Social Network Analysis: When Social Relationship is the Dependent Variable

Anabel Quan-Haase
Faculty of Information and Media Studies
Sociology
Overview of Presentation

- General overview of the social network approach
- Key terminology
- Uniqueness of collecting and analyzing social network data
- Recoding and importing data into STATA
- Example of analysis
- Understanding limitations and problems
- Conclusions
What is SNA?

Social network analysis is focused on uncovering the patterning of people’s interaction....Network analysts believe that how an individual lives depends in large part on how that individual is tied into the larger web of social connections. Many believe, moreover, that the success or failure of societies and organizations often depends on the patterning of their internal structure (Freeman, 1998, November 11).
Graph these case-sensitive comma-separated phrases: social network, social networks, social network analysis
between 1930 and 2008 from the corpus English with smoothing of 3.

Search lots of books

Run your own experiment! Raw data is available for download here.

What is Unique about SNA?

Social science research and theory tends to focus on social actors’:

• attributes
• attitudes
• opinions
• behavior

Focus is on individual level of analysis, less on network-structural level.
a whole is not simply the sum of its parts
Key Terminology

• 1. Social structure
• 2. Social network
• 3. Social actors
• 4. Social ties or relations
• 5. Additional terms of relevance:
  – Nodes & edges
  – Directed graphs vs. undirected graphs
  – Ego
  – Alter
  – Homophily
1. Social Structure

- Sociological inquiry consists of understanding the constraining influence of social structure on social action
- BUT; how do we study social structure?
2. Social Network

Social Structure as Social Network
3. Social Actors

- The actors considered in a social network are exclusively social (alternatively referred to as agents, nodes, or social entities).
- These include individuals, organizations, institutions, nations, or groups (Wasserman & Faust, 1994).
Blurred Social Actors

• Social actors can therefore be distinguished from non-social actors – e.g., neurons comprising a neural network.

• On occasion, the distinction between a social and a non-social actor is not absolute. For example, computer networks represent a hybrid type of network.
One-mode vs. Two-mode

• Most social network analysis methods allow only one type of social actor (for instance, individuals or corporations) in their analysis; these are referred to as one-mode networks (Wasserman & Faust, 1994).

• However, methods exist which allow two different types of social actors in their analysis; these are referred to as two-mode networks. For instance, a study may simultaneously analyze corporations and their directors.

• Two-mode networks may also include social actors from distinct networks, for example, a network comprised of adults and a network comprised of children.

• Two-mode networks allow for comparison between different types and sets of social actors.
4. Ties

- Ties are links that connect social actors, and are the main focus of social network analysis. Ties are seen as “channels for transfer or “flow” of resources (either material or nonmaterial)” (Wasserman & Faust, 1994, p. 4).
Simple Ties

• Naturally occurring ties among social actors are inherently complex and consist of numerous different interaction activities.

• However, unlike ethnographers network analysts do not focus on the complexity of interactions among individuals (Burt, 1983).

• Instead, social network analysts focus more on the pattern of relations amongst individuals and to do so simplify the inherent complexity of social relationships by categorizing interactions into different broad types. The types can be manifold. For example, a pair of social actors may have friendship, working, cooperation, or citation ties.
5. Additional Terms

• Nodes & edges
• Directed graphs vs. undirected graphs
• Ego
• Alter
• Homophily
Types of Network Analysis

• *Ego-centered/Socio-centered Social Networks*
• *Community-centered social networks*
Ego-centered/Socio-centered Social Networks
Community-Centered Social Networks

Face-to-face (1/week)
Uniqueness of Collecting and Analyzing Social Network Data

- Relational data
- Boundary specification and sampling
- Interdependence of data points
- Complexity of data collection
  - Behavioral
  - Self-report
- Key Sources
- Visualization
  - Sylva (CulturePlex)
Internet Resources of Social Network Analysis

• Center for the Study of Group Processes
  http://lime.weeg.uiowa.edu/~grpproc/

• INSNA International Network of Social Network Analysis
  http://www.heinz.cmu.edu/project/INSNA/

• Barry Wellman’s Homepage
  http://www.chass.utoronto.ca/~wellman/index.html

• NodeXL
  http://nodexl.codeplex.com/
Example


- Collaborators: Department of Sociology and Anthropology, University of Haifa, Israel.
Hypotheses

• **H1**: Perceived tie closeness is expected to be positively associated with the frequency of communication via IM and the multiplexity of communication topics.

• **H2**: Same-sex relationships are expected to communicate more frequently via IM and on a larger number of topics than mixed-sex relationships.

Methods

• The two dependent variables examined in this study are dyadic, with the unit of observation being relational.

• We defined $i$ as ego, the focal individual in the study, and $j$ as the network contact listed in the name generator: $j = 1$ or $2$. 
The Name Generator

• An efficient means to study ego-networks.
• Provides quick insight into ties.
• Allows to study several ties at one time in point.
The next couple of questions examine your relationship with instant messaging contacts.

Consider all your communication instances of the past **WEEK** that took place via instant messaging. Considering every single day of the week, please write down the names of **TWO** people who you communicated with most frequently. This list can include friends on or off campus, acquaintances, professors and family. If you communicated with strangers, you can also write them down. All these names will remain confidential.

(Modified question from Burt, 1992)

1. _____________________________________________
2. _____________________________________________
   ....
Frequency of Communication

- The first dependent variable measured the frequency of communication via IM for each pair \((i, j)\) on an eight-point scale from 1 = “never” to 8 = “several times a day”. 
Topic Multiplexity

• The second dependent variable, IM topic multiplexity, measured the number of topics discussed by each pair \((i, j)\), ranging from 0 to 11 topics. The measure of multiplexity was developed and tested in previous studies (Marsden & Campbell, 1984; Mesch & Talmud, 2006).

• Respondents were provided with a list of 11 topics of conversation from which to choose, such as school, family, and personal. Positive responses were coded 1, and a scale was created based on the sum of all the topics discussed among each pair \((i, j)\).
Independent Variables:
Age and Sex of Alter

In this section, a series of questions will be asked about each of the people you have listed above. Please indicate which response applies best.

What is the approximate age of each person listed?
- Less than 10 years
- 10-14 years
- 15-18 years
- 18-25 years
- 25-30 years
- More than 30 years

What is the sex of each person listed?
- Female
- Male
Relational Variables: Closeness & Distance

Closeness
How close do you feel to each person listed?
- Very close
- Somewhat close
- Neutral
- Somewhat distant
- Distant

Distance
Where is each person listed located?
- On campus
- Off campus
- In London
- Within 3 hours driving distance
- Within 6 hours driving distance
- Within Canada, but outside Ontario
- Within North America, but outside Canada
- Outside North America
Are We More Connected?

• A key question in the literature evolves around our levels of connectivity and the role of technology.
• Are we less connected, or are these connections of lesser quality?


Topic Multiplexity

Which topics do you discuss with each of the 6 friends in the IM list?

- Classes and/or professors in the university
- About relationship with parents
- About mutual friends
- Sports and movies
- About family problems
- About personal problems
- Favorite television programs
- Romantic relationships
- Music
- Religion/Spirituality
- Fashion and clothes
- Food and Diet
- Politics
- Places to go out
- Computers and technology
Recoding and importing data into STATA

Excel ➔ • 1. Reconfiguring data from standard social science data into relational data.

SPSS ➔ • 2. Recoding homophily variables.

STATA ➔ • 3. GLMS analysis.
Standard Data => Relational Data

Ego = i

<table>
<thead>
<tr>
<th>Case</th>
<th>Age_Ego</th>
<th>Sex_Ego</th>
<th>Age_Alter1</th>
<th>Age_Alter2</th>
<th>Sex_Alter1</th>
<th>Sex_Alter2</th>
<th>Closeness_Alter1</th>
<th>Closeness_Alter2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26</td>
<td>1</td>
<td>22</td>
<td>30</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>0</td>
<td>20</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>1</td>
<td>21</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>Age_Ego</th>
<th>Age_Alter</th>
<th>Sex_Ego</th>
<th>Sex_Alter</th>
<th>Closeness</th>
<th>Distance</th>
<th>Topic Multiplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>26</td>
<td>22</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>1b</td>
<td>26</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2a</td>
<td>19</td>
<td>20</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2b</td>
<td>19</td>
<td>23</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>3a</td>
<td>23</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>3b</td>
<td>23</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>
Homophily

Ego = i   Alter = i, j

Calculate new homophily variable for sex:
If ego and alter same sex than equal 1
If ego and alter different sex than equal 0

Independence of Observations

• Given that the dyadic relations between contact $i$ and contact $j$ are not independent observations from the relations between contacts $i$ and $k$, we utilized a clustering procedure with Robust Standard Error Estimates (RSE) to control for the dependence of the observations. In this way, we analyzed both contacts in a single model, controlling for joint variance resulting from the clustered contacts (Cameron & Trivedi, 2010; Mok, Wellman, & Carrasco, 2010; Wellman & Frank, 2001).
Clustering Variable

Ego=$i$

<table>
<thead>
<tr>
<th>case</th>
<th>age_ego</th>
<th>age_alter</th>
<th>sex_ego</th>
<th>sex_alter</th>
<th>closeness</th>
<th>distance</th>
<th>topic_multiplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>26</td>
<td>22</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>1b</td>
<td>26</td>
<td>30</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2a</td>
<td>19</td>
<td>20</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2b</td>
<td>19</td>
<td>23</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>3a</td>
<td>23</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>3b</td>
<td>23</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>
Analysis in STATA

This approach allows us to control for the interdependence of cases.

glm DV IV1 IV2,  
family(gaussian) link(identity) vce(cluster id)
### Table 1. Relationship type as a percentage of IM contacts for Israel and Canada

<table>
<thead>
<tr>
<th>Relationship type</th>
<th>First IM contact</th>
<th>Second IM contact</th>
<th>First IM contact</th>
<th>Second IM contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>11%</td>
<td>10%</td>
<td>16%</td>
<td>20%</td>
</tr>
<tr>
<td>Close friend</td>
<td>48%</td>
<td>44%</td>
<td>42%</td>
<td>49%</td>
</tr>
<tr>
<td>Distant friend</td>
<td>23%</td>
<td>35%</td>
<td>12%</td>
<td>16%</td>
</tr>
<tr>
<td>Romantic partner</td>
<td>14%</td>
<td>5%</td>
<td>21%</td>
<td>6%</td>
</tr>
<tr>
<td>Online tie</td>
<td>4%</td>
<td>9%</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

*Note. Israel N = 492; Canada N = 293*
Table 2. GLM results for the final models predicting IM topic multiplexity for Israel and Canada

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Israel</th>
<th></th>
<th>Canada</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>RSE</td>
<td>z</td>
<td>β</td>
</tr>
<tr>
<td>Age</td>
<td>-0.09*</td>
<td>.04</td>
<td>-2.25</td>
<td>-0.05</td>
</tr>
<tr>
<td>Gender (1=male)</td>
<td>-0.11*</td>
<td>.25</td>
<td>-2.44</td>
<td>-0.08</td>
</tr>
<tr>
<td>Relationship status (1=single)</td>
<td>0.0</td>
<td>.25</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Gender similarity (1=same)</td>
<td>-0.09*</td>
<td>.21</td>
<td>-2.22</td>
<td>0.05</td>
</tr>
<tr>
<td>Propinquity (1=nearby)</td>
<td>0.03</td>
<td>.22</td>
<td>0.82</td>
<td>-0.03</td>
</tr>
<tr>
<td>Hours of daily IM use</td>
<td>0.15**</td>
<td>.15</td>
<td>2.72</td>
<td>0.14**</td>
</tr>
<tr>
<td>Number of IM regular contacts</td>
<td>0.12*</td>
<td>.14</td>
<td>2.47</td>
<td>0.09*</td>
</tr>
<tr>
<td>Online relationship</td>
<td>-0.08</td>
<td>.68</td>
<td>-1.74</td>
<td>-0.04</td>
</tr>
<tr>
<td>Family</td>
<td>-0.18***</td>
<td>.35</td>
<td>-4.64</td>
<td>-0.21***</td>
</tr>
<tr>
<td>Distant friend</td>
<td>-0.14**</td>
<td>.27</td>
<td>-3.26</td>
<td>-0.09</td>
</tr>
<tr>
<td>Partner</td>
<td>0.01</td>
<td>.39</td>
<td>0.14</td>
<td>0.19***</td>
</tr>
<tr>
<td>Close friend</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>0.13*</td>
</tr>
<tr>
<td>Duration of relationship</td>
<td>0.04</td>
<td>.11</td>
<td>0.92</td>
<td>0.22***</td>
</tr>
<tr>
<td>Closeness</td>
<td>0.24**</td>
<td>.17</td>
<td>5.60</td>
<td>0.22***</td>
</tr>
<tr>
<td>Constant</td>
<td>2.0</td>
<td>1.23</td>
<td>1.62</td>
<td>6.33</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>.23***</td>
<td>.23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *** $p < .001$, ** $p < .01$, * $p < .05$, two-tailed.

Israel $F(13, 633) = 16.17$; Canada $F(13, 513) = 13.81$.

Israel $N = 647$; Canada $N = 526$.
Other Approaches


Limitations of Social Network Analysis

- Boundary specification
- Data source
- Definition of social actors
- No distinct method
Conclusions

• Powerful for research questions that focus on the relationship and the flow of resources.

• Statistical software does not yet accommodate all forms of social network analysis.

• A number of considerations are necessary:
  – boundary specification;
  – interdependence;
  – specifications of other statistical methods.
Thank you!

“With enough effort and perseverance: Anything is possible”