What are social networks? .......................................................... 1
A set of actors and the ties or relations among them.

Representations of social networks (Chapter 3 in W&F) ...................... 1

Forms of social network data (Chapter 2 in W&F) ................................. 3

Statistical Issues to consider .................................................... 6

Influence: How Actors’ Interactions Affect their Beliefs and Behaviors
......................................................................................................... 7
The Formal Model of Influence -- the Network Effect 8;
Toy Data (as found in SAS example) 10; Influence Exercise 11; Structure of W 12;
Estimation of D for cross-sectional data (not recommended) 14

Selection
How Actors Choose Others with whom to Interact .............................. 15
The Selection Model 16; Selection Exercise 17; Graphical Images of Selection
Processes 18; Estimation of 20 and 21 21

Clustering and Graphical Representations ........................................ 26
The goal: to identify patterns in the network 26; Theory for defining cluster
membership 26; Criteria for Determining Clusters 27; Result of clustering to
identify cohesive subgroups 29; Sampling Distribution for 1 30; Other Graphical
Representations 32

Centrality: The strength of the Connection between an Actor and the Network
........................................................................................................ 33

Centralization -- the Centrality of the System .................................... 35

Review in context of teacher’s use of computers ................................. 36

Confidentiality/Ethical issues in Collecting Network Data ....................... 37

Logistics of Data Collection .......................................................... 38

Organizing data entry .................................................................... 39

Resources for Networks ............................................................... 40
Books 40; Introductory On the Web 40; Web Clearinghouses 40; Individual Web
Pages 40

Network Software ....................................................................... 41

Four references (which contain most of the references above): .................. 41
What are social networks?

A set of actors and the ties or relations among them.

1) close colleagues (tie) among teachers (actors)
2) help (relation) one teacher (actor) provides to another
3) communication (tie) between people (actors) in an organization
4) friendships (tie) among politicians (actors)
5) links (tie) among web cites (actors)
6) referrals (relation) among social service agencies (actors)

Representations of social networks (Chapter 3 in W&F)

Sociomatrix or adjacency matrix (W&F) page 151

Friendships among the French financial elite

```

+-----------+
| 1 |......111. |
| 25 |..1....11. |
| 14 | .1.1.1.11.|
| 15 | ..1..1.11.|
| 4  | .........11.|
| 26 |..11...11.|
| 13 | 1......11.|
| 17 |11111111.11|
| 19 |11111111...|
| 20 |.......1..|
```

Notation:

- $x_{ij}$, takes a value of 1 if $i$ nominates $j$, 0 otherwise: $x_{1\, 25}=0$, $x_{1\, 13}=1$

Ken uses:

- $w_{ii'}$, takes a value of 1 if $i$ nominates $i'$, 0 otherwise: $w_{1\, 25}=0$, $w_{1\, 13}=1$

Graphical representation: sociogram
Figure 1: Sociocentric Data

Crystallized Sociogram: Friendships Among the French Financial Elite

Lines indicate friendships: solid within subgroups, dotted between subgroups
numbers represent actors
Rgt,Cen,Soc,Non = political parties;
B=Banker, T=treasury; E=Ecole National D’administration
Forms of social network data (Chapter 2 in W&F)

Directionality: if there is a difference between $i \in i'$ and $i' \in i$.
- Directed: liking, giving information to, referring clients
- Non-directed: economic exchange, discussion, physical distances

Representations:
- Arrow heads in sociograms, symmetry in matrix, double arrow $\overleftrightarrow{E}$ versus $E$ in notation, or $w_{ij} = 1$ if $i \in j$. Note that the letter $w$ is arbitrary. Sometimes the letter $x$ is used.

Multiple relations
- Interacting with regard to different issues, buying or selling different products, social service organizations sharing information, referring clients etc.

Valued relations
- Data indicate the extent to which $i$ and $i'$ engage in interactions, number of clients one agency refers to another, number of times users traverse a link from one web site to another.

Centricity (W&F, page 41)
- Sociocentric: the types of systems we have been describing, where alters and egos can be interchanged and alters may overlap

What is the effect of social support on people’s level of depression? (The people can be considered separately. Systemic level propositions not possible)

- Egocentric: each ego’s ties are considered independent of one another

What factors predict whether or not you receive social support from a particular friend or relative? (Wellman & Wortley)
Figure 2

Ego Centric Data: Alters Not Connected

Egos are in squares, alters represented by circles. A line indicates alter supports ego.
Also available at: http://www.chass.utoronto.ca/~wellman/publications/index.html

Bi-modal data
From: Identifying positions from affiliation networks: Preserving the duality of people and events In Press,
Sam Field, Kenneth A. Frank (equal first author), Kathryn Schiller, Catherine Riegle-Crumb and Chandra Muller
Compare with Faust’s chapter 7 in Carrington et al.
Statistical Issues to consider

Dependencies among observations
\( w_{12} \) not independent of \( w_{23}, w_{13} \)

Causality in cross-sectional data
Is it selection or influence?

Multiple levels

Pairs within nominators and nominees
Alters within egos
People within subgroups within organizations

Sample and population (?)
Influence: How Actors’ Interactions Affect their Beliefs and Behaviors

Examples of Research Questions

How does a teacher’s interactions affect her implementation of innovations?
How does a banker’s interactions affect her profitability?
How does an adolescent’s interactions affect her delinquency, alcohol use or engagement in school?

Theoretical Mechanisms

Normative/conformity
Abelson & Bernstein, 1976; Burt, 1987; Carley, 1989; Helbing, 1994; Friedkin 1994

Information
Carley, 1991; Coleman, 1964; Kaufer & Carley, 1993; Pfeffer, 1982

Dual processes

Evidence of Effects

Face-to-face contact
Homans 1950; Festinger, 1950; Mitchell, 1973

Written communication
Kaufer & Carley, 1993

E-mail exchanges
Freeman, 1986

In schools
Frank et al., 2004; Fuller & Izu, 1986; Grant, 1988; Hanson, 1978; Lightfoot, 1983; Rosenholtz & Simpson, 1990; Zeichner & Gore, 1989

http://www.nap.edu/books/0309089522/html/
The Formal Model of Influence -- the Network Effect


Define $w_{ii'}$ to indicate the extent or existence of a relation between individuals $i$ and $i'$ as perceived by $i$.

$y_i$ to represent an attitude or behavior, of actor $i$ that might be influenced through relations with others.

$E_i w_{ii'} y_i'$, the sum of the attributes of others to whom actor $i$ is related.

A model for $n$ people representing the influence of others on $y_i$ through interpersonal relations:

$$y_{it} = \rho \left[ \sum_{i'=1, \atop i' \neq i}^{n} w_{ii'} y_{i't-1} \right] + \gamma y_{i't-1} + e_{it} \quad .$$

(1)

Errors are assumed iid normal, with mean zero and variance ($\sigma^2$).

In words:

**use of computers time 2**

$$y_{2i} = \rho [use \ of \ first \ colleague \ time \ 1] + \rho [use \ of \ second \ colleague \ time \ 1] + \rho [use \ of \ last \ colleague \ time \ 1] + \gamma (use \ time \ 1)_i + error \ time \ 2_i$$

(2)
Influence Via Friendship on Academic Engagement

Friend AB \( (W_{ab}) \)
Friend AC \( (W_{ac}) \)
Friend AD \( (W_{ad}) \)

Prior for A
Toy Data (as found in SAS example)

person 3 interacts with 2, 4, and 6:

\[
y_{t3} = \rho \left[ \frac{(y_{t1})_2 + (y_{t1})_4 + (y_{t1})_6}{3} \right] + \gamma (y_{t1})_3 + \text{error}_{t3}. \tag{3}
\]

\(D = .15 \) and \( \gamma = .67 \). For person 3, \( y \) at time 1 = 1.2 and \( y \) at time 2 = 1.

Persons 2, 4, and 6, have levels of \( y \) at time 1 of 2.6, -.5, and -1 respectively

\[
1 = .13 \left[ \frac{2.6 - .5 - 1}{3} \right] + .68(1.2) + .21. \tag{4}
\]
Influence Exercise

Assume Bob talks to Sue with frequency 3, to Lisa with frequency 2 and not at all to Jane. Last year (at time 1), Sue’s delinquency behavior was a 10, Lisa’s was a 5 and Jane’s was 2.

1) What is the sum Bob’s exposure to delinquency through his peers?

\[ E_{i} = \sum_{i'} \left( w_{ii'} y_{i',t-1} \right) \]

the sum of the attributes of others to whom an actor is related.

\[ E_{i} = \sum_{i'} \left( w_{ii'} y_{i',t-1} \right) = \text{frequency that Bob talks to Lisa} \times \text{Lisa’s behavior at time 1} + \text{frequency that Bob talks to Sue} \times \text{Sue’s behavior at time 1} + \text{frequency that Bob talks to Jane} \times \text{Jane’s behavior at time 1} + \]

2) What is the mean of the influence of Bob’s peers? Hint (Mean=sum/n, but what should n be?)

3) Following the equation:

\[ y_{it} = \rho \left[ \sum_{i' \neq i}^{n} w_{ii'} y_{i',t-1} \right] + \gamma y_{i,t-1} + e_{it} \]  

\[ \text{In words, what are } D \text{ and } \gamma? \]

4) Write down a model including separate terms for influence from within versus from outside a subgroup

5) Write down a model that including separate terms for influence of others’ attitudes and others’ behaviors

SOFTWARE: http://www.msu.edu/~kenfrank/software.htm
Structure of W

Cohesion -- direct connections/communication

Examples:
Students’ educational and aspirations decisions are influenced through direct discussions

Adolescents’ delinquency is influenced by the delinquency of their friends (Haynie 2004).

Structural Equivalence -- common roles/comparison & comparison
(Burt, 1987; Merton, 1957; Nadel, 1957; Radcliffe-Brown, 1940).

Examples
students who occupy similar positions defined by curricular tracks may develop similar educational aspirations
(Bowles & Gintis, 1976; Hansell & Karweit, 1983).

Businesses who sell to similar others may adopt similar practices
(Burt 2000)

Direct Influence versus Indirect Influence (Leenders)
Are you influenced by those who you do not talk to, but with whom you share intermediaries?

http://www.sciencedirect.com/science?_ob=JournalURL&_cdi=5969&_auth=y&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=0dbd43b8d4784bc1532be7b6c056be81
A Redundant effect of $e_{i't}$ on $y_{it+2}$

$e_{i't} \rightarrow y_{i't} \rightarrow y_{i't+1} \rightarrow y_{i't+2}$

$(\rho w_{i'i't \rightarrow t+1}) \quad (\gamma) \quad (\rho w_{i'i't+1 \rightarrow t+2})

y_{i't+1}$

Hypothesis: behaviors are more influenced by competition & comparison, beliefs more influenced by communication

Interpretation of $D$ from cross-sectional data
not necessarily an “effect” – descriptive
(Leenders 1995; Marsden and Friedkin 1994, page 13)
Estimation of $D$ for cross-sectional data (not recommended)

Define
\[ y \text{ to be an } n \times 1 \text{ vector of attributes,} \]
\[ W \text{ an } n \times n \text{ matrix representing relations,} \]
\[ e \text{ an } n \times 1 \text{ vector of error terms:} \]

Then in matrix notation, the model is:
\[ y = \rho Wy + e. \]

Challenge: $y$ on both sides!

Ordinary least squares estimates of $D$ are biased (Anselin, 1988; Ord, 1975).

Maximum Likelihood (Ord 1975)
Write the likelihood for $y$ and transform to $e$

\[ y - \rho Wy = e = (I - \rho W)y = e. \quad (7) \]

Perfect fit is if $y$ is an eigen vector of $W$. Therefore can get big values of $D$ with perfect fit if $y$ aligns with a minor eigen vector of $W$. That’s bad. So $1/\text{max} < D < 1/\text{min}$. This also makes sure likelihood is defined.

See dissertation by JiQiang Xu at MSU for one estimation solution.

But cannot differentiate selection from influence from cross-sectional data. $D$ is overestimated.
Selection
How Actors Choose Others with whom to Interact

Examples of Research Questions

How do teachers decide to whom to provide help?
How do bankers decide to whom to loan money?
How do social service agencies choose other agents to refer clients to?

Theoretical Mechanisms (references from Frank & Fahrbach 1999)

Balance seeking/homophily -- seeking to interact with others like yourself

Information seeking
  Goal oriented
    Alchian & Demsetz, 1972; Burt, 1992; Katz & Kahn, 1978
  Reduce uncertainty
    Radner, 1986
  Power oriented
    Burns & Stalker, 1961; Pettigrew, 1972

Better understanding
  Rosen, 1961

Curiosity
  Freedman, 1965

Inoculate
  McGuire & Papageorgis, 1962

Evidence of Effects

Adolescents select friends who are like themselves
Matsueda and Anderson; Moody 2002

Teachers who want to be innovative interact with other innovators
Darling-Hammond & McLaughlin, 1995; Lieberman, 1995; Zeichner & Gore, 1989

The Selection Model

Generic (ignoring time sequence)

\[
\log \left( \frac{p[w_{ii'}=1]}{1-p[w_{ii'}=1]} \right) = \theta_0 + \theta_1 [(-1)|y_i - y_{i'}|].
\]

\(w_{ii'}\) represents whether \(i\) and \(i'\) are related (e.g., talked, close colleagues, etc);
\(=1\) if yes, \(0\) if no.

\(y_i\) represents a characteristic of person \(i\) (e.g., gender)

\(|y_i - y_{i'}| = \text{absolute value of difference}. \text{ Big difference implies large } |y_i - y_{i'}|. \text{ similarity } = (-1) |y_i - y_{i'}|. \)

\(\theta_1\) represents the effect of similarity in the above selection model

\(\theta_0\) represents an intercept, or baseline density
Selection Exercise

1) Write a model of whether two actors talked as a function of whether they are the same race and whether they are the same gender.

\[ w_{ii'}, \text{ represents whether } i \text{ and } i' \text{ talked,} \]
\[ y_i \text{ represents the gender of } i \text{ (0 if male, 1 if female), and} \]
\[ z_i \text{ represents the race of } i \text{ (0 if white, 1 if African American)} \]
(You’ll need one term for effects associated with gender, and another for race)

2) Assume Bob that and Lisa are African American and that Jane and Bill are white. Bill and Bob are Male and Lisa and Jane are female.

Calculate the independent variables based on similarity of race and gender for Bob with each of his interaction partners:

(Bob, Lisa): same gender = _______; same race =________
(Bob, Jane): same gender =________; same race =________
(Bob, Bill): same gender =________; same race =________

3) Assuming the values of the \(2\)'s are positive and that the effect of race is stronger than that of gender, who is Bob most likely to talk to?

4) Include a term capturing the interaction of similarity of race and gender
Graphical Images of Selection Processes

---

**Expansiveness**
Characteristics of nominator (i):
age, gender, gregariousness

**Attractiveness**
Characteristics of nominee (i')
Age, gender, charisma

**Commonality**
Characteristics of both (i,i'):
same age, same gender, similar beliefs

---

**Tie to be modeled**
---

**Reciprocity**
---

---

**Stars**

**Triads (cycle)**

---

---

Question: which ties are dependent on which other ties?

Answers:

P1: only the other part of the dyad


P2: Ties involving same nominator (sender) or nominee (receiver):

(cross-nested multilevel models)


SOFTWARE [http://stat.gamma.rug.nl/stocnet/](http://stat.gamma.rug.nl/stocnet/)


p*: Only those that share a common actor: triads and mstars


[http://kentucky.psych.uiuc.edu/pstar/index.html](http://kentucky.psych.uiuc.edu/pstar/index.html)

*See Wasserman and Robbins’ chapter 8 in Carrington et al.*

Neighborhoods: those in a common neighborhood – sharing intermediate ties:


All are part of Exponential Random Graph Models (ERGM)


[http://stat.gamma.rug.nl/sniijders/](http://stat.gamma.rug.nl/sniijders/)
Estimation of $2_0$ and $2_1$

Use the example of $w_{ii'}$ being whether one teacher helped another during the interval $t-h$ to $t$: $Help_{ii' \mid t-h \rightarrow t}$ represents whether teacher $i'$ provided Help to $i$ during interval $t-h\Xi t$.

Naive: Help is dichotomous, so, use logistic regression:

$$\log \left( \frac{p[Help_{ii' \mid t-h \rightarrow t} = 1]}{1 - p[Help_{ii' \mid t-h \rightarrow t} = 1]} \right) = \theta_0 - \theta_1 |y_{i \mid t-h} - y_{i' \mid t-h}|.$$  

(9)

Similarity of attributes captured by $-|y_{i \mid t-h} - y_{i' \mid t-h}|$.

Likelihood function: $p(A \text{ and } B) = p(A) \times p(B)$ if $A \text{ and } B$ are independent. NO! $Help_{ii'}$ is not independent of $Help_{ii'}$!

2) Control for Dependencies through $p^*$:

$$\left( \frac{P[Help_{ii'} = 1 \mid \text{all other Helpful (or absence of Helpful) acts in the network}]}{1 - P[Help_{ii'} = 1 \mid \text{all other Helpful (or absence of Helpful) acts in the network}]} \right) = \theta_0^* + \theta_1(\text{Same Gender}_{ii'}) + \theta_2(\text{Same Grade}_{ii'}) + \theta_3(\text{Close Colleagues}_{ii'}) + \Theta^*_x.$$  

(10)

Vector of characteristics might includes

- number of others, $iO$ who helped both $i$ and $i'$ ($iO\cap i\cap i'$),
- number of others, $iO$ who were helped by $i$ and helped $i'$ ($i \in iO \cap \cap i'$), etc.

Vector may also include tendency for $i'$ to provide help (in terms of the number of others whom $i'$ helped), and the tendency for $i$ to receive help (in terms of the number of others from whom $i$ received help). But these characteristics may control out the effects of individuals, such as identification with collective, etc., reducing them to individual differences.
p2 via Multilevel Models

Level 1, the pair $i'i'$:

$$
\log \left( \frac{P[\text{Help}_{i'i'} t-h \rightarrow i^1]}{1-P[\text{Help}_{i'i'} t-h \rightarrow i^1]} \right) = \\
\theta_0i'i' + \theta_{0i} + \\
\theta_1 i'i' \text{ or } i \ (\text{Close Colleagues}_{i'i'}) + \\
\theta_2 i'i' \text{ or } i \ (\text{Same Grade}_{i'i'}) + \\
\theta_3 i'i' \text{ or } i \ (\text{Same Gender}_{i'i'}) +
$$

(11)

Level 2a, nesting within the provider, $i'$

$$
\theta_{0i'} = \gamma_{00}^{(i')} + \gamma_{01}^{(i')} \text{identification with collective}_{i'} \text{,} + u_{0i'} \\
\theta_{1i'} = \gamma_{10}^{(i')} + \gamma_{11}^{(i')} \text{identification with collective}_{i'} \text{,} + u_{1i'} \\
\theta_{2i'} = \gamma_{20}^{(i')} + u_{2i'} \\
\theta_{3i'} = \gamma_{30}^{(i')} + u_{3i'}
$$

(12)

Level 2b, nesting within the receiver, $i$

$$
\theta_{0i} = \gamma_{00}^{(i)} + \gamma_{01}^{(i)} \text{identification with collective}_{i} + \nu_{0i} \\
\theta_{1i} = \gamma_{10}^{(i)} + \gamma_{11}^{(i)} \text{identification with collective}_{i} + \nu_{1i} \\
\theta_{2i} = \gamma_{20}^{(i)} + \nu_{2i} \\
\theta_{3i} = \gamma_{30}^{(i)} + \nu_{3i}
$$

(13)

The term $u_{0i'}$ accounts for the unique effect of actor $i'$ to provide help. But not all $u_{0i'}$ are estimated. They are considered random, $N(0, \psi)$. Therefore only $\psi$ needs to be estimated. Standard errors for effects such as $(\psi_{02}$ are then determined relative to $\psi$ with degrees of freedom roughly based on the number of providers of help, $i'$. 

Running p2:

Using Van Duijn’s p2:
go to: http://stat.gamma.rug.nl/stocnet/
go to downloads and save stocnet in c:\stocnet (follow directions if you install somewhere else).

Unzip into c:\stocnet
run stocnet.exe

Manual available @ http://stat.gamma.rug.nl/stocnet/downloads/manualp2.pdf

To run:

[remember to keep clicking APPLY]

file/new session
add network data s50-network1.dat
add actor data s50-alcohol.dat

select model/p2
put network1 (s50-network1) into digraph
put file1 (s50-alcohol.dat) into attributes

specify model with actor attributes on network parameters
apply
run.
Example output for p2 for Toy data

P2 IGLS
toylb.out

September 14, 2005, 5:01:52 PM

General Information:

Digraph: C:\stocnet\temp\~toylb.dat
September 14, 2005, 5:01:52 PM

Number of valid tie indicator observations: 30

Convergence criterion: 0.0001 reached after 14 iterations.

Random effects:

<table>
<thead>
<tr>
<th>parameter</th>
<th>standard estimate</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>sender variance</td>
<td>0.1749</td>
<td>0.9372</td>
</tr>
<tr>
<td>receiver variance</td>
<td>1.8614</td>
<td>1.5930</td>
</tr>
<tr>
<td>sender receiver covariance</td>
<td>-0.5163</td>
<td>0.8605</td>
</tr>
</tbody>
</table>

Fixed effects:

Overall effects:

<table>
<thead>
<tr>
<th>parameter</th>
<th>standard estimate</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>2.1326</td>
<td>2.0955</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>1.4799</td>
<td>1.7894</td>
</tr>
</tbody>
</table>

Overall covariate effects:

Overall effects of covariates including diff and absdiff manipulations.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Wald test statistic</th>
<th>df</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute1</td>
<td>3.9484</td>
<td>1</td>
<td>0.0469</td>
</tr>
</tbody>
</table>

Specific covariate effects:

Density covariates:

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Wald test statistic</th>
<th>df</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>abs_diff_Attribute1</td>
<td>-1.5300</td>
<td>0.7700</td>
<td></td>
</tr>
</tbody>
</table>

This last term models whether difference in attribute 1 predicts density.
Alternatives for running p2

In sas:
   download Sam Field’s p2 via sas from my web site:
   download glimmix from my web site and save to c:\
   run glimmix.sas in sas
   run Sam’s program (p2_explore.sas)
   Note it generates its own ego and alter files (see data i and data j) and network data (a5), but these could be read in.

Can also do using Peter Hoff’s R routine
http://cran.cnr.berkeley.edu/
Clustering and Graphical Representations

*The goal: to identify patterns in the network*

rearrange rows and columns of social network matrix to reveal clustering
plot actors and ties in two dimensions to reveal clustering

*Theory for defining cluster membership*

cohesion (clusters are called subgroups): an actor should be in a cluster if the actor has demonstrated a preference for engaging in ties with members of the cluster. *Result:* ties are concentrated within subgroups

structural equivalence (blocks): an actor should be in a cluster if the actor engages in a similar pattern of ties as members of that cluster. *Result:* blocks represent positions, but ties not necessarily concentrated within blocks.
Criteria for Determining Clusters

Structural Equivalence:
Factor analyze sociomatrix (Katz & Kahn)
iteratively rearrange and revalue rows and columns (CONCORR -- White el al., 1976)

Cohesion
utilize fixed criteria (e.g., must be connected to at least \( k \) others in clusters, or
must be minimal path length from \( k \) others, etc).

use flexible criterion -- preference relative to group sizes and number of ties:

Model based cohesion

\[
samegroup_{iN} = \begin{cases} 
1 & \text{if actors } i \text{ and } iN \text{ are members of the same subgroup,} \\
0 & \text{otherwise.}
\end{cases}
\]

Then \( Z_1 \) represents subgroups salience:

\[
\log \left( \frac{P[w_{ii'} = 1]}{1 - P[w_{ii'} = 1]} \right) = \theta_0 + \theta_1 samegroup_{ii'} \quad . \quad (14)
\]

So ...... Maximize \( Z_1 \) (odds ratio)
## Odds Ratio for Association Between Common Subgroup Membership and The Occurrence of Ties Between Actors

<table>
<thead>
<tr>
<th>Subgroup Membership</th>
<th>Tie Occurring</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No $w_{iiN} = 0$</td>
<td>Yes $w_{iiN} = 1$</td>
<td></td>
</tr>
<tr>
<td>Different 0</td>
<td>A</td>
<td>B</td>
<td>$n(n-1) - E_g n_g (n_g - 1)$</td>
</tr>
<tr>
<td>Same 1</td>
<td>C</td>
<td>D</td>
<td>$E_g n_g (n_g - 1)$</td>
</tr>
</tbody>
</table>

$E_i N$ represents the number of actors in subgroup $g$.

Odds ratio = $AD/BC = (\text{Absence of ties outside of subgroups}) \times (\text{presence of ties within subgroups})$

(Presence of ties outside of subgroups) * (absence of ties within subgroups)
Table 1
Partitioned Friendships Among the French Financial Elite

<table>
<thead>
<tr>
<th>N</th>
<th>Group And Actor Id</th>
<th>Group ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>AAAAAAAA</td>
<td>BBBB</td>
</tr>
<tr>
<td></td>
<td>11222222</td>
<td>1111</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>1 A</td>
<td>8 A....1.11</td>
<td>1....1</td>
</tr>
<tr>
<td>1 A</td>
<td>9 A....1..11</td>
<td>1....1</td>
</tr>
<tr>
<td>1 A</td>
<td>12 A....1111</td>
<td>.......</td>
</tr>
<tr>
<td>1 A</td>
<td>19 A....1..11</td>
<td>.......</td>
</tr>
<tr>
<td>1 A</td>
<td>20 A....1..11</td>
<td>.......</td>
</tr>
<tr>
<td>1 A</td>
<td>21 A....1..11</td>
<td>.......</td>
</tr>
<tr>
<td>1 A</td>
<td>24 A....1..11</td>
<td>.......</td>
</tr>
<tr>
<td>1 A</td>
<td>25 A....1..1A1</td>
<td>.......</td>
</tr>
<tr>
<td>1 A</td>
<td>26 A....1..11</td>
<td>.......</td>
</tr>
<tr>
<td>1 A</td>
<td>27 A....1..11</td>
<td>.......</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>2 B</td>
<td>2 B....11..1</td>
<td>B1....1</td>
</tr>
<tr>
<td>2 B</td>
<td>5 B....11..1</td>
<td>B1....1</td>
</tr>
<tr>
<td>2 B</td>
<td>6 B....11..1</td>
<td>B1....1</td>
</tr>
<tr>
<td>2 B</td>
<td>10 B....11..1</td>
<td>B1....1</td>
</tr>
<tr>
<td>2 B</td>
<td>13 B....11..1</td>
<td>B1....1</td>
</tr>
<tr>
<td>2 B</td>
<td>14 B....11..1</td>
<td>B1....1</td>
</tr>
<tr>
<td>2 B</td>
<td>17 B....11..1</td>
<td>B1....1</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>3 C</td>
<td>1 C....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>3 C</td>
<td>4 C....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>3 C</td>
<td>15 C....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>3 C</td>
<td>16 C....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>3 C</td>
<td>18 C....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>3 C</td>
<td>22 C....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>3 C</td>
<td>28 C....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>4 D</td>
<td>3 D....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>4 D</td>
<td>7 D....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>4 D</td>
<td>11 D....11..1</td>
<td>.......</td>
</tr>
<tr>
<td>4 D</td>
<td>23 D....11..1</td>
<td>.......</td>
</tr>
</tbody>
</table>
Sampling Distribution for $\hat{z}_1$

Clustering algorithms will always identify clusters, but are ties more concentrated within subgroup boundaries than is likely to occur by application of the algorithm to random data?

Define $\hat{z}_{1\text{base}}$ as size of $\hat{z}_1$ when algorithm applied to random data

Compare models

$$\log \left( \frac{P[w_{ii'} = 1]}{1-P[w_{ii'} = 1]} \right) = \theta_0 + \theta_{1\text{ base}} \text{ samegroup}_{ii'} \quad . \quad (15)$$

Versus

$$\log \left( \frac{P[w_{ii'} = 1]}{1-P[w_{ii'} = 1]} \right) = \theta_0 + \theta_{1\text{ base}} \text{ samegroup}_{ii'} + \theta_{1\text{ subgroup processes}} \text{ samegrou} \quad (16)$$

Output

PREDICTED THETA (1 base) BASED ON SIMULATIONS.
VALUE BASED ON UNWEIGHTED DATA.

0.65962

ESTIMATE OF THETA (1 subgroup processes)

0.43653

THE TOTAL THETA1 IS:

1.09616

Thus

$$\log \left( \frac{P[X_{ii'} = 1]}{1-P[X_{ii'} = 1]} \right) = \theta_0 + .66 \text{ samegroup}_{ii'} + .44 \text{ samegroup}_{ii'} \quad . \quad (17)$$

LRT BASED ON PREDICTED THETA (1 base)

17.92620

COMPARE TO CHI-SQUARE ON 1 DF
P-VALUE (LESS THAN OR EQUAL TO):

0.00100
Did the Algorithm Recover the Correct Subgroups?

Many algorithms search for optimal subgroups. KliqueFinder does not, but how different are the subgroups it finds from the optimal or known subgroups?

Output

PREDICTED ACCURACY: LOG ODDS OF COMMON SUBGROUP MEMBERSHIP, + OR - .5734 (FOR A 95% CI)

1.6767

The Log odds applies to the following table:

<table>
<thead>
<tr>
<th>OBSERVED SUBGROUP</th>
<th>DIFFERENT</th>
<th>SAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIFFERENT</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>KNOWN SUBGROUP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIFFERENT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAME</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

THE LOGODDS TRANSLATES TO AN ODDS RATIO OF

5.3480

WHICH INDICATES THE INCREASE IN THE ODDS THAT KLIQUEFINDER WILL ASSIGN TWO ACTORS TO THE SAME SUBGROUP IF THEY ARE TRULY IN THE SAME SUBGROUP.
Other Graphical Representations
See Freeman’s chapter 12 in Carrington et al.

*Multidimensional Scaling (netdraw in ucinet)*

*Correspondence analysis (for two mode data in ucinet)*

*3-D graphics*  ([http://www.heinz.cmu.edu/project/INSNA/gal_inf.html](http://www.heinz.cmu.edu/project/INSNA/gal_inf.html))
Centrality: The strength of the Connection between an Actor and the Network

*Freeman (1978/79) direct connections*


Degree: number of ties to node $i$

Betweenness: proportion of geodisics (connecting paths) between $j$ and $k$ that go through $i$.

Closeness: total number of edges required to link $i$ to all others

*Bonacich (1972): eigen vector*


The centrality of a given person ($e_i$) depends on the centrality of the people to whom the person is tied ($w_{ii'}=1$ if $i$ and $i'$ are related, 0 otherwise):

$$
\lambda e_i = \sum_{j=1}^{n} w_{ii'} e_j 
$$

$$
\rightarrow \lambda e = We 
$$

The elements in $e$ then represent the components of the eigen vector of $W$. 
Bonacich Centrality Revised

(AJS, 1986):

\[ c_i(\alpha, \beta) = \sum_{j=1}^{n} (\alpha + \beta c_j)w_{ij} \]

\[ \rightarrow c(\alpha, \beta) = \alpha (I - \beta W)^{-1} W1 \]

where 1 is a column of ones. When \( \beta < 1 / \beta_{\text{max}} \):

\[ c(\alpha, \beta) = \alpha \sum_{k=0}^{\infty} \beta^k W^{k+1} = \alpha (W1 + \beta W^2 + \beta^2 W^3 + ...) \]

Centrality as a function of direct (W1) and indirect (W2, W3 ...) effects.

When \( \beta > 0 \), centrality is like old measure, accounting for the centrality of people you are connected to. When \( \beta < 0 \), power is a function of “bargaining position” -- see Burt’s constraint. Be someone who brokers between others who have few choices.

Centralization -- the Centrality of the System

How does the pattern of communication in organization A differ from that in organization B, and how are these patterns formed by characteristics external to the organization?

Freeman: distribution of centrality
Compare measures against the maximal measure in the graph
-- but what if there is more than one actor who is highly extreme in centrality?


\[
W = \begin{bmatrix}
A & B & C \\
A & 9 & 9 \\
B & 9 & 1 \\
C & 9 & 1 \\
\end{bmatrix}
\]

Actor 1 is “too” central. The data cannot be easily represented in euclidean space -- cannot be easily resolved into a small number of dimensions. There is a negative eigen root (eigen root=eigen value\(^2\)), and a complex eigen value.

Measure of centralization:

Warp=(sum of positive eigen roots)/(sum of all eigen roots).

If system can be neatly represented in Euclidean space -- if no one actor is “overly” central - - then no negative eigen roots, warp=1. To the extent that one or more actors are overly central the warp will be >1.

Question: use \(W\) or \(W'W\)

**Warp is produced by KliqueFinder**
Review in context of teacher’s use of computers

Selection: Who chooses to be colleagues with whom, who helps whom with computers?

Influence: How do teachers influence each other’s use of computers?

Graphical representation: How is diffusion structured by subgroups?

Centrality: Do more central teachers have greater access to expertise?

Centralization: Which social structures best convey diffusion?
Confidentiality/Ethical issues in Collecting Network Data

Need names on survey

Data can be confidential but not anonymous (especially for longitudinal)

(All issues of social networks available via science direct)

Who benefits from network analysis? Who bears the cost?


Issues to raise when dealing with Human Subjects Board:

Klovdahl, Alden S. Social network research and human subjects protection: Towards more effective infectious disease control *Pages 119-137*

Hint on Human Subjects boards: they like precedents. Once you have one network study accepted, refer to it when submitting others!
Logistics of Data Collection

Need for longitudinal data to disentangle selection from influence
(Matsueda and Anderson 1998; Leenders 1995).

Time constraints: how long does a network question take?
Without roster: 2-3 minutes
With roster: 5-10 minutes (depending on size of network)

High response rates (70% or more) needed to characterize system, influence
incentives: school, individual
administer in collective settings (e.g., staff meeting)
do not be perceived to be affiliated with principal

Network data without survey?
Sensors
Participation in events (two-mode)
on-line e-mails
web links

Marsden in Carrington et al., follow up on
Organizing data entry
check out: http://www.classroomsociometrics.com/

Your name Lisa Jones (person 1)

Please list your closest colleagues (*network 1*) at xxx and the frequency with which you interact with each person.

<table>
<thead>
<tr>
<th>Name</th>
<th>Yearly</th>
<th>Monthly</th>
<th>Weekly</th>
<th>Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bob Jones</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2. Sue Meyer</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Data entered (chooser, chosen, frequency)
1 2 2
1 3 4

Your name Bob Jones (person 2)

Please list your closest colleagues (*network 1*) at Hueco and the frequency with which you interact with each person.

<table>
<thead>
<tr>
<th>Name</th>
<th>Yearly</th>
<th>Monthly</th>
<th>Weekly</th>
<th>Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lisa Jones</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2. Lin Freeman</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Data entered (chooser, chosen, frequency)
2 1 2
2 4 4
Resources for Networks

Books

Peter J. Carrington, John Scott, Stanley Wasserman “Models and Methods in Social Network Analysis” Cambridge, order from Amazon on-line.


Introductory On the Web

Borgatti’s slide show: http://www.analytictech.com/networks/intro/index.html

David Knoke’s intro to social network methods: http://www.soc.umn.edu/%7Eknoke/pages/SOC8412.htm


Jim Moody’s course: http://www.sociology.ohio-state.edu/jwm/

Web Clearinghouses

International social network analysis web page: http://www.insna.org/


Individual Web Pages:

Phil Bonacich http://www.sscnet.ucla.edu/soc/faculty/bonacich/home.htm

Ron Breiger (http://www.u.arizona.edu/~breiger/):

Ronald Burt (google Ron Burt):
http://portal.chicagogsb.edu/portal//server.pt/gateway/PTARGS_0_2_332_207_0_43/http%3B/portal.chicagogsb.edu/Facultycourse/Portlet/FacultyDetail.aspx?&min_year=20044&max_year=20063&person_id=30400

Ken Frank http://www.msu.edu/~kenfrank/
Linton Freeman http://moreno.ss.uci.edu/lin.html
James Moody http://www.sociology.ohio-state.edu/jwm/
Tom Snijders http://stat.gamma.rug.nl/snijders/
Barry Wellman: http://www.chass.utoronto.ca/~wellman/

Network Software

Huisman and Van Duijn chapter 13 in Carrington et al

Social Networks web site
http://www.insna.org/INSNA/soft_inf.html

Four references (which contain most of the references above):


