
<td>1.08</td>
</tr>
</tbody>
</table>

- Edge centrality:
  - We show it on the next slide
  - Edges or nodes with high “betweenness” might be on paths between blocks or clusters in the network...
Edge betweenness...

A -> B 3.83  F -> D 1.00
A -> G 4.83  F -> E 4.50
A -> J 3.08  F -> H 1.00
B -> A 3.00  G -> A 1.00
B -> H 5.58  G -> B 2.25
B -> I 3.50  G -> C 5.50
C -> D 3.38  G -> H 3.58
C -> F 3.38  G -> J 1.00
C -> H 1.00  H -> C 3.25
C -> I 6.00  H -> D 7.12
D -> E 8.50  H -> F 7.12
E -> C 4.00  I -> C 3.00
E -> D 1.00  I -> H 2.50
E -> F 2.00  J -> A 1.75
E -> H 2.00  J -> G 2.50
F -> C 2.00  J -> H 5.83
Edges with high edge-betweenness might be connecting communities (E-B communities)

A random walk of, say, 4 steps, should get stuck in a community (walktrap communities)
Digression to R...

- Pick one or more schools, and explore the features we have been talking about with it/them. Make some comparisons!
Models for Social Networks

- We will (mostly) skip over classic social network models
  - $P_1$ models
  - $P_2$ models
  - $P_\ast$ or “Exponential Random Graph Models” (ERGMs)

- **Instead we concentrate on scalable generative models:**
  - *Predicting ties from external, observable covariates*
  - *Using latent variables to model tie behavior that is not predicted from observable covariates*
Conditionally Independent Dyad (CID) models

- CIDnetworks models are based on a normal mixed effects regression framework:
  \[ Y_{ij}^* = X_{ij} \beta + U_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2) \]
- \( X_{ij} \)'s are edge covariates
- \( U_{ij} \) is a random effect, i.e. latent/unobserved structure
  - *Allows for clustering, block/community structure, transitivity...*
- The CIDnetworks package models
  - Continuous tie weights, as \( Y_{ij} = Y_{ij}^* \)
  - Ordinal tie weights, as \( Y_{ij} = 0, 1, 2, \ldots \) depending on interval \( Y_{ij}^* \) falls into
    (ordered probit model)
  - 0/1 ties, as \( Y_{ij} = 1_{(Y_{ij}^*>0)} \) (probit / normal ogive model)