

Network Data

Monday, August 29, 2022 12:33



Advanced Network Analysis 1. Network Data

Weihua An



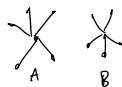
Outline

- ▶ Study Design
 - ▶ Type of networks
 - ▶ Type of relationships
- ▶ Network Data Collection
 - ▶ Surveys
 - ▶ Observations
 - ▶ Other methods
- ▶ Network Data Representation
 - ▶ Adjacent matrix
 - ▶ Edge list
- ▶ Basic Network Analysis

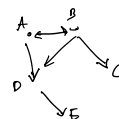
Study Design

- ▶ Types of networks
 - ▶ Egocentric networks
 - ▶ Sociocentric networks
 - ▶ Cognitive networks
- ▶ Type of relationships
 - ▶ Instrumental: advising, collaborator
 - ▶ Affective: friendship, romantic
 - ▶ Multiplex: a mixture of different types of relationships

Egocentric networks
Ask A & B their friends,
know only local network



Sociocentric Networks
Ask each person in class
to nominate their best
friend



Cognitive Networks
Each person has their own
report of the network
Mesh all the versions together



We mainly focus
on this one

check out AddHealth dataset

Network Data Collection

- ▶ Survey questions
 - ▶ Name generator: ask for the names of contacts
 - ▶ Name interpreter: provide information on contacts
 - ▶ Position generator: name contacts with particular positions
- ▶ Observation
- ▶ Other methods
 - ▶ Internet: Facebook, Twitter, Youtube, etc.
 - ▶ Biographies

Name Generator (1985 and 2004 GSS)

"From time to time, most people discuss important matters with other people. Looking back over the last six months who are the people with whom you discussed matters important to you? Just tell me their first names or initials. IF LESS THAN 5 NAMES MENTIONED, PROBE: Anyone else?"

Name Generator and Interpreter (An 2011)

AK. Please list your closest friends in this school (including your own class). (Up to ten)

Index	Name	Code (Check)	Class Index	Strength of Friendship (1 to 10, from least strong to strongest)	Distance between two of your households (meters)	Whether smoke?
						1 Yes 2 No 9 Don't know
AKa						
AKb						
AKc						
AKd						
AKe						
AKf						
AKg						
AKh						
AKi						
AKj						

Position Generator (Lin 1999)

Table 2 Position generator for measuring accessed social capital: an example

Here is a list of jobs (show card). Would you please tell me if you happen to know someone (on a first-name basis) having each job?

Job 1. Do you know anyone having this job? 2. How long have you known this person (no. of years)? 3. What is your relationship with this person? 4. How close are you with this person? 5. His/her gender 6. His/her job

Job A _____

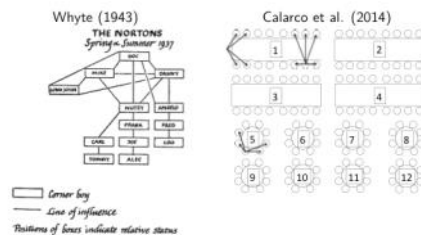
Job B _____

Job C _____

etc. _____

*If you know more than one person, think of the one person whom you have known the longest.

Observations



Possible Problems in Network Data Collection

- Problems (Marsden 2005; Handcock and Gile 2007)
 - Recall error: forgetting friends (20%) (Brewer and Webster 1999)
 - Respondent burden (usually capped at 10 in a fixed choice design)
 - Clustering bias: tend to name people in clustering and not necessarily in order of social proximity
- Consequences
 - May affect the structural properties of the social networks, such as density, number of cliques, centralization, and centralities (Brewer and Webster 1999)
 - Discordant reports (An and Schramski 2015)

Network Data

How do we code the data?

Friends



Adjacency matrix

	A	B	C
A	0	1	0
B	1	0	0
C	1	0	0

Edge list

A	B
A	B
B	A
C	A

C nominates A

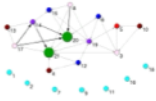
Datasets

★ Final Project Data

- The UCI Network Data Repository
- Linton Freeman's Data Collection
- UNICNET Datasets
- Stanford Large Network Dataset Collection
- Search "network" at AER, APSR, or PNAS
- The General Social Survey
- National Longitudinal Study of Adolescent Health

Basic Network Analysis

A Co-Advising Network



- Node Level
 - Indegree, outdegree, closeness, betweenness, and eigenvector centrality
- Dyad Level → 2 people
 - Reciprocity, geodistance
- Group Level
 - Triad, cliques, component
- Network Level
 - Density, centralization, transitivity

Mathematical Definitions

- Degree centrality: the number of nodes a focal node is connected to.

$$D_i = \sum_j w_{ij}$$

where $w_{ij} = 1$ if there is a tie between i and j ; otherwise $w_{ij} = 0$.

- Closeness centrality: the speed a node can reach to all other nodes.

$$C_i = \left[\sum_j d_{ij} \right]^{-1}$$

where d_{ij} is the geodistance (i.e., the shortest path) between i and j .

- Betweenness centrality: the number of shortest paths through a node.

$$B_i = \sum_{j,k} \frac{g_{jk}}{g_{jk}^i}$$

where g_{jk} is the number of shortest paths between j and k .

- Eigenvector centrality: the extent of connecting to important alters.

$$E_i = \frac{1}{\lambda} \sum_j w_{ij} E_j$$

$$\lambda E = ME$$

where M is the adjacency matrix and λ the eigenvalue.

Mathematical Definitions

- Density: the proportion of observed ties out of all possible ties.

$$D(G) = \frac{\sum_{i,j} w_{ij}}{n(n-1)}$$

- Centralization: the extent to which ties are concentrated. It is bounded between 0 and 1 with larger values indicating more centralization.

$$C(G) = \sum_i \left| \frac{M - C_i}{(n-2)(n-1)} \right|$$

where C_i is the degree of node i , M is the maximum among C_i .

- Transitivity: the proportion of transitive triangles out of possible triangles.

$$T(G) = \frac{\sum_{i,j,k} w_{ij} w_{jk} w_{ik}}{\sum_{i,j,k} w_{ij} w_{jk}}$$