Rise in popularity of network research

Number of articles on social networks indexed by Google Scholar
Lots of applied interest

- Health sciences
  - Epidemiology and patient support

- Management consulting companies
  - Boston Consulting Group (BCG)
  - Booz Allen Hamilton
  - McKinsey (through ex-student Rob Cross)
  - Arthur Andersen
  - CFAR (specialists in hospitals)

- Other companies
  - Merck, Pfizer, Novartis
  - BankBoston
  - Towers Perrin
  - Price Waterhouse (forensics & change)

- US govt
  - JWAC, US Army HTS, DTRA, NSA (both blue and red team work)
  - Civilian management
It’s not exactly new

1700s- Euler
1930s- Moreno’s Sociometry
  Hawthorne studies
1940s Psychologists
  Clique formally defined
1950s & 60s Anthropologists
  Kinship analysis; society as network
1970s Rise of Sociologists
  Small Worlds, Strength of weak ties; Social Networks;
  INSNA;
  Sunbelt conference
1980s IBM computation
  Computer programs developed
1990s Multi-disciplinary diffusion
  Spread of network analysis to multiple fields; Social
  capital & embeddedness in vogue
2000s Physicists’ “new science”
  Scale free, small worlds, etc.

Number of articles on social networks indexed by Google Scholar
Why do we care?
Nature organizes itself as networks

“network science”

Aspartame sweetener

Protein reactions

Neurons in the brain

Food web
... and so do humans

An effective way of organizing

Organization chart

Plumbing system

Road system

Trade relations among nations

Regular permutation group arising from Arunta marriage system.
Networks are everywhere

*or are they just a way of seeing the world?

• A molecule is a network of atoms
• A brain is a network of neurons
• A body contains many networks, including the circulatory system
• Genes form regulatory networks that turn other genes on and off
• Firms are networks of individuals, passing along information, orders and coordinating efforts
• Buildings contain many networks, including heating/cooling, plumbing, electrical
• Economies are networks of firms and other agents buying and selling
• Countries contain many networks, e.g., transportation systems, phone systems
• The internet is a network
• Ecosystems are networks of species eating each other, creating environments for each other, etc.
Characterizing SNA

Characterizing network theorizing
Contextual

• Importance of an individual’s environment
  • To explain individual outcomes, must take into account the node’s social environment in addition to internal characteristics
  • In SNA, the environment is conceptualized as network
  • An emphasis on structure relative to agency
  • Consistent with an open systems perspective
• The contrast is with an essentialist/dispositional perspective
  • Predict individual’s outcomes using other characteristics of the individual
  • Employee’s success a function of ability and motivation
Relational

• Traditionally, social science has focus on attributes of individuals to predict individual outcomes
  • Income as a function of education
• SNA puts the focus on relationships between individuals

Variables (attributes)

<table>
<thead>
<tr>
<th>Age</th>
<th>Sex</th>
<th>Education</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cases (entities)

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Canonical data matrices
• Person by attribute
• Person by person
Structural

• It’s not just relational (ties) but structural (pattern of ties)
  • To understand function, need to know more than list of elements. It’s how they are connected
  • Non-reductionist, emergent flavor

• Indirect effects propagate
  • Power grids
Positional

• A node’s position in a network determines in part the opportunities and constraints that it will face
  • Risk of news, risk of infection
  • Sense of identity
  • Individual social capital

• **Backcloth / traffic distinction**
  • Social ties provide conduits along which traffic can flow
  • A node’s position in the network has significant implications for ...
    • How early it encounters something flowing
    • How frequently it receives what is flowing
    • With what certainty it is reached
Summarizing the network perspective

**Key Dimensions**

- **Contextual**
  - It’s the environment, stupid!

- **Relational**
  - By environment, we mean ties to others

- **Structural**
  - It’s a network
  - Concepts and metrics for characterizing the network

- **Positional**
  - Location, location, location

**The Flow Model**

- Network theory largely ...
  - Regards flows as the key mechanism underlying outcomes
  - Assumes the data we collect are about the roads that enable flows
  - Most of the conceptual machinery (e.g., centrality measures) is about calculating expected flows given the network structure and given some assumptions about how things flow
Network research has sought to explain ...

• Homogeneity
  • Why people have similar beliefs, behaviors, and belongings
  • Generic network explanation: contagion, diffusion, interpersonal influence processes
    • Contagion of obesity, happiness, etc
    • Diffusion of innovations
    • Spread of disease
    • Fads and fashion
    • Social conformity

• Achievement and reward
  • Why some people are more successful than others
  • Generic network explanation: social capital
    • Ties provide access to resources
    • Certain positions in social structures can be exploited for gain
Big Idea #1 -- Contagion

• Individuals influence each other
  • Infect each other with diseases, ideas, behaviors

• As a result, we observe network autocorrelation – the tendency for adjacent nodes to have similar characteristics such as opinions, ways of dressing, food preferences

• Flows of information, money

• The case of AIDS

Discovery of HIV: Sexual contacts among gay men w/ unusual cancers, traced by Bill Darrow of the CDC
Big Idea #2 – Social capital

• Why are some individuals more successful than others?
  • Attributes such as intelligence, motivation
    • Human capital
  • Who they know, who they owe
    • Social capital
  • Social ties provide access to resources the individual doesn’t own/control directly
Types of network research
Antecedents and consequences

Mainstream Network Research

• Antecedents
  • Social processes that give rise to social ties, interactions, exchanges
    • And higher level constructs like popularity or network structure
  • Theory of networks

• Consequences
  • Mechanisms that translate ties into outcomes
    • Not just ties but network position and network structure
  • Network theory

Cognitive SNA

• Antecedents
  • How ties & network structures are perceived by 3rd parties

• Consequences
  • Consequences of these perceptions
  • E.g., Being perceived to be friends with a high status other affects judgments of your influence (even more than actual friendship with high status other)
Levels of analysis -- Organized by most to least number of units

• Dyad level – $O(n^2)$
  • Units are pairs of persons
  • Variables are things like presence of absence of a certain kind of tie between each pair of persons in network

• Node level – $O(n)$
  • Units are persons
  • Variables are things like the number of friends each person has

• Group/network level – $O(1)$
  • Units are whole networks (e.g., teams, firms or countries)
  • Variables are things like the density of trust ties, or the average number of degrees of separation between members of the group
## Types of studies

<table>
<thead>
<tr>
<th></th>
<th>Dyad Level</th>
<th>Node Level</th>
<th>Group Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theory of Networks (Antecedents)</strong></td>
<td>Understanding who becomes friends with whom</td>
<td>Explaining why some people are more liked than others</td>
<td>Explaining why some groups have more centralized network structures</td>
</tr>
<tr>
<td><strong>Network Theory (Consequences)</strong></td>
<td>Predicting similarity of opinion as a function of friendship</td>
<td>Explaining why some employees rise through the ranks faster than others as a function of social ties</td>
<td>Predicting team performance as a function of structure of trust network within team</td>
</tr>
</tbody>
</table>
Research designs
Whole network or sociocentric design

• Start with a set of people (typically a “natural” group such as a gang or a department)

• Collect data on the presence/absence (or strength) of ties of various kinds among all pairs of members of the set
  • Who doesn’t like whom; How frequently each pair of persons have a conversation
  • Typically collected via survey: respondent presented with roster of people to select/rate

• Issues
  • The set of persons needs to be some kind of census – can’t randomly pick sample of 100 persons from the population of all Americans
  • The set can’t be too big
  • Problems with inferential validity – how to generalize results?
Personal network or egocentric design

• Select random sample of respondents/subjects
  • Call them egos
• For each subject, identify the set of persons in that subject’s life
  • Call them alters
• For each alter, determine their individual characteristics
  • E.g., ask ego how old the alter is, whether they use drugs, etc.
• For each alter, determine the nature of the relationship with ego
  • E.g., ask ego how often they talk to alter, whether alter is a neighbor, etc.
• For pairs alters, determine their relationships to each other
  • E.g., ask ego whether alter 1 is friends with alter 2, etc.
Cognitive social structures (CSS) design

• A blend of whole network and personal network designs
• Start with natural group of persons as in whole network design
• Ask each person to indicate not only their own relationship with each other person, but also their perception of the relationships among all pairs of persons
• Result is a perceived network from each member of the network
• Issues
  • Tedious for the respondent – can only be used with small groups
  • Extremely rich data. Can calculate accuracy of each person’s perceptions. Study effects of social perceptions
## Design comparisons

<table>
<thead>
<tr>
<th>Full</th>
<th>Personal</th>
<th>Cognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can compute all of the stats you can compute with personal design, plus more</td>
<td>Can use random samples and standard statistics to study large populations</td>
<td>Can do everything you can do with full network</td>
</tr>
<tr>
<td>Compute global network measures like centrality</td>
<td>Can characterize node’s network neighbor, e.g. demographic composition of friends</td>
<td>Can study perception of networks and how this impacts ego outcomes</td>
</tr>
<tr>
<td>Introduces significant challenges for statistical significance due to autocorrelation</td>
<td>Respondents (and alters) can be anonymous</td>
<td>If survey-based, very tedious data collection -- requires small networks.</td>
</tr>
<tr>
<td>Tie data can be richer than in Full because of few names and anonymity</td>
<td>Alter data is from ego’s pov</td>
<td></td>
</tr>
</tbody>
</table>

25 April 2016 (c) 2015 Stephen P Borgatti
Relational states

- Relational states include ...
  - Co-ties such as co-members of P&T
  - Kinship ties like son of
  - Other role-based ties like boss of, student of, friend of

- Relational states have an always-on character
  - While they hold, they hold continuously

- Relational states are things you are
  - I am my father’s son
Relational events

• Relational events include ...
  • Have meeting with, send email to, ask question of
  • sex with, inject with, shake hands with
  • Transactions, e.g., a sale

• Relational events are discrete and transitory
  • They happen, then they are gone

• Relational events are things you count up, not things you are
  • # of lunches together vs son of

• Old definitions of networks sound like events, as in “recurring patterns of relations”
  • but include relations like friendship as examples
Building networks on events versus states

- Co-authorship network from Moody (2002)
  - Line shown between two nodes if they publish a paper together
- Note connectedness of network, ability for information to flow everyone to everyone
- But some node located more advantageously than others
  - Structural holes; centrality

Courtesy Jim Moody  
25 April 2016 (c) 2016 Stephen P. Borgatti
The network is misleading

• Can a broker play people off each other if the ties to them are at different times?

• Can information flow along paths in which links occur at different times?

• The cumulative network we had drawn was fictional
Flows

- What is transmitted, adopted, transferred, copied etc. when entities interact
  - Disease
  - Information
  - Goods
- Flows are consequence of interaction
- Flows are theoretically crucial, but often implied rather than measured
Relations among types of “tie”

• Phenomena to the left carry or enable the phenomena to the right
  • Backcloth that enables/constrains traffic (Atkin, 1972)

• Interactions can be enabled/obligated by social relations
  • Mediate effect of relations on what happens to nodes as result of receiving flows

• But causality also runs other way
  • A flow can affect relationship
Basic Concepts

Statistical Horizons ● Social Network Analysis
Steve Borgatti ● LINKS Center ● University of Kentucky

http://tinyurl.com/statisticalhorizons2016
Graph theoretic representation

- A network is represented by a graph (a mathematical object, not a picture)
- A graph $G(V,E)$ consists of a set of nodes $V$ and a set of edges (ties) $E$
  - If the ties are directed, we often call the edges “arcs”
- If $u$ and $v$ are nodes in $V$, we can indicate the presence of a tie one of two ways:
  - $u \rightarrow v$ (if directed) or $u - v$ (if undirected) $(u,v) \in E$
    - If the network is undirected then writing $(u,v) \in E$ is the same as $(v,u) \in E$, and we can also write $u - v$ to indicate the tie
- Note that in a directed graph, when $(u,v) \in E$ and $(v,u) \in E$, then we say the $(u,v)$ tie is reciprocated
  - If all ties are reciprocated, the graph is still directed, but empirically indistinguishable from an undirected graph
Matrix representation

• A network is represented by an adjacency matrix $X$
• $X$ is (often) a square matrix in which both rows and columns refer to nodes, and $x_{ij} = 1$ (or $x(i,j) = 1$) indicates the presence of a tie from $i$ to $j$ and $x_{ij} = 0$ indicates no tie was reported from $i$ to $j$
  • The row person does it to the column person. E.g., if $A$ represents Liking, then $x(b,a) = 1$ means that $b$ likes $a$
• If we are being careful, we distinguish $x_{ij} = 0$ (which indicates the absence of a tie) from $x_{ij} =$ missing value (which means we don’t know if there is a tie or not)
• If the network is directed, then the matrix need not be symmetric, meaning that $x_{ij}$ may not equal $x_{ji}$ for all $i$ and $j$

\[
\begin{array}{cccc}
  a & b & c & d \\
  a & 1 & 0 & 0 \\
  b & 1 & 1 & 0 \\
  c & 0 & 1 & 1 \\
  d & 1 & 0 & 1 \\
\end{array}
\]
Graph traversals

• Moving from node to node across links, what kinds of sequences are there?
  • Walks
    • Unrestricted traversals. At any node can go anywhere, including retracing your steps exactly as you’ve done before
  • Trails
    • You can’t retrace edges (links), but you can revisit nodes
  • Paths
    • You can’t revisit nodes
  • Every path is a trail and every trail is a walk
## How do things move?

<table>
<thead>
<tr>
<th></th>
<th>Paths</th>
<th>Trails</th>
<th>Walks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move</td>
<td>Snail mail</td>
<td>Used paperback</td>
<td>Dollar bill</td>
</tr>
<tr>
<td>Copy</td>
<td>Virus</td>
<td>Gossip</td>
<td>Emotion</td>
</tr>
</tbody>
</table>

- Path – can’t revisit a node or a line
- Trail – can’t revisit a line
- Walk – unrestricted
- Every path is a trail, every trail is a walk
Length & distance

- Length of a path (or any walk) is the number of links it has
- A shortest path to from A to B is called a geodesic
- The geodesic distance (aka graph-theoretic distance) between two nodes is the length of the shortest path
- Diameter of a network is longest geodesic distance

Distance from 5 to 8 is 2, because the shortest path (5-1-8) has two links
Geodesic Distance Matrix

Average distance or characteristic path length (CPL) is a measure of cohesiveness
Component

• Maximal sets of nodes in which every node can reach every other by some path (no matter how long)
  • Coherent fragments of a graph
• A graph with a single component is called a connected graph
• Weak vs strong components
  • A weak component is where we ignore the direction of the arcs

It is relations (types of tie) that define different networks, not components. A network that has two components remains one (disconnected) network.
A network with 4 weak components

Who you go to so that you can say ‘I ran it by ____, and she says ...’

Data drawn from Cross, Borgatti & Parker 2001.
1 weak component, 4 strong components
Cutpoints

• Nodes which, if deleted, would increase the number of components in the network
  • Removing Biff would disconnect the network (create 2 components)
Bridge

• An edge which, if removed, would increase the number of components in the network
Local Bridge of Degree K

• An edge that connects nodes that would otherwise be a minimum of $k$ steps apart
  • The A—B tie is local bridge of degree 5
• Loss of relationship between A and B would effectively, though not actually, disconnect A from B
A bipartite network
Bipartite network represented as matrix

<table>
<thead>
<tr>
<th>Names of Participants of Group I</th>
<th>Code Numbers and Dates of Social Events Reported in Old City Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mrs. Evelyn Jefferson</td>
<td>X</td>
</tr>
<tr>
<td>2. Miss Laura Mandeville</td>
<td>X</td>
</tr>
<tr>
<td>3. Miss Theresa Anderson</td>
<td>X</td>
</tr>
<tr>
<td>4. Miss Brenda Rogers</td>
<td>X</td>
</tr>
<tr>
<td>5. Miss Charlotte McDowd</td>
<td>X</td>
</tr>
<tr>
<td>6. Miss Frances Anderson</td>
<td>X</td>
</tr>
<tr>
<td>7. Miss Eleanor Nye</td>
<td>X</td>
</tr>
<tr>
<td>8. Miss Pearl Oglethorpe</td>
<td>X</td>
</tr>
<tr>
<td>10. Miss Verne Sanderson</td>
<td>X</td>
</tr>
<tr>
<td>11. Miss Myra Liddell</td>
<td>X</td>
</tr>
<tr>
<td>12. Miss Katherine Rogers</td>
<td>X</td>
</tr>
<tr>
<td>15. Mrs. Helen Lloyd</td>
<td>X</td>
</tr>
<tr>
<td>16. Mrs. Dorothy Murchison</td>
<td>X</td>
</tr>
<tr>
<td>17. Mrs. Olivia Carleton</td>
<td>X</td>
</tr>
<tr>
<td>18. Mrs. Flore Price</td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 1. Davis, Gardner and Gardner (1941) *Deep South* women-by-events matrix.
Family of Theoretical Constructs

**Group level** (properties of groups / networks)

- Cohesion
  - Density
  - Average path length
  - Fragmentation

**Node level** (properties of nodes)

- Centrality
  - Degree
  - Closeness
  - Betweenness

**Dyad level** (properties of dyads)

- Proximity
  - Adjacency
  - Simmelian tie
  - Geodesic distance

- Equivalence
  - Structural equivalence
  - Regular equivalence

- Structural Role identification

- Subgroup identification

- Shape
  - Scale-freeness
  - Core peripheriness

- Small-worldness
  - Density
  - Average path length
  - Fragmentation

- Proximity
  - Adjacency
  - Simmelian tie
  - Geodesic distance

- Equivalence
  - Structural equivalence
  - Regular equivalence

25 April 2016
Notes on Network Surveys

Statistical Horizons ● Social Network Analysis
Steve Borgatti ● LINKS Center ● University of Kentucky

http://tinyurl.com/statisticalhorizons2016
Key issues

• Whole network designs need good response rate – say, 90%
• We want truthful data
• As a result ...
  • Careful attention to questionnaire design
    • Length, question wording, attractiveness
  • Build trust & inspire interest
  • If you want to use network analysis ever again, handle the data ethically and carefully
Roster vs Write-in

Roster method (closed-ended)
• Boundaries are known and all actors listed
• Becomes cumbersome as networks grow in size
• Fewer concerns about respondent recall and accuracy
• Each actor has approximately an equal chance of being selected

Write-in method (open-ended)
• More subject to recall error
• Can use a fixed choice method limiting the number of actors elicited
• Each actor in the network does not have an equal chance of being chosen given recall and freelisting issues
• Can make getting valued ties more complicated
• Better for face-to-face interviews where probing can be used
Serial vs parallel

• Serial
  • Focuses attention on the tie
  • Tends to keep definition of “friend” the same across all alters

• Parallel
  • Focus on the alter
  • One question can affect the other
Binary or valued?

**Binary**
- Easy
- Fast
- Limited discrimination
- Lets respondents make own decisions about cutoffs

**Valued**
- Cognitively difficult – results may not be meaningful
- Tiring
- Very slow
- More nuanced, if accurate
- Some network procedures can’t handle valued data
## Asking frequencies or amounts

<table>
<thead>
<tr>
<th>Absolute rating</th>
<th>Relative ranking</th>
<th>Sequential choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>“How often do you talk to each person, on average?”</td>
<td>“How often do you speak to each person on the list below?”</td>
<td>1. Who do you talk to at least once every few months? (check all that apply)</td>
</tr>
<tr>
<td>1. Once a year or less</td>
<td>1. Very infrequently</td>
<td>2. Who do you talk to at least once every few weeks?</td>
</tr>
<tr>
<td>2. Every few months</td>
<td>2. Somewhat infrequently</td>
<td>3. Who do you talk to at least once a week?</td>
</tr>
<tr>
<td>3. Every few weeks</td>
<td>3. About average</td>
<td>4. Who do you talk to every day?</td>
</tr>
<tr>
<td>4. Once a week</td>
<td>4. Somewhat frequently</td>
<td></td>
</tr>
<tr>
<td>5. Every day</td>
<td>5. Very frequently</td>
<td></td>
</tr>
</tbody>
</table>

- Need to do pre-testing to determine appropriate time scale
- Danger of getting no variance
- Assumes a lot from resps

- Requires less of respondents; easier task
- Is automatically normalized within respondent
  - Removes response set issues
  - Makes it hard to compare values across respondents (in different rows of data matrix)

- Same data as absolute rating
  - less tiring for respondent
  - But questionnaire may look longer
- With online surveys, can pipe responses so that respondent only sees names checked off in previous question
- final question will have few names to react to
Testing Hypotheses

Statistical Horizons ● Social Network Analysis
Steve Borgatti ● LINKS Center ● University of Kentucky

http://tinyurl.com/statisticalhorizons2016
Classic example

• Centrality $\rightarrow$ performance, rewards, knowledge
  • Ordinary regression $\text{Perf} = b_0 + b_1 \cdot \text{Cent} + b_2 \cdot \text{Exper} + b_3 \cdot \text{Educ}$

• Centrality $\rightarrow$ disease
  • Disease is 0/1 var
  • Logistic regression so that $\log(b_1)$ gives increase in odds of having disease as a function of one unit increase in centrality

• Indegree in support net $\rightarrow$ quality of life
Canonical hypotheses of SNA

• Position in network determines in part the opportunities and constraints an actor will face
  • Social capital

• Actors are influenced in their views/behaviors by those they are connected to
  • Smoking as a function of number of friends who are smokers (Siena)
  • Contagion/Diffusion/Social Influence
Testing

• Normally, we use ordinary regression tools to test these kinds of hypotheses

• Data example
  • Predicting performance (supervisor rating) from network position
  • Use Stata for the regression
Statistical issues

• Sampling – non-random, non-samples
• Distributions – non-normal error distribution
• Autocorrelation – correlated error terms, non-independence of observations
Logic of a permutation test

• Given an outcome (e.g., high correlation between X and Y), calculate how likely it would be to occur if X and Y were independent

• Can enumerate all the ways an experiment could come out
  • Suppose we have four wines: Argentine, Australian, French, and Napa.
  • Alleged wine expert is supposed to identify which is which. After tasting, says bottle 1 is Argentina, b2 is Australia, b3 is Napa and b4 is France.
    • 50% correct.
  • Suppose the expert has no expertise: his choices are actually independent of actual origin. So how likely is it to get 50% correct in that case?
    • Solution is to write down all possible ways assignments of bottles to origins.
All Permutations

- 29% of permutations have at least 2 correct
- Pretty good chance of getting 2 right even if you know nothing about wine

<table>
<thead>
<tr>
<th># Correct</th>
<th>Freq</th>
<th>Prop</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
<td>0.38</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>0.33</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Total:</td>
<td>24</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note that it is impossible to get 3 out of 4 right.
Large sample permutation test

• Calculate observed statistic (e.g., corr(X,Y) or difference in means)

• Repeat 10,000 times:
  • Randomly permute values of one variable relative to the others
    • We know these values are independent of the other variable
  • Calculate statistic and record whether it was greater than or equal to the observed

• P-value is proportion of times the statistic was greater than or equal to the observed

<table>
<thead>
<tr>
<th># of Items</th>
<th># of Perms</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
</tr>
<tr>
<td>10</td>
<td>3,628,800</td>
</tr>
<tr>
<td>20</td>
<td>2.43E+18</td>
</tr>
<tr>
<td>40</td>
<td>8.16E+47</td>
</tr>
</tbody>
</table>
Antecedent or consequence?

• Centrality $\rightarrow$ performance
  • Looking at the consequences of the network variable
    • What are the benefits of social capital?

• Personality $\rightarrow$ Centrality
  • Looking at the antecedents of network variables
    • How do you get to be central?

• Statistically, these are indistinguishable
Network variables as antecedents or consequences

- Network var $\rightarrow$ non-network var
  - Centrality $\rightarrow$ performance
- Non-network var $\rightarrow$ network var
  - Extraversion $\rightarrow$ Centrality
- Network var $\rightarrow$ Network var
  - Degree centrality $\rightarrow$ betweenness centrality
  - (largely of methodological interest)
- Non-network var $\rightarrow$ non-network var
  - Age $\rightarrow$ protective behaviors
  - Mainstream, non-network research

<table>
<thead>
<tr>
<th>Indep Var (X)</th>
<th>Dependent Var (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Var</td>
<td>Deg $\rightarrow$ Bet</td>
</tr>
<tr>
<td>Non-Network Var</td>
<td>Personality $\rightarrow$ Cent</td>
</tr>
</tbody>
</table>
Levels of analysis

• Dyad level
  • Cases are pairs of nodes, vars are characteristics of the relationship between pair
  • Friendship, proximity of homes

• Node level
  • Cases are nodes, vars are node attributes
  • Centrality, gender

• Group/network level
  • Cases are entire networks, vars are properties of the group as a whole
  • Cohesion, heterogeneity

• Mixed level Node-Dyad
  • Gender (node-level attribute) predicts Friendship (dyad-level)
  • To test, we convert node-level attribute to dyad level
    • For each pair, 1 = same gender, 0 = different gender
## Types of network hypotheses

<table>
<thead>
<tr>
<th>Level</th>
<th>Example</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node level</td>
<td>• Centrality $\rightarrow$ knowledge</td>
<td>• Consequences of network</td>
</tr>
<tr>
<td></td>
<td>• Extraversion $\rightarrow$ Centrality</td>
<td>• Antecedents of network</td>
</tr>
<tr>
<td></td>
<td>• Degree centrality $\rightarrow$ Betweenness centrality</td>
<td>• Network-network</td>
</tr>
<tr>
<td></td>
<td>• Extraversion $\rightarrow$ Extraversion</td>
<td></td>
</tr>
<tr>
<td>Group/Network level</td>
<td>• Centralization $\rightarrow$ team performance</td>
<td>• Consequences of network</td>
</tr>
<tr>
<td></td>
<td>• Heterogeneity in individual abilities $\rightarrow$ centralization</td>
<td>• Antecedents of network</td>
</tr>
<tr>
<td></td>
<td>of team</td>
<td>• Network-network</td>
</tr>
<tr>
<td></td>
<td>• Density $\rightarrow$ centralization</td>
<td></td>
</tr>
<tr>
<td>Dyad level</td>
<td>• Friendship ties $\rightarrow$ similarity of views on emerging issue</td>
<td>• Consequences of network</td>
</tr>
<tr>
<td></td>
<td>• Influence/contagion/diffusion</td>
<td>• Antecedents of network</td>
</tr>
<tr>
<td></td>
<td>• Sameness of gender $\rightarrow$ interaction at conferences</td>
<td>• Network-network</td>
</tr>
<tr>
<td></td>
<td>• homophily</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Friendship ties $\rightarrow$ advice ties</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Multiplexity (one tie leads to another)</td>
<td></td>
</tr>
</tbody>
</table>
Group-level hypothesis

• Bavelas-Leavitt experiments
• Statistically uncomplicated – use regular statistics
Dyad-level hypotheses

• Allen (1977) physical proximity $\rightarrow$ amount of communication

<table>
<thead>
<tr>
<th>Proximity</th>
<th>Jim</th>
<th>Jill</th>
<th>Jen</th>
<th>Joe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jim</td>
<td>-3</td>
<td>9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Jill</td>
<td>3</td>
<td>-1</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Jen</td>
<td>9</td>
<td>1</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Joe</td>
<td>2</td>
<td>15</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communication</th>
<th>Jim</th>
<th>Jill</th>
<th>Jen</th>
<th>Joe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jim</td>
<td>-9</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Jill</td>
<td>9</td>
<td>-8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Jen</td>
<td>1</td>
<td>8</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Joe</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dyad</th>
<th>Prox</th>
<th>Comm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jim - Jill</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Jim - Jen</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Jim - Joe</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Jill - Jen</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Jill - Joe</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>Jen - Joe</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

• How to correlate two matrices?
  • String each matrix out as a long column with $n(n-1)/2$ entries each
  • Can’t just permute one column against the other – triadic autocorrelation
Triadic autocorrelation

- Affective ties, as well similarities and proximities tend to have transitivity effects
  - \( \text{dist}(\text{jill}, \text{jim}) \) cannot be larger than \( p(\text{jill}, \text{joe}) + p(\text{joe}, \text{jim}) \)
  - Creates dependencies among triples of cells
Solution

• Re-order rows and corresponding columns of the matrices in order to produce new variables that have same constraints as real variables and are necessarily independent

• Call this approach QAP correlation (and QAP regression, etc)
  • Correlate matrices (this is the observed test statistic)
  • Permute rows/cols of one matrix. Re-correlate. Repeat 10,000 times
  • P-value is the proportion of correlations that are as large as the observed

<table>
<thead>
<tr>
<th></th>
<th>jim</th>
<th>jill</th>
<th>jen</th>
<th>joe</th>
</tr>
</thead>
<tbody>
<tr>
<td>jim</td>
<td>0</td>
<td>50</td>
<td>61</td>
<td>57</td>
</tr>
<tr>
<td>jill</td>
<td>50</td>
<td>0</td>
<td>85</td>
<td>41</td>
</tr>
<tr>
<td>jen</td>
<td>61</td>
<td>85</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>joe</td>
<td>57</td>
<td>41</td>
<td>54</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>jen</th>
<th>jill</th>
<th>jim</th>
<th>joe</th>
</tr>
</thead>
<tbody>
<tr>
<td>jen</td>
<td>0</td>
<td>85</td>
<td>61</td>
<td>54</td>
</tr>
<tr>
<td>jill</td>
<td>85</td>
<td>0</td>
<td>50</td>
<td>41</td>
</tr>
<tr>
<td>jim</td>
<td>61</td>
<td>50</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td>joe</td>
<td>54</td>
<td>41</td>
<td>57</td>
<td>0</td>
</tr>
</tbody>
</table>

No dependencies are broken when permuting in this way

50 ≤ 41 + 57
85 ≤ 41 + 54
Example

• Padgett and Ansell data on ties among Florentine families during the Renaissance
  • Marriage ties and business ties
  • Do marriages among families grease the wheels for business ties?
Mixed Dyadic/Node-level

- Two types distinguished by direction of causality
  - Homophily (selection)
    - People tend to interact with / form positive ties with people who are similar to themselves in socially significant ways, such as gender, race, education
  - Contagion (influence)
    - Through interaction, people come to have similar views

- We test these by converting the node attributes (eg gender) into dyadic variables
  - Categorical node variables (e.g., gender)
    - X_{ij} = 1 if node I and node j are same gender, X_{ij} = 0 otherwise
  - Continuous node variables (e.g., age)
    - X_{ij} = difference in age between node I and node j

- Now we correlate the derived matrix X with the network tie matrix
Examples

• Examining the correlation between gender and interaction in the campnet dataset
• Examining the correlation between socializing and choice of karate club
MR-QAP (regression)

• In the campnet example, we also believe role (instructor, participant) affect who interacts with whom
  • Interaction = b0 + b1*same sex + b2*same role
The MRQAP approach was developed by Hubert (1987) and Krackhardt (1987, 1988).

The basic idea is to apply regular regression coefficients and OLS linear regression analysis to dyadic data collected in square matrices; compute *p*-values by a *permutational approach*:

- the null distribution is obtained by permuting *X* values and *Y* values with respect to each other, permuting rows and columns (‘actors’) simultaneously so that the network structure is respected.
MRQAP – cont.

• It was shown by Dekker, Krackhardt and Snijders (2007) how to do this correctly when controlling for other variables (permute residuals; use pivotal statistics).

• This does not model network structure, but controls for it.

• The MRQAP approach is especially useful if one is not interested in network structure per se, but wishes to study linear relations between independent and dyadic dependent variables in a network setting.
Comparison of stochastic models for dyads

• Logistic regression and $p_1$ are inadequate. Assume dyadic independence
• $p_2$
  • does not model network structure except reciprocity
  • gives nicely interpretable regression coefficients and variance-covariance parameters.
• MRQAP
  • does not inherently model network structure (need to include appropriate variables)
  • It always works, regardless of network structure
• ERGM
  • is the only model that represents details of (micro) network structure.
  • Requires considerable experience in model specification
QAP & ERGM
Statistical Horizons ● Social Network Analysis
Steve Borgatti ● LINKS Center ● University of Kentucky
http://tinyurl.com/statisticalhorizons2016
QAP

• Observations in a dyadic regression are not independent – existence of one tie may affect another
  • Row/column dependencies
• Can’t use classical inferential stats
• QAP method allows you to correlate (or regress) relations while controlling for all dependencies
  • You don’t need to know what the dependencies are
  • But it also doesn’t tell you what those dependencies might be

<table>
<thead>
<tr>
<th></th>
<th>Jim</th>
<th>Jill</th>
<th>Jen</th>
<th>Joe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jim</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Jill</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jen</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Joe</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>jim</th>
<th>jill</th>
<th>jen</th>
<th>joe</th>
</tr>
</thead>
<tbody>
<tr>
<td>jim</td>
<td>0</td>
<td>50</td>
<td>61</td>
<td>57</td>
</tr>
<tr>
<td>jill</td>
<td>50</td>
<td>0</td>
<td>85</td>
<td>41</td>
</tr>
<tr>
<td>jen</td>
<td>61</td>
<td>85</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>joe</td>
<td>57</td>
<td>41</td>
<td>54</td>
<td>0</td>
</tr>
</tbody>
</table>
Some dependencies

• Reciprocity
• Common node
• Preferential attachment
• Transitivity
ERGM

• Exponential random graph models
• Instead of controlling for dependencies, ERGMs seek to model them
• An ERGM can tell you whether a network shows a greater than expected tendency to have
  • Reciprocity
  • Transitivity
  • Preferential attachment
  • Homophily, attribute similarity
  • Etc
• And each is assessed while controlling for all the rest
  • Is apparent transitivity just artifact of homophily?
Basic statistical model

\[ P(Y) \propto \prod_{k=1}^{K} \theta_k g_k(y) \]

Probability of the graph \[ \text{coefficient*covariate} \]

Simplest model: Bernoulli random graph (Erdős-Rényi)

All ties \( x_{ij} \) are equally probable and independent (iid)

So the probability of the graph just depends on the cumulative probability of each tie:

\[ \theta \sum_{i=1}^{n} y_{ij} \]
Re-expressed in terms of \( p(\text{tie}) \)

\[
P(Y = y) = \exp \left\{ \sum_{k=1}^{K} \theta_k g_k(y) \right\} / \kappa(\theta)
\]

\[
P(y_{ij} = 1 \mid Y^{(ij)}) = P(Y^+) / \left\{ P(Y^+) + P(Y^-) \right\}
\]

\[
\text{logit}[P(y_{ij} = 1 \mid Y^{(ij)}]) = \theta_1 \partial_1(y^{(ij)}) + \theta_2 \partial_2(y^{(ij)}) + \ldots + \theta_K \partial_K(y^{(ij)})
\]

Conditional log odds of the tie.

\( \delta \) is the "change statistic", the change in the value of the covariate \( g(y) \) when the \( ij \) tie changes from 1 to 0

So \( \theta \) is the impact of the covariate on the log-odds of a tie
QAP vs ERGM

• In ERGM you can have dyadic covariate, so you can see if marriage ties tend to be associated with business ties
  • But you can’t build a linear model, as in business ties = intercept + marriage ties + proximity + political ties

• QAP not as well-suited to testing overall tendencies, such as tendencies to form triangles

• With ERGM, model specification is critical, and no way to introduce new dependencies not anticipated by statistician

• Estimation of ERGM models is computationally tricky
Cohesive Subgroups

Statistical Horizons ● Social Network Analysis
Steve Borgatti ● LINKS Center ● University of Kentucky
http://tinyurl.com/statisticalhorizons2016
A rose by any other name ...

• in multivariate statistics, we call it clustering
• physicists and computer scientists call it finding communities
• mathematicians call it graph partitioning
• Marketers call it segmentation
• social network people have generally talked about detecting cohesive subgroups
Two styles

• Old style is to (mathematically) define a kind of subgroup, then write algorithms to find them
  • Clique (Luce and Perry, 1949). Maximal set of actors each of whom named the other as a friend in a sociometric interview. \{a,b,f\} is only clique
  • N-clique (Mokken, 1979). An n-clique is a maximal subset such that each member is no more than n links away from any other \{a,b,c,f,e\} is a 2-clique
  • K-plex (Seidman and Foster, 1978). A k-plex is a maximal subset of since N such that each member is connected to at least N-k others. \{a,b,e,d\} in bottom graph is 2-plex

• Typically, subgroups are overlapping
New style

• Find are partition of nodes such that there are lots of ties within classes and few ties between classes
  • A partition exhaustively divides the nodes into mutually exclusive classes
• Each class of the partition corresponds to a subgroup
Campnet Example

Partition into 3 classes:

1: HOLLY MICHAEL BILL DON HARRY
2: CAROL PAM PAT JENNIE PAULINE ANN
3: BRAZEY LEE JOHN GERY STEVE BERT RUSS
Johnson’s Hierarchical Clustering -- HiClus

• Output is a set of nested **partitions**, starting with identity partition and ending with the complete partition
  • Partition represented as vector that associates each node with one and only one “group” (mutually exclusive) – is a cluster id variable
  • Identity partition is where every node gets its own cluster
  • Complete partition is where there is only one cluster and every node belongs to it

• Different flavors of hiclus based on how distance from a cluster to outside point/node is defined
  • Single linkage; connectedness; minimum
  • Complete linkage; diameter; maximum
  • Average, median, etc.
Closest distance is NY-BOS = 206, so merge these.
Closest pair is DC to BOS/ NY combo @ 233. So merge these.

<table>
<thead>
<tr>
<th></th>
<th>BOS</th>
<th>NY</th>
<th>DC</th>
<th>MIA</th>
<th>CHI</th>
<th>SEA</th>
<th>SF</th>
<th>LA</th>
<th>DEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOS/ NY</td>
<td>0</td>
<td>233</td>
<td>1308</td>
<td>802</td>
<td>2815</td>
<td>2934</td>
<td>2786</td>
<td>1771</td>
<td></td>
</tr>
<tr>
<td>DC</td>
<td>233</td>
<td>0</td>
<td>1075</td>
<td>671</td>
<td>2684</td>
<td>2799</td>
<td>2631</td>
<td>1616</td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>1308</td>
<td>1075</td>
<td>0</td>
<td>1329</td>
<td>3273</td>
<td>3053</td>
<td>2687</td>
<td>2037</td>
<td></td>
</tr>
<tr>
<td>CHI</td>
<td>802</td>
<td>671</td>
<td>1329</td>
<td>0</td>
<td>2013</td>
<td>2142</td>
<td>2054</td>
<td>996</td>
<td></td>
</tr>
<tr>
<td>SEA</td>
<td>2815</td>
<td>2684</td>
<td>3273</td>
<td>2013</td>
<td>0</td>
<td>808</td>
<td>1131</td>
<td>1307</td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>2934</td>
<td>2799</td>
<td>3053</td>
<td>2142</td>
<td>808</td>
<td>0</td>
<td>379</td>
<td>1235</td>
<td></td>
</tr>
<tr>
<td>LA</td>
<td>2786</td>
<td>2631</td>
<td>2687</td>
<td>2054</td>
<td>1131</td>
<td>379</td>
<td>0</td>
<td>1059</td>
<td></td>
</tr>
<tr>
<td>DEN</td>
<td>1771</td>
<td>1616</td>
<td>2037</td>
<td>996</td>
<td>1307</td>
<td>1235</td>
<td>1059</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BOS/NY/DC/CHI</td>
<td>MIA</td>
<td>SEA</td>
<td>SF/LA</td>
<td>DEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS/NY/DC/CHI</td>
<td>0</td>
<td>1075</td>
<td>2013</td>
<td>2054</td>
<td>996</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>1075</td>
<td>0</td>
<td>3273</td>
<td>2687</td>
<td>2037</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEA</td>
<td>2013</td>
<td>3273</td>
<td>0</td>
<td>808</td>
<td>1307</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF/LA</td>
<td>2054</td>
<td>2687</td>
<td>808</td>
<td>0</td>
<td>1059</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEN</td>
<td>996</td>
<td>2037</td>
<td>1307</td>
<td>1059</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BOS/NY/DC/CHI</td>
<td>MIA</td>
<td>SF/LA/SEA</td>
<td>DEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>------</td>
<td>-----------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS/NY/DC/CHI</td>
<td>0</td>
<td>1075</td>
<td>2013</td>
<td>996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>1075</td>
<td>0</td>
<td>2687</td>
<td>2037</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF/LA/SEA</td>
<td>2054</td>
<td>2687</td>
<td>0</td>
<td>1059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEN</td>
<td>996</td>
<td>2037</td>
<td>1059</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BOS/NY/DC/CHI/DEN</td>
<td>MIA</td>
<td>SF/LA/SEA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------</td>
<td>---------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOS/NY/DC/CHI/DEN</td>
<td>0</td>
<td>1075</td>
<td>1059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIA</td>
<td>1075</td>
<td>0</td>
<td>2687</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF/LA/SEA</td>
<td>1059</td>
<td>2687</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Applying HiClus to Network Data

• Wants symmetric data
• Doesn’t work well with matrix of 0s and 1s – not enough variation to play with
• So first we calculate one of the many dyadic properties, and then cluster on that
  • Geodesic distance
  • Friends in common
  • Maximum flow
  • Symmetrize if needed

Geodesic Distances

```
1  1  1  1  1  1  1  1
  1  2  3  4  5  6  7  8
H  B  C  P  P  J  P  A  M  B  L  D  J  H  G  S  B  R
- - - - - - - - - - - - - - - - - -
1   HOLLY  0  4  2  1  1  2  2  2  1  2  4  1  3  1  2  3  4  3
2  BRAZEY  4  0  5  5  6  4  5  3  4  1  4  3  4  2  1  1  2
3   CAROL  2  5  0  1  1  2  1  2  3  4  5  3  2  3  3  4  4  3
4   PAM    1  5  1  0  2  1  1  2  3  5  2  2  2  3  4  4  3
5    PAT   1  5  1  2  0  1  1  2  2  3  5  2  2  2  3  4  4  3
6  JENNIE  2  6  2  1  1  0  2  1  3  4  6  3  3  3  4  5  5  4
7  PAULINE 2  4  1  1  1  2  0  1  3  4  4  3  1  3  2  3  3  2
8   ANN    2  5  2  1  2  1  1  0  3  4  5  3  2  3  3  4  4  3
9 MICHAEL  1  3  3  2  2  3  3  3  0  1  3  1  2  1  1  2  3  2
10  BILL   2  4  4  3  3  4  4  1  0  4  1  3  1  2  3  4  3
11  LEE    4  1  5  5  6  4  5  3  4  0  4  3  4  2  1  1  2
12  DON    1  4  3  2  2  3  3  3  1  1  4  0  3  1  2  3  4  3
13  JOHN   3  3  2  2  2  3  1  2  2  3  3  3  0  3  1  2  2  1
14  HARRY  1  4  3  2  2  3  3  3  1  1  4  1  3  0  2  3  4  3
15  GERY   2  2  3  3  3  4  2  3  1  2  2  2  1  2  0  1  2  1
16  STEVE  3  1  4  4  4  5  3  4  2  3  1  3  2  3  1  0  1  1
17   BERT  4  1  4  4  4  5  3  4  3  4  1  4  2  4  2  1  0  1
18  RUSS   3  2  3  3  3  4  2  3  2  2  3  1  3  1  1  1  0
```
Hiclus of geo distances

Level 5 3 7 4 1 6 8 0 9 4 2 1 7 2 6 5 3 8

1.000 XXXXX XXX XXX XXXXXXX XXXXXXX XXXXX
1.333 XXXXX XXXXXXX XXXXXXX XXXXXXX XXXXX
1.457 XXXXXXXXXXXXXXX XXXXXXX XXXXXXXXXXXXX
1.481 XXXXXXXXXXXXXXX XXXXXXXXXXXXXX
2.723 XXXXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXX
3.142 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

25 April 2016 (c) 2016 Stephen P. Borgatti
ncd

Q = 0.4896
Girvan-Newman

• Calculate edge-betweenness for graph
• Remove edge with highest edge betweenness
• If number of components increases, record partition
• Recalculate edge betweenness & repeat until all nodes are isolates or maximum number of clusters reached/exceeded
Factions – for binary data

- Combinatorial optimization (tabu search) to find partition that minimizes Hamming distance between observed adjacency matrix and ideal matrix
  - i.e., counts errors

\[
\begin{array}{cccc}
8 & 7 & 9 & 1 \\
8 & 7 & 9 & 1 \\
\hline
1 & 0 & 3 & 2 \\
1 & 1 & 3 & 2 \\
\hline
6 & 4 & 5 & 6 \\
6 & 4 & 5 & 6 \\
\end{array}
\]

\[
\begin{array}{cccc}
8 & 7 & 9 & 1 \\
8 & 7 & 9 & 1 \\
\hline
1 & 0 & 3 & 2 \\
1 & 1 & 3 & 2 \\
\hline
6 & 4 & 5 & 6 \\
6 & 4 & 5 & 6 \\
\end{array}
\]

8 errors

Ideal

8 errors

Best fitting
Factions, 3 & 4 cluster solutions

- 3 cluster: 8 errors, $Q = 0.4896$
- 4 cluster: 6 errors, $Q = 0.4792$
- In this case, $Q$ seems less intuitive
## Campnet Example

**Group Assignments:**

1: HOLLY MICHAEL BILL DON HARRY  
2: CAROL PAM PAT JENNIE PAULINE ANN  
3: BRAZEY LEE JOHN GERY STEVE BERT RUSS

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 1 | HOLLY | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 10| BILL   | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 12 | DON   | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 14 | HARRY | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 9  | MICHAEL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4  | PAM   | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6  | JENNIE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8  | ANN   | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7  | PAULINE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5  | PAT   | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 3  | CAROL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 11 | LEE   | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 13 | JOHN  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2  | BRAZEY | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 15 | GERY  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 16 | STEVE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 17 | BERT  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 18 | RUSS  | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

---

25 April 2016 (c) 2016 Stephen P. Borgatti
Clique

- A maximal complete subgraph
  - Everyone is adjacent to everyone else
  - Avg distance & Diameter is 1
  - Density is 1

- Limitations
  - Undirected

- Issues
  - Multiple, overlapping
  - Too strict

10 cliques found.

1: HOLLY MICHAEL DON HARRY
2: BRAZEY LEE STEVE BERT
3: CAROL PAT PAULINE
4: CAROL PAM PAULINE
5: PAM JENNIE ANN
6: PAM PAULINE ANN
7: MICHAEL BILL DON HARRY
8: JOHN GERY RUSS
9: GERY STEVE RUSS
10: STEVE BERT RUSS
Overlap

• A graph with 21 vertices can have as many as 2187 cliques
• Secondary analysis looks at the overlap matrices, e.g.,
  • form a node-by-node matrix O such that $o_{ij}$ gives number of cliques i and j are both in – co-membership matrix
  • Run hierarchical clustering on O
Help with the rice harvest

Villages in northern Thailand studied by Entwisle & Faust

Ties indicated helping with the rice harvest. Nodes are households, located by GPS coordinates.

Village 1

Village 2

(c) 2015 Stephen P. Borgatti
Which networks are good for what?

- Consequences of these structures for the organization and for nodes

- Antecedents of these structures. How do they come about, whether by change or on purpose?
Centralization

• Extent to which network revolves around a single node

• Defined as
  • the sum of differences between the centrality of the most central node, and the centrality of every other node, divided by normalizing constant to make it run between 0 and 1

  \[ C = \frac{\sum_i d_{\text{max}} - d_i}{(n-1)(n-2)} \]

  • For a perfect star w/ 5 nodes, the center has \( d = 4 \), and all others have \( d = 1 \), so the numerator is \( 0 + 4(3) \) which is \((5-1)(5-2)\), so centralization = 1.0
Core/Periphery

• Extent to which there is a “core” of people that holds the network together, such that
  • Core people are well connected to other core people, in general
  • Periphery people are connected to core people
  • Periphery people are NOT connected to other periphery people
Core/Periphery
Finding Core/Periphery Structures

• Two approaches
  • Discrete/blockmodeling
    • Use combinatorial optimization to partition nodes into core and periphery sets such that core-core ties are maximized and periphery-periphery ties are minimized
  • Continuous
    • Calculate coreness of each node by modeling existence/strength of ties between pair of nodes as function of coreness of each
Categorical Approach

• Use combinatorial optimization to partition nodes into core and periphery sets such that
  • core-core ties are maximized
  • periphery-periphery ties are minimized
  • Core to Periphery: unspecified, but normally expect in-between value
## Categorical Results

| KAMWEFU | 1 1 1 1 1 1 1 | 1 1 1 1 1 1 1 |
| NKUMBULA | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| ABRAHAM | 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| BASTINGS | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| CHIVATA | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| MEHAK | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| NKOLAYA | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| SUKO | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| KALAMBA | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| MUKUBWA | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| NKOLOYA | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| ZULU | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| CHIPATA | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| MESHAK | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| NKOLOYA | 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| ABRAHAM | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| HASTINGS | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
| JOSEPH | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1 1 1 1 |
|  |
|  |
|  |

### Density Matrix

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.699</td>
<td>0.235</td>
</tr>
<tr>
<td>0.235</td>
<td>0.173</td>
</tr>
</tbody>
</table>

Kaptail-kapfts2

---

4/25/2016 (c) 2015 Stephen P. Borgatti
Continuous approach

- Discrete model effectively creates binary coreness variable such that ties between I and j given by product coreness of each
  - If $c_i$ and $c_j = 1$ then $X_{ij} = 1$, and if $c_i$ and $c_j = 0$ then $X_{ij} = 0$
- So we generalize to continuous coreness scores such that prob/strength of a tie between I and j is a function of the coreness of each
  - $X_{ij} = f(c_i * c_j)$
    - If both have high coreness, then tied to each other
    - If both have low coreness, then not tied
- So we use a least-squares type procedure to find scores $c$ to minimize
  - Fitting a model of ties
    $$\sum_{i,j} (x_{ij} - c_i c_j)^2$$
Continuous coreness

Colors based on the discrete model. Sizes based on continuous model.
Measure cpness

• Both discrete and continuous approaches fit a model to the data, i.e., predict ties
  • Discrete
    • If $c_i = 1$ and $c_j = 1$ then $x_{ij} = 1$
    • If $c_i = 0$ and $c_j = 0$ then $x_{ij} = 0$
  • Continuous
    • $\text{Prob}(x_{ij}) = f(c_i*c_j)$

• So in both cases we can measure goodness of fit
  • Degree to which data conforms to idealized cp structure
  • Simply correlate data matrix with idealized matrix.
C/P Structures & Morale

Study by Jeff Johnson of a South Pole scientific team over 8 months

C/P structure seems to affect morale

C/P Structures & Morale

Month

Group Morale

Core/Periphery-ness
Local Position

Statistical Horizons • Social Network Analysis
Steve Borgatti • LINKS Center • University of Kentucky

http://tinyurl.com/statisticalhorizons2016
Overview

• Characterizing a node’s ego network – their network neighborhood
• Measures of network composition, heterogeneity, homophily
• Structural holes
• First-order neighborhood
  • All nodes ("alters") directly connected to ego
    • In either direction?
  • Ties among those nodes
• 2\textsuperscript{nd} order neighborhood
  • All nodes within two links of ego
    • Respecting direction?
Two sources of ego net data

• Full network research designs
  • Start with a set of actors, typically members of a group
    • E.g., all employees of Central Bank
  • Measure ties among them – e.g., who is friends with whom

• Personal network research designs – see ENET program
  • Sample a set of actors (called egos) from any population – e.g., call up random phone numbers
  • Ask each respondent about the people in their lives (called alters)
  • Ask the respondent about alter attributes – gender, age, etc.
  • Ideally, ask the respondent about ties among the alters.
Egonet size

- Number of ties
  - Typically, of all kinds
  - If just one kind, is same as degree

- Outdegree
  - People named by ego

- Indegree
  - People who name ego
  - (only in full network designs)

\[ d_{i}^{\text{out}} = \sum_{j} x_{ij} \]

\[ d_{j}^{\text{in}} = \sum_{i} x_{ij} \]

Indegree x Outdegree in “seeks advice” network
Valued ties

• Degree formula works fine for valued ties
  • Is the sum of tie strengths with others
    • Bigger value indicates more connectedness to others in the dept, neighborhood, etc.
  • Does introduce a question of normalization

• Variants include
  • Average: just a rescaling of sum, but easier to understand
  • Median: more robust version of average
  • Maximum: how strong is your strongest tie?
    • Psychological health ← have a best friend?

• Strength of weak ties theory
  • Seems to say less connected to others the more novel information
Egonet composition

• Summarizing who is in a person’s ego network
  • Examining attributes of ego’s alters

• Categorical attributes
  • Gender, race

• Continuous attributes
  • Age, income

• Objective is to summarize distribution of attribute values among ego’s alters
Categorical alter attributes

• Counts of the number of alters of different types
  • How many gay people is ego friends with?
  • How many white people does ego seek advice from?
  • Expressed as either frequencies or proportions
  • Can characterize ego by modal type of attribute
    • Bill has a male network

• Heterogeneity
  • How evenly distributed are ego’s alters with respect to the attribute categories
    • Equal number of each gender? Mostly African-Americans?
    • Diversity of ego’s contacts

<table>
<thead>
<tr>
<th></th>
<th>freq</th>
<th>prop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>14</td>
<td>0.875</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>Jane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>0.5</td>
</tr>
<tr>
<td>Female</td>
<td>8</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>1</td>
</tr>
</tbody>
</table>
Categorical Heterogeneity

• Blau / Herfindhal / Hirschman index
  • $p_k$ is proportion of ego’s alters that fall in category k
  \[
  H = 1 - \sum_k p_k^2
  \]
  • $H = 0$ if all alters in one category
  • $H = 1 - \frac{1}{K}$ if all categories have equal frequency
    • The measure cannot reach 1.0 unless there are infinite categories

<table>
<thead>
<tr>
<th></th>
<th>freq</th>
<th>prop</th>
<th>prop^2</th>
<th>K</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>14</td>
<td>0.875</td>
<td>0.765625</td>
<td>H</td>
<td>0.219</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.125</td>
<td>0.015625</td>
<td>IQV</td>
<td>0.438</td>
</tr>
<tr>
<td>Female</td>
<td>16</td>
<td>1</td>
<td>0.78125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>0.5</td>
<td>0.25</td>
<td>H</td>
<td>0.5</td>
</tr>
<tr>
<td>Female</td>
<td>16</td>
<td>1</td>
<td>0.5</td>
<td>IQV</td>
<td>1</td>
</tr>
</tbody>
</table>
Categorical Heterogeneity

• Agresti’s IQV
  • Divide H by $1 - 1/K$ so that measure runs between 0 and 1
    • Gives amount of diversity given the number of categories
  • But H could be seen as best measure of diversity, because it is not satisfied until the number of categories $\rightarrow \infty$, which would imply massive diversity

$$IQV = \frac{H}{1 - \frac{1}{K}}$$

<table>
<thead>
<tr>
<th></th>
<th>freq</th>
<th>prop</th>
<th>prop^2</th>
<th>K = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting</td>
<td>14</td>
<td>0.560</td>
<td>0.314</td>
<td>H = 0.618</td>
</tr>
<tr>
<td>Economics</td>
<td>5</td>
<td>0.200</td>
<td>0.040</td>
<td>IQV = 0.772</td>
</tr>
<tr>
<td>Finance</td>
<td>4</td>
<td>0.160</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>1</td>
<td>0.040</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Marketing</td>
<td>1</td>
<td>0.040</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>total:</td>
<td>25</td>
<td>1.000</td>
<td>0.382</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>freq</th>
<th>prop</th>
<th>prop^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting</td>
<td>5</td>
<td>0.200</td>
<td>0.040</td>
</tr>
<tr>
<td>Economics</td>
<td>5</td>
<td>0.200</td>
<td>0.040</td>
</tr>
<tr>
<td>Finance</td>
<td>5</td>
<td>0.200</td>
<td>0.040</td>
</tr>
<tr>
<td>Management</td>
<td>5</td>
<td>0.200</td>
<td>0.040</td>
</tr>
<tr>
<td>Marketing</td>
<td>5</td>
<td>0.200</td>
<td>0.040</td>
</tr>
<tr>
<td>total:</td>
<td>25</td>
<td>1.000</td>
<td>0.200</td>
</tr>
</tbody>
</table>
Continuous alter attributes

• Central tendency
  • Median or mean
  • Total
  • Can use weighted mean where weight is the strength of tie to the alter

• Dispersion
  • Standard deviation
    • Bill’s alters have an average income of $36K with low SD
    • Jane’s alters have an average of income of $36K with high SD
    • Who has an advantage in terms of access to resources?
  • Gini coefficient on binned data
Homophily

- Tendency for people to form positive ties with people similar to themselves on socially significant attributes
  - Race
  - Gender
  - Age
  - Education
  - Social class
  - Interests, activities, etc – social foci (Feld)

- Why?
### A. Age Respondent

<table>
<thead>
<tr>
<th>Alter</th>
<th>&lt; 30</th>
<th>30–39</th>
<th>40–49</th>
<th>50–59</th>
<th>60 or over</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 30</td>
<td>567</td>
<td>186</td>
<td>183</td>
<td>155</td>
<td>56</td>
</tr>
<tr>
<td>30–39</td>
<td>191</td>
<td>501</td>
<td>171</td>
<td>128</td>
<td>106</td>
</tr>
<tr>
<td>40–49</td>
<td>88</td>
<td>170</td>
<td>246</td>
<td>84</td>
<td>70</td>
</tr>
<tr>
<td>50–59</td>
<td>84</td>
<td>100</td>
<td>121</td>
<td>210</td>
<td>108</td>
</tr>
<tr>
<td>60 or over</td>
<td>34</td>
<td>127</td>
<td>138</td>
<td>212</td>
<td>387</td>
</tr>
</tbody>
</table>

### B. Education Respondent

<table>
<thead>
<tr>
<th>Alter</th>
<th>1–6</th>
<th>7–9</th>
<th>10–12</th>
<th>H.S. Grad.</th>
<th>Some College</th>
<th>Assoc. Degree</th>
<th>Bach. Degree</th>
<th>Grad. Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–6 years</td>
<td>16</td>
<td>20</td>
<td>9</td>
<td>36</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>7–9 years</td>
<td>10</td>
<td>54</td>
<td>43</td>
<td>105</td>
<td>24</td>
<td>7</td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>10–12 years</td>
<td>15</td>
<td>50</td>
<td>89</td>
<td>156</td>
<td>50</td>
<td>10</td>
<td>31</td>
<td>19</td>
</tr>
<tr>
<td>H.S. Grad</td>
<td>14</td>
<td>70</td>
<td>183</td>
<td>658</td>
<td>250</td>
<td>39</td>
<td>115</td>
<td>85</td>
</tr>
<tr>
<td>Some Coll.</td>
<td>6</td>
<td>28</td>
<td>76</td>
<td>276</td>
<td>225</td>
<td>44</td>
<td>137</td>
<td>93</td>
</tr>
<tr>
<td>Assoc. Deg.</td>
<td>2</td>
<td>2</td>
<td>13</td>
<td>55</td>
<td>38</td>
<td>22</td>
<td>45</td>
<td>35</td>
</tr>
<tr>
<td>Bach. Deg.</td>
<td>4</td>
<td>19</td>
<td>11</td>
<td>130</td>
<td>128</td>
<td>27</td>
<td>230</td>
<td>121</td>
</tr>
<tr>
<td>Grad. Deg.</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>50</td>
<td>30</td>
<td>12</td>
<td>92</td>
<td>149</td>
</tr>
</tbody>
</table>
### C. Race/ethnicity

<table>
<thead>
<tr>
<th>Race/ethnicity</th>
<th>Alter</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Black</td>
<td>Hispanic</td>
<td>Other</td>
</tr>
<tr>
<td>White</td>
<td>3806</td>
<td>29</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Black</td>
<td>40</td>
<td>283</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Hispanic</td>
<td>66</td>
<td>6</td>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>21</td>
<td>5</td>
<td>3</td>
<td>34</td>
</tr>
</tbody>
</table>

### D. Religious preference

<table>
<thead>
<tr>
<th>Religious preference</th>
<th>Alter</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Protestant</td>
<td>Catholic</td>
<td>Jewish</td>
<td>None</td>
<td>Other</td>
</tr>
<tr>
<td>Protestant</td>
<td>2129</td>
<td>305</td>
<td>22</td>
<td>83</td>
<td>30</td>
</tr>
<tr>
<td>Catholic</td>
<td>241</td>
<td>790</td>
<td>24</td>
<td>41</td>
<td>13</td>
</tr>
<tr>
<td>Jewish</td>
<td>13</td>
<td>7</td>
<td>68</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>None</td>
<td>92</td>
<td>66</td>
<td>12</td>
<td>131</td>
<td>14</td>
</tr>
<tr>
<td>Other</td>
<td>27</td>
<td>11</td>
<td>1</td>
<td>4</td>
<td>37</td>
</tr>
</tbody>
</table>

### E. Sex

<table>
<thead>
<tr>
<th>Sex</th>
<th>Alter</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
<td>1245</td>
<td>748</td>
</tr>
<tr>
<td>Female</td>
<td>970</td>
<td>1515</td>
</tr>
</tbody>
</table>
- Propinquity effects
- Homophily with respect to spatial coordinates

**Figure 1a.** Number of local phone calls by distance between parties. (Mayer, 1978)

**Figure 1b.** Number of electronic messages by distance between parties. (Eveland and Blikson, 1987)

**Figure 1c.** Probability of communication by distance between potential communicators. (Allen, 1977)

**Figure 1d.** Communication frequency by distance between collaborators.
Random Network in Two Dimensions

Homophilous Network in Two Dimensions

SocioDemographic Space
Measuring homophily

• Measuring the extent to which egos resemble their alters
  • Men’s friends tend to be men, women’s friends women

• Selection (homophily) or influence?
  • Is ego seeking out similars (selection/homophily), or is ego influencing alters to become like self (influence)?
  • In case of gender, probably selection
  • In case of smoking, probably both

• Multiple measures
  • In general can’t distinguish selection from influence
Measures of node-level homophily/influence

• **Pct of matches**
  • What proportion of ego’s alters have same attribute value as ego does?
    • If ego is male, what proportion of alters are male?
      • If 0, then perfectly heterophilous. If 1, then perfectly homophilous

• **E-I index (Krackhardt and Stern)**
  • \( \frac{E-I}{E+I} \), where E (external) is number of alters different from ego, and I (internal) is number of alters same as ego
  • Is a linear rescaling of % matches, which is \( \frac{I}{I+E} \)

• Note that if most people are white, then if people chose without regard for color, most whites will be homophilous, and most blacks will be heterophilous
Importance of non-ties (whole network designs only)

• % matches and E-I take into account only ties that are present
  • They don’t count non-ties (choices to not befriend someone)

<table>
<thead>
<tr>
<th>Same group</th>
<th>Has Tie</th>
<th>1</th>
<th>0</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>50</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>75</td>
<td>90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are 90 potential alters. Ego has a tie with 15 of them. Of these 15, 10 are the same race/gender/etc as ego

• The above measures only use the values in the “1” column, so it looks like this person prefers own kind 2 to 1
• But if you look at who they are not tied with (0 column), you see they also prefer to NOT have a tie with their own kind – also at odds of 2 to 1
  • So they have no preference for own kind. It is just that there are more of them
Odds ratio and Yule’s Q

• Odds ratio is just \[ \frac{a/b}{c/d} = \frac{ad}{bc} \]
  - When this is 1, then we have independence – no preference for ingroup ties
  - Odds ratio is unbounded

• Yule’s Q bounds the odds ratio between -1 and 1
  \[ \frac{ad-bc}{ad+bc} \]
  - where 0 indicates no association, 1 is perfect homophily and -1 is perfect heterophily

• Key advantage of these measures is they are not fooled by group sizes

<table>
<thead>
<tr>
<th>Has Tie</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same group</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

Yule’s Q = 0

<table>
<thead>
<tr>
<th>Has Tie</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same group</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>550</td>
<td>110</td>
</tr>
</tbody>
</table>

Yule’s Q = 0

(a) 2015 Stephen P. Borgatti
Yule’s Q and the correlation coefficient

- Yule’s Q is \( \frac{ad-bc}{ad+bc} \)
- Pearson correlation is \( \frac{ad - bc}{\sqrt{(a + c)(b + d)(a + b)(c + d)}} \)

- Just a different denominator from Yule’s Q
- Yule’s Q has advantage that it equals 1 if nodes only have ties to same group, whereas correlation adds condition that nodes must have ties to all members of their group
  - In real settings, correlation can never achieve it’s maximum value of 1
Yule’s Q example

• Yule’s Q gives 1.0 to both nodes e and j, but corr only gives 1.0 for e

• On average, the reds are more homophilous

<table>
<thead>
<tr>
<th>% Same</th>
<th>El Index</th>
<th>Matches</th>
<th>Cohen Kappa</th>
<th>Yules Q</th>
<th>Corr</th>
<th>fInGro</th>
<th>fOutGr</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.000</td>
<td>1.000</td>
<td>0.444</td>
<td>-1.000</td>
<td>-0.216</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>b</td>
<td>0.000</td>
<td>1.000</td>
<td>0.444</td>
<td>-1.000</td>
<td>-0.216</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c</td>
<td>0.500</td>
<td>0.000</td>
<td>0.556</td>
<td>0.143</td>
<td>0.053</td>
<td>0.060</td>
<td>1</td>
</tr>
<tr>
<td>d</td>
<td>0.500</td>
<td>0.000</td>
<td>0.556</td>
<td>0.143</td>
<td>0.053</td>
<td>0.060</td>
<td>1</td>
</tr>
<tr>
<td>e</td>
<td>1.000</td>
<td>-1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>4</td>
</tr>
<tr>
<td>f</td>
<td>0.500</td>
<td>0.000</td>
<td>0.556</td>
<td>0.143</td>
<td>0.053</td>
<td>0.060</td>
<td>1</td>
</tr>
<tr>
<td>g</td>
<td>0.500</td>
<td>0.000</td>
<td>0.556</td>
<td>0.143</td>
<td>0.053</td>
<td>0.060</td>
<td>1</td>
</tr>
<tr>
<td>h</td>
<td>0.500</td>
<td>0.000</td>
<td>0.556</td>
<td>0.143</td>
<td>0.053</td>
<td>0.060</td>
<td>1</td>
</tr>
<tr>
<td>i</td>
<td>0.500</td>
<td>0.000</td>
<td>0.556</td>
<td>0.143</td>
<td>0.053</td>
<td>0.060</td>
<td>1</td>
</tr>
<tr>
<td>j</td>
<td>1.000</td>
<td>-1.000</td>
<td>0.778</td>
<td>1.000</td>
<td>0.526</td>
<td>0.598</td>
<td>2</td>
</tr>
</tbody>
</table>

Why does h have a positive score? Shouldn’t it be a perfect 0.0? No, because it shuns 4/5 reds but only 3/4 blues
Summarizing measures

• Measures that only look at a person’s ties ...
   • Don’t correct for availability of different groups
   • Measure realized levels of contact with ingroup vs outgroup members
   • Are the only measure available in personal network designs, where the alters aren’t interviewed

• Measures that look at both ties and non-ties ...
   • Correct for availability
   • Measure underlying preferences for others, so that even having just two fellow Eskimo friends is indicative of homophily
   • Can only be calculated on data collected in whole network research designs
Local Social Capital

Statistical Horizons ● Social Network Analysis
Steve Borgatti ● LINKS Center ● University of Kentucky

http://tinyurl.com/statisticalhorizons2016
Social resource theory

• Associated with Nan Lin
• Resources are necessary for achievement
• But individuals need not own resources in order to make use of them
• Access to resources through social ties
  • Abbreviated social resources
• Having friends in high places
  • Getting dispensation from local officials to rezone area
• Having friends with money to invest in your projects
Measures

• Most common measure is sum of resources of alters

\[ c_i = \sum_j w_{ij} r_j \]

\( r_j \) is amount of resource controlled by alter j, 
\( w_{ij} \) is the influence i has on j. e.g., dependence of j on i for 
\( c_i \) is the total amount of resource that i can indirectly control

• Another measure is maximum resource of any alter, weighted by influence

\[ c_i = \max_j (w_{ij} r_j) \]
Structural holes

• Burt ’92 theory of individual social capital
• Structural advantage
  • Not based on the attributes of ego’s alters, but on the structure of the ego network
• Specifically, the lack of ties among alters
• Benefits
  • Autonomy
  • Control
  • Information
Autonomy

• Independence
• Freedom of action
• Fluid identity
Control Benefits

White House Diary Data, Carter Presidency

Year 1

Year 4

Data courtesy of Michael Link
The vision advantage

Information benefits

Brokerage across Structural Holes 

Achievement & Rewards (What benefits?)

Adaptive Implementation (How to frame it & who should be involved?)

Creativity & Learning (What should be done?)

Alternative Perspective (how would this problem look from the perspective of a different group, or groups — thinking "out of the box" is often less valuable than seeing the problem as it would look if you were inside a specific "other box")

Best Practice (something they think or do could be valuable in my operations)

Analogy (something about the way they think or behave has implications for how I can enhance the value of my operations; i.e., look for the value of juxtapositioning two clusters, not reasons why the two are different so as to be irrelevant to one another — you often find what you look for)

Synergy (resources in our separate operations can be combined to create a valuable new idea/practice/product)

Information & Success

Cultural interventions, relationship building

Data warehousing, systems architecture

Information flow within virtual group

Changes Made

• Cross-staffed new internal projects
  • white papers, database development

• Established cross-selling sales goals
  • managers accountable for selling projects with both kinds of expertise

• New communication vehicles
  • project tracking db; weekly email update

• Personnel changes
9 Months Later

Note: Different EV – same initials.
Measures of Structural Holes

- Burt’s effective size
- Burt’s constraint
Effective Size

\[ m_{jq} = \text{j's interaction with q divided by j's strongest relation with anyone} \]

\[ p_{iq} = \text{proportion of i's energy invested in relation with q} \]

\[ ES_i = \sum_j \left[ 1 - \sum_q p_{iq} m_{jq} \right], \quad q \neq i, j \]

\[ ES_i = \sum_j \sum_q p_{iq} m_{jq}, \quad q \neq i, j \]

• Effective size is network size (N) minus redundancy in network

Figure 1. Adapted from Burt (1995:56)
Simplifying effective size for case of 1/0 data

- \( M_{jq} = j\)'s interaction with \( q \) divided by \( j\)'s strongest tie with anyone
  - So this is always 1 if \( j \) has tie to \( q \) and 0 otherwise
- \( P_{iq} = \) proportion of \( i\)'s energy invested in relationship with \( q \)
  - So this is a constant \( 1/N \) where \( N \) is ego’s network size
- Effective size reduces to network size minus the average degree among the alters

\[
ES_i = \sum_j \left( 1 - \sum_q p_{iq} m_{jq} \right), \quad q \neq i, j
\]

\[
ES_i = \sum_j \left( 1 - \frac{1}{n} \sum_q m_{jq} \right), \quad q \neq i, j
\]

\[
ES_i = \sum_j \left( 1 - \frac{1}{n} \sum_q m_{jq} \right), \quad q \neq i, j
\]

Effective Size of \( G \) = Number of \( G\)'s Alters – Sum of Redundancy of \( G\)'s alters

\[
= 6 - 1.33 = 4.67
\]

Sized by Effective Size
Constraint

M_{jq} = j’s interaction with q divided by j’s strongest relationship with anyone
So this is always 1 if j has tie to q and 0 otherwise

P_{iq} = proportion of i’s energy invested in relationship with q
So this is a constant 1/N where N is network size

\[ c_{ij} = \left( p_{ij} + \sum_{q} p_{iq} p_{qj} \right)^2, \quad q \neq i, j \]
\[ c_i = \sum_j c_{ij} \]

- Reverse-coded measure of structural holes: large values = fewer holes
- Alter j constrains i to the extent that
  - i has invested in j
  - i has invested in people (q) who have invested heavily in j. That is, i’s investment in q leads back to j.
- Even if i withdraws from j, others in i’s network still invested in j
- Overall constraint is the extent to which i is invested in those that other contacts of i are invested in
Gould & Fernandez brokerage roles

• Gould & Fernandez (1989)
• Broker is middle node of directed triad (note: a is NOT connected to c)
• What if nodes belong to different groups?
  • Categorical node attribute such as dept, ethnic group
G&F Brokerage Roles

- Coordinator
- Representative
- Gatekeeper
- Consultant

- We can count how often a node enacts each kind of brokerage role
Caveats on G & F brokerage roles

• Just because B is in the structural position to be a representative doesn’t mean she ever does the associated behaviors
  • Is the tie “gives information to” or something irrelevant like “likes”?  
  • Even if tie is gives info to, no evidence that what A gives B is the same thing that B gives C

\begin{tikzpicture}
    
    
    
    
    \node [left] (A) at (0,0) {A};
    \node [right] (B) at (1,1) {B};
    \node [right] (C) at (1,-1) {C};

    \draw [->] (A) to (B);
    \draw [->] (B) to (C);

    \node [below] at (B) {Representative};

\end{tikzpicture}
Centrality

Statistical Horizons ● Social Network Analysis
Steve Borgatti ● LINKS Center ● University of Kentucky
http://tinyurl.com/statisticalhorizons2016
What is centrality?

• Structural importance
• An aspect of a node’s position in a network
• Measures or constructs?
• Centrality & centralization
• Big 4
  • Degree
  • Closeness
  • Betweenness
  • Eigenvector
Big 4 in undirected graphs
Degree centrality

• Barely a centrality measure, as you don’t need to know the structure of the network to calculate it

• Number of ties a node has
  • In most cases, this is also number of nodes the node is adjacent to

• Interpreted as exposure and influence

• Highly correlated w/ many outcomes
Degree as row sums or averages

\[ d_i = \sum_j a_{ij} \]

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>Sum</th>
<th>Avg</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>1</td>
<td>.2</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>.6</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>.4</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>.6</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>.4</td>
</tr>
<tr>
<td>f</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>.2</td>
</tr>
</tbody>
</table>
Turbo-charging degree

• Degree is a count of the number of nodes you are connected to
  • Treats all nodes equally

• What if you wanted to weight the nodes by how many nodes they
  were connected to?

\[ t_i = \sum_{j} a_{ij} d_j \]

• But why stop there? Can keep iterating ...
### Iterated Degree

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
<th>D8</th>
<th>D9</th>
<th>D10</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>1</td>
<td>3</td>
<td>6</td>
<td>16</td>
<td>35</td>
<td>86</td>
<td>195</td>
<td>465</td>
<td>1071</td>
<td>2524</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>16</td>
<td>35</td>
<td>86</td>
<td>195</td>
<td>465</td>
<td>1071</td>
<td>2524</td>
<td>5854</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>13</td>
<td>32</td>
<td>73</td>
<td>173</td>
<td>401</td>
<td>940</td>
<td>2190</td>
<td>5117</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>16</td>
<td>38</td>
<td>87</td>
<td>206</td>
<td>475</td>
<td>1119</td>
<td>2593</td>
<td>6086</td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>20</td>
<td>47</td>
<td>107</td>
<td>253</td>
<td>582</td>
<td>1372</td>
<td>3175</td>
</tr>
<tr>
<td>f</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>20</td>
<td>47</td>
<td>107</td>
<td>253</td>
<td>582</td>
<td>1372</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>28</td>
<td>64</td>
<td>150</td>
<td>348</td>
<td>814</td>
<td>1896</td>
<td>4430</td>
<td>10332</td>
<td>24128</td>
</tr>
</tbody>
</table>

### Adjacency Matrix

- **A**
  - Degree: 8.3, 10.7, 9.4, 10.7, 10.1, 10.6, 10.3, 10.5, 10.4, 10.5
- **B**
- **C**
- **D**
- **E**
- **F**
  - Degree: 8.3, 7.1, 6.3, 6.0, 5.7, 5.8, 5.6, 5.7, 5.6, 5.7
Eigenvector

• Principal eigenvector of network adjacency matrix A

\[ \mathbf{A} \mathbf{v} = \lambda \mathbf{v} \]

\( \mathbf{v} \) is the eigenvector, \( \lambda \) is the associated eigenvalue

• A node has high eigenvector score to the extent it is connected to many nodes who themselves have high scores

• Often interpreted as popularity or status – have ties not just to many others but many well-connected others
  - A kind of turbo-charged degree centrality
Eigenvector

- Node $d$ has the highest eigenvector centrality
Issues with eigenvector

• Can’t use with disconnected networks
• In clumpy networks, it favors the nodes in the larger cliques
• Fails as a measure of risk/exposure because it doesn’t take into account the fact that an alter’s high degree might be because of ties with nodes that ego is already connected to
  • So shouldn’t give that alter any weight, because they are not adding to exposure
• Perhaps better as a measure of clique membership
Closeness

- Sum of distances from node to all others
- Inverse measure of centrality
- Often interpreted as index of time-until-arrival of stuff flowing through network
  - In gossip network, persons strong in closeness centrality hear things early
### Closeness as marginals of distance matrix

<table>
<thead>
<tr>
<th>ID</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>
<th>k</th>
<th>l</th>
<th>m</th>
<th>n</th>
<th>o</th>
<th>p</th>
<th>q</th>
<th>r</th>
<th>s</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>62</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>d</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>g</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>49</td>
</tr>
<tr>
<td>h</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>i</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>j</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>k</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>l</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>m</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>n</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>o</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>p</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>q</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>r</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>s</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>sum</td>
<td>62</td>
<td>61</td>
<td>59</td>
<td>62</td>
<td>59</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
</tr>
</tbody>
</table>

Node i has the best closeness

Average distance would be more interpretable

4/25/2016

(c) 2015 Steve Borgatti
Bavelas-Leavitt experiments

<table>
<thead>
<tr>
<th>FPT</th>
<th>3</th>
<th>5</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>50.4</td>
<td>53.2</td>
<td>35.4</td>
<td>32</td>
</tr>
<tr>
<td>No. of errors</td>
<td>7.6</td>
<td>2.8</td>
<td>0</td>
<td>0.6</td>
</tr>
<tr>
<td>No. of msgs</td>
<td>high</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
</tbody>
</table>
Issues & variants of closeness

• When graphs are disconnected, distances between some nodes are undefined
  • What to do?

• Typical solutions
  • Replace undefined distances with a large number, then calculate as usual
    • Use diameter + 1 (1 larger than longest observed distance)
    • Use N (1 larger than longest possible distance)
  • Replace entries $d_{ij}$ of distance matrix with $1/d_{ij}$, and set $1/d_{ij}$ to zero when $d_{ij}$ is undefined
  • Calculate closeness only within components, and don’t compare across components (probably a bad idea)
Betweenness

• Loosely, the extent to which a node is along the shortest paths of between all pairs of nodes

\[ b_k = \sum_{i,j} \frac{g_{ikj}}{g_{ij}} \]

* \( g_{ij} \) is number of geodesic paths from i to j
* \( g_{ikj} \) is number of geodesics from i to j that pass through k

• More correctly, \( b_k \) is the share of geodesics between pairs of nodes that pass through k

• Often interpreted as control over flows (gatekeeping), correlated with power

• Also seen as index of frequency something reaches node
Betweenness

• Node $h$ has the highest betweenness
Betweenness – cont.

• Often discussed in terms of identifying liaisons, gatekeepers, “secretary power”
• Global network cohesion is highly dependent on high betweenness nodes.
  • (But) networks that contain high betweenness nodes are brittle
• Nodes with high betweenness and low degree are often overlooked by network members themselves
  • Degree is highly visible, betweenness may not be
The emergence of Moscow

• Pitts (1979) study of 12th century Russia and the later emergence of Moscow

• Why did Moscow come to dominate?
  • Great man theory
  • Resource richness
Emergence of Moscow

• Rivers enable trade between city-states
  • System of rivers creates network of who can trade directly and indirectly with whom
  • What happens in the network is a function of global paths and position
  • Moscow very high in betweenness centrality

Nodes have high betweenness to the extent they are along the shortest paths between pairs of nodes
Duality of closeness & betweenness

- Dependency matrix $D$, where $d_{ij} =$ number of times* that $i$ needs to go through $j$ to reach someone via a shortest path
- Column totals of $D$ equal betweenness times 2
- Row totals of $D$ equal closeness minus $n-1$

*Note: The asterisk indicates an element of the matrix $D$. The specific meaning of this element is not provided within the text.
Big 4 in directed graphs
Degree

• Out-degree is the number of outgoing ties a node has
  • For positive ties, a measure of gregariousness
  • In survey data, often seen as potentially biased

• In-degree is the number of incoming ties
  • For positive ties, popularity, prestige, status, influence, etc
Directed degree – cont.

- Outdegree and indegree correspond to the row and column sums of the adjacency matrix
  - Outdegree = row sums

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

- Outdegree

- Indegree
### Indegree and outdegree

<table>
<thead>
<tr>
<th></th>
<th>Braze</th>
<th>Caro</th>
<th>Jenni</th>
<th>Pauli</th>
<th>Mich</th>
<th>Bill</th>
<th>Lee</th>
<th>Don</th>
<th>John</th>
<th>Harry</th>
<th>Gery</th>
<th>Steve</th>
<th>Bert</th>
<th>Russ</th>
<th>Outdeg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holly</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Braze</td>
<td>0</td>
<td>0</td>
<td>0</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>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Carol</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Pam</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Pat</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Jennie</td>
<td>0</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Pauline</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ann</td>
<td>0</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Michael</td>
<td>1</td>
<td>0</td>
<td>0</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>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Bill</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Lee</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Don</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>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>John</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</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>1</td>
<td>3</td>
</tr>
<tr>
<td>Harry</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>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Gery</td>
<td>0</td>
<td>0</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>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Steve</td>
<td>0</td>
<td>0</td>
<td>0</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>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Bert</td>
<td>0</td>
<td>0</td>
<td>0</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>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Russ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</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>1</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Table Notes
- **Indeg**: Indegree values for each node.
- **Outdeg**: Outdegree values for each node.

**25 April 2016**

Copyright (c) 2013 by Steve Borgatti. All rights reserved.
Plot indegree vs outdegree

• Suppose relation is seeks advice from
  • Outdegree = how many people you seek help from
  • Indegree = how many people seek you out
Closeness

• Out-closeness: total distance of a node to all others
  • Measure of how well/badly positioned a node is to diffuse things to others
• In-closeness: total distance to a node from all others
  • Measure of how well positioned a node is to receive things early

• Problems
  • In directed networks, many nodes can’t be reached from others, creating many undefined distances
• How to get row or col sums when you have undefined distances?

<table>
<thead>
<tr>
<th></th>
<th>HOLLY</th>
<th>BRAZEY</th>
<th>CAROL</th>
<th>PAM</th>
<th>PAT</th>
<th>JENNIE</th>
<th>PAULINE</th>
<th>ANN</th>
<th>MICHAEL</th>
<th>BILL</th>
<th>LEE</th>
<th>DON</th>
<th>JOHN</th>
<th>HARRY</th>
<th>GERY</th>
<th>STEVE</th>
<th>BERT</th>
<th>RUSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Betweenness

• With betweenness, there is no need for separate in and out versions
  • A node is between two others if it is along a directed path from one to the other

F gets no points for being between E and B, because there is no directed path from E to B
  • B has only outgoing arrows, so no way to get to B
Directed Eigenvector

• In principle, similar to degree:
  • Out-eigenvector (known as right eigenvector) gives a high score to those who send to many people who themselves send to many people who ...
    • If the relation is influences, then high score means you influence the influencers
  • In-eigenvector (left eigenvector) gives high score to those who receive from people who receive from many people who receive from ...
    • For the respects relation, a high score indicates you are respected by the well respected

• In practice, is often not calculable or gives wacky answers

\[ r_i = \frac{1}{\lambda} \sum_j a_{ij} r_j \quad \text{right} \]
\[ l_j = \frac{1}{\lambda} \sum_i a_{ij} r_i \quad \text{left} \]

(c) 2015 Stephen P Borgatti
Beta centrality (aka Bonacich power)

• Defined as: \( p = (I - \beta R)^{-1} R1 \)
  - \( R \) is the adjacency matrix; \((I - \beta R)^{-1}\) is a new matrix derived from \( R \)
  - \( \beta \) is a parameter chosen by the user

• When \(-1/\lambda < \beta < 1/\lambda\), where \( \lambda \) is largest eigenvalue of \( R \), \( p \) can be seen as the row sums of this sum of matrices:
  \[
  R^+ = b^0 R^1 + b^1 R^2 + b^2 R^3 + b^3 R^4 + \ldots
  \]
  \[
  P = R^+ 1
  \]
  - \( R^2 \) gives the number of walks of exactly 2 steps between every pair of nodes
  - \( R^3 \) gives the number of walks length 3 between all pairs of nodes, etc.

• Beta centrality measures # of walks of all lengths, weighted inversely by length, that emanate from a node

\( R^+ \) is no. of walks, wtd inversely by length, btw each pair of nodes
\( R^+1 \) is a column vector giving the sum of each row of \( R^+ \)
The $\beta$ parameter in beta centrality

- When $\beta$ is 0, beta centrality equals degree
  - Only paths of length 1 (direct connections) matter
- As $\beta$ increases from 0, longer paths are given increasing weight
- When $\beta$ is as close to $1/\lambda$ as possible, beta centrality equals eigenvector centrality
- When $\beta$ gets larger than $1/\lambda$, beta centrality becomes uninterpretable

\[ R^+ = b^0 R^1 + b^1 R^2 + b^2 R^3 + b^3 R^4 + \ldots \]
\[ P = R^+ 1 \]
Issues with beta centrality

• Often highly related to degree
• How to choose beta?
• But ... it works great with directed graphs
Directed beta centrality

• Beta centrality is the solution to the directed eigenvector problem.
  • When it is possible to compute eigenvector centrality, running beta centrality with $\beta \approx 1/\lambda$ gives same result
  • When it is not possible to compute eigenvector centrality, beta centrality is fine (except for disconnected graphs)

• Out-beta centrality
  • Score measures number of walks of all lengths emanating from a node, weighted inversely by length
  • Also indicates extent to which that the node sends to many nodes who themselves send to many nodes …

• In-beta centrality
  • Measures # of walks of all lengths that arrive at a node, weighted inversely by length
  • Also indicates extent to which node receives from nodes who themselves are targets
Beta centrality on difficult graphs

- For network I, with beta = 0.8, node 5 has the most power.
- In net III (with beta = 0.8), node a has most power followed by b, c, d and e in order.
An application of beta centrality
# Measures and type of network

<table>
<thead>
<tr>
<th>Graph</th>
<th>Degree</th>
<th>Eigenvector</th>
<th>Beta Centrality</th>
<th>Closeness</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undirected</td>
<td>Ok</td>
<td>Ok</td>
<td>Ok</td>
<td>Ok</td>
<td>Ok</td>
</tr>
<tr>
<td>Directed</td>
<td>Ok</td>
<td>Very problematic</td>
<td>Ok</td>
<td>Problematic&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ok</td>
</tr>
<tr>
<td>Valued</td>
<td>Ok</td>
<td>Ok</td>
<td>Ok</td>
<td>No&lt;sup&gt;b&lt;/sup&gt;</td>
<td>No&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Disconnected</td>
<td>Ok</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Ok</td>
</tr>
</tbody>
</table>

<sup>a</sup> only a problem because directed graphs are often disconnected -- have unreachable nodes  
<sup>b</sup> there are ways to do it in ucinet, but not commonly accepted  
<sup>c</sup> not possible in Ucinet, but in principle can be done easily with values that represent costs or distances