SOCIAL NETWORK ANALYSIS
WITH APPLICATIONS TO
MEDICINE AND HEALTH POLICY

James O’Malley, Ph.D.
Department of Biomedical Data Science
The Dartmouth Institute for Health Policy and Clinical Practice
Geisel School of Medicine at Dartmouth
Email: James.OMalley@Dartmouth.edu
Outline

• Definition of a social network (5 min)
• Examples of networks (5 min)
• Overview of types of network analyses (10 min)
• Core topics
  1. Network descriptive statistics (20 min)
     • Application: Comparison of multiple networks of organizations or regions
  2. Statistical analysis of relational data (“sociology”) (30 min)
  3. Statistical analyses of individual outcomes given network (“medicine”) (10 min)
• Questions interspersed (≈ 10 min)
Definition of a social network

- A social network consists of one or more sets of actors—also known as “units,” “nodes,” or “vertices”—together with the possibly directed relationships or social ties among them.
Components of a social network

• **Actors:**
  - Individual persons (e.g., patients or clinicians)
  - Organizations (e.g., hospitals)
  - Health states (e.g., diseases)
  - Work products (e.g., academic papers)

• **Social ties:**
  - Communication
  - Influence
  - Trust or affect (e.g., friendship)
  - Affiliations (e.g., co-authors)

• **Attributes:**
  - Actors, relationships, or both
Example networks

- Physician network within a health clinic
- Physician network within a health referral region
- Hospital network in health referral regions
- Biological networks
- Human disease networks
- International trade networks
- Networks of congress based on sponsorship of bills
- Honduran villages
- School classroom networks
- Company networks in Japan
- Social network in Framingham Heart Study
Physician social network within clinic

Spring embedder algorithm determines positions of actors (Fruchterman and Reingold 1991)

Network of physicians in a Boston hospital (O’Malley and Marsden, 2008)
Networks of Physicians in 4 Health Referral Regions (Landon et al 2012, JAMA)
Network of Hospitals in Two Adjourning Health Referral Regions (Moen et al, 2015)

Gary, IN (181) – blue
South Bend, IN (187) – green

Ties: physicians treat two or more relevant patients
Biological network (genes, proteins)

Gene network

Adapted from: Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)
Human Disease Network (edges reflect shared genes)

Adapted from: Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)
Layers of Networks in Medicine

World trade networks

Social network: Spread of Obesity in Framingham Heart Study (Christakis and Fowler, 2007)
Overview: Three “Important” Problems in Social Network Analysis

I. Do social network characteristics correlate with other variables of interest
   • Example, do social network characteristics of a health care organization correlate with utilization, quality and cost?

II. What factors affect the formation/dissolution of relationships; structure of society
   • Examples, similarity of characteristics (homophily), reinforcement of relationships

III. Do individuals influence one another
   • Example, diffusion of treatment tendencies or innovations across physicians (Coleman, 1957, 1966)
Preview I: Network is observational unit

- Network determines the outcomes, the predictors, or both
- Example: Egocentric network analysis
- Reasonable to think of the units on which networks are evaluated as independent
- For each network, generate summary measures that are used as either the outcome(s) or predictor variable(s) in a subsequent analysis involving other variables
- “Standard” statistical analysis; need to account for clustering if units observed multiple times

\[ Y_1, X_1 \quad Y_2, X_2 \quad \ldots \quad Y_n, X_n \]
Preview II: Analysis of network structure

- Observed network is the outcome
- Often only observe network once (cross-sectional data)
  - Multiple observations (e.g., longitudinal data) becoming more common
- Are global network properties explained by local configurations or sub-networks?
  - Closed dyads: reciprocity
  - Closed triads: transitivity, 3-cycles, other …
- Are individuals with particular characteristics more likely to form ties (homophily, assortative mixing, social selection)?
- Do (latent) communities underly the network?
- Non-standard and challenging statistical analyses required!
Preview III: Social influence analysis

- Network defines predictors
- Do individuals influence one another (``social influence'')
- Example, adoption of a new medical technology
- **Endogenous peer effects**: Does the behavior of my peers affect my own behavior?
- **Exogenous peer effects**: Does the treatment received by my peers affect my outcome (above and beyond my treatment)?
  - Violates stable unit treatment value assumption (SUTVA)

\[ Y, X = Y_{\text{peer}}, X_{\text{peer}} \]
Preview III: cont.

• Overlapping groups of individuals yield the predictor(s) of individuals’ outcomes!

• Complicated simultaneous equations model

• Longitudinal data helps with identification of causal effects
  • Reverse causality, simultaneity, …
  • Avoids reliance on strong parametric assumptions

• Statistical analysis challenging if seek causal claim when dyads not formed at random!
I. Describing Networks

- Size
- Components; connectivity
- Degree distribution; power law?
- Centrality
  - Eigenvalue (recursive definition; Page Rank)
  - Betweenness
  - Closeness
- Centralization = \text{var}(\text{centrality measure})
- Reciprocity
- Clustering
- Transitivity
- Many others (K-coreness, similarity, structural balance …)
Notation

- $Y_i$ and $X_i$ denote an outcome and predictor trait of individual $i$
- $A_{ij}$ denotes the relationship from actor $i$ to actor $j$
- The array of relationships forms an adjacency matrix, denoted $A$
  - By convention $\text{diag}(A) = 0$
  - In a non-directed network, $A = A^T$
- In a **weighted network** we often distinguish between ties (relationship versus no relationship) and relationship strength
  - Binary network measures relevant to a weighted network
  - Some binary network measures have weighted counterparts

Not today!
Binary Relational Data

- Letters = actors; numbers = relationship status
- Binary (1, 0) network of N actors:
  - Dyads (pairs of actors) have $2^2 = 4$ possible states
  - $N(N-1)/2$ dyads
  - $4^{N(N-1)/2}$ possible networks

Digraph

\[ \begin{array}{ccccccc}
A & B & C & D & E & F & G \\
\hline
A & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
B & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
C & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
D & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
E & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
F & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
G & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array} \]
Network Statistics (Directed network)

- **Size:**
  - Number of nodes: $N$
  - Number of edges: $L = \sum_{i \neq j} A_{ij}$

- **Density:** Ratio of edges to possible edges
  - $\mu = L/((N(N - 1))$

- **Degree of actor $i$**
  - Out-degree: $d^\text{out}_i = \sum_{j \neq i} A_{ij}$
  - In-degree: $d^\text{in}_i = \sum_{j \neq i} A_{ji}$

- **Mutual or reciprocated tie indicator:** $M_{ij} = A_{ij}A_{ji} = M_{ji}$
  - Number of mutual dyads: $M = \sum_{i<j} A_{ij}A_{ji}$
Degree distributions for $N = 33$ physicians in a medical practice

E.g., Edge denotes occurrence of important discussions about women's health issues
Properties of density and degree

- Average out-degree = average in-degree
- Average-degree = N*Density
- Corr(out-degree, in-degree)
  - Are expansive people popular?
- ≠ preferential attachment (popular actors forming ties with each other), a between actor phenomena
- Degree is a measure of an actors centrality
Power law distribution

- In network science there is extensive interest in whether the degree distribution of a network follows a power law.
- Under a power law, the probability actor $i$ has degree $k$:
  $$\Pr(\text{Degree}_i = k) \propto k^{-\lambda}$$
  for a parameter $\lambda$, typically $\lambda > 2$.
- Derived from a model of how networks are created:
  - Probability a new actor forms a tie with a given actor is proportional to the degree of that actor.
  - “Generative model”
- Power-law distributions/networks are also known as “scale-free distributions/networks” (Barabasi and Albert 1999, Newman 2010).
Centralization

- Heterogeneity of actors’ network prominence; extreme cases:

- Degree centralization measures:
  - $\text{StdDev}(\{d_i\}_{1:N})$ or $\text{Var}(\{d_i\}_{1:N})$
  - $C_D = \frac{\sum_i (\max_i d_i - d_i) - d_{max}}{(N-1)(N-2)}$, where $d_{max}$ is the maximum observed degree
  - Standardizing by maximum aids comparisons between networks
Three other measures of centrality

• Closeness centrality
  • The inverse of the mean of the geodesic distances from a given actor to all others

• Betweenness centrality
  • The proportion of times an actor is in an intermediary position along the geodesic paths linking pairs of other actors

• Eigenvector centrality
  • Assumes connections to central actors indicate greater prominence than similar-strength connections to peripheral actors
  • Used by Google Page Rank
  • Given by elements of eigenvector of largest eigenvalue of adjacency matrix

• Averaging over actors → network-level centrality; variability over actors → network centralization
Illustration of (betweenness) centrality

Bigger node implies greater centrality
Paths and Components

- Two actors are connected if there is a path linking them
  - Relevant to network meta-analysis
- Two actors are in separate components if there is no path linking them
- Number of components = number of groups of actors with no path between them
- A network is connected if it only contains a single component (i.e., a path between all pairs of actors exists)

From Erika Moen
Distance

- The length of a path is the number of edges on it
- Ties have path length 1
- **Shortest path between two actors = geodesic distance**
  - Infinite if no path exists between actors
  - Motivates analysis of largest connected component
- Size, degree, and density measures are well-defined irrespective of whether the network is connected but many network measures are not
  - Distance based measures are problematic if multiple components
Mutuality

• The Dyad census is a count of the number of (Mutual, Asymmetric, Unconnected/Null) dyads

• The count of the number of mutual dyads is a network statistic
  - Not centered; hard to interpret
  - Standard deviation interpretable because it is centered about the mean, unlike the 2-norm

• An index of mutuality or reciprocity should reflect difference from “analogous” network without mutuality
  → Seek the null distribution of number of mutual ties

• Equivalently, test if an actor’s in-degree is independent of their out-degree
Mutuality Index

• If each actor randomly chose $d$ friends from the N-1 other actors, the probability a dyad is mutual is given by
  \[ \Pr(\text{Mutual}) = \frac{d^2}{(N - 1)^2} \]

• Thus, in absence of mutuality:
  \[ E[M] = \frac{Nd^2}{2(N - 1)} \]
  where $M$ = observed number of mutual dyads and $N$ = number of actors

• Quantify within dyad dependence (mutuality) by defining $\rho$:
  \[ \Pr(A_{ij} = 1|A_{ji} = 1) = \Pr(A_{ij} = 1) + \rho \Pr(A_{ij} = 0) \]

• No mutuality (independence) if $\rho = 0$; certain tie if $\rho = 1$

• Methods-of-moments estimator:
  \[ \hat{\rho} = \frac{2(N-1)M-Nd^2}{Nd(N-1-d)} \]

Max$\{M\} = \frac{Nd}{2}$ → max index =1
Mutuality Index cont.

- If the degree of the actor varies, the probability actors $i$ and $j$ are a mutual dyad equals: $x_{i+}x_{+j}/(N-1)^2$
- Mutuality index generalizes to:
  \[
  \hat{\rho} = \frac{2(N-1)^2-L^2+L_2}{L(N-1)^2-L^2+L_2},
  \]
- where $L$ and $L_2$ are the sum and the sum of squares of the actor outdegrees, respectively
Clustering

- Extent to which the network neighbors of a given node are directly connected to one another.

- Clustering coefficient = probability that any two randomly chosen network neighbors of the individual at the center are directly connected to one another, = 0.4 above
Transitivity ("A friend of a friend is a friend")

• Sociologists → Triads are an important building block of society
• Triadic clustering is a special form of clustering
• Undirected network: the count of triangles is the basis of transitivity
• Directed network:
  • $4^3 = 64$ states of a triad
  • 16 triad groups that are non-isomorphic, embodying several sociological constructs
• The transitive triad (shown above for actor A) is perhaps of greatest interest
The 16 non-isomorphic triads (Frank, 1986)
Transitivity statistic

- Triad census is the empirical distribution of the 16 non-isomorphic triads
- Test for transitivity should condition on simpler explanations of triadic clustering
  - The in- and out-degree distribution and the dyad census
- To test for transitivity we need only consider the frequencies of the intransitive triad and the transitive triad
  - Condition on the dyad census (M = mutual, A = asymmetric, U = unconnected) and degree-distributions. Using randomly drawn networks with no triadic dependence generate null distribution of counts
  - Compare to corresponding observed statistic and reject if p-value is small (e.g., < 0.025)
Application: Comparison of Multiple Networks

• **Question**: Do network statistics correlate with differences in health care costs or intensity of care between institutional units or geographically-defined regions?

• **Example**: The network of physicians in hospitals or health referral regions across the US
  - Conjecture: physician influence and diffusion attitudes/practice-patterns operate through professional relationships

• “Standard” statistical analysis as have observations on multiple units that you want to compare
Bipartite Physician Hospital Networks

• Ideal: Directly observe ties between physicians
• In lieu of direct measurement use patients to infer professional relationships between providers
  • “Surrogate” relationship: Overlap of care for the same patients
• Assign physicians a “hospital” according to where the patients they care for are hospitalized the most
• Then form a network of physicians for each hospital
Surrogate Professional Relationships: Patient Sharing Imputed from Insurance Claims

One patient’s medical visits during a period of time:

Patient – physician complexities
- Multiple encounters
- Different medical reasons for encounters
- Varying importance of encounters

Physician – physician complexities
- Multiple overlapping patients
- Different patient medical conditions
- Different level of care requirements across patients

Contribution to professional relationship between A and B
Bipartite projection

\[ B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \]

\[ A = BB^T = \begin{pmatrix} 2 & 1 & 2 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 1 & 2 & 1 & 1 & 0 \\ 1 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 1 & 2 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \]
Bipartite or affiliation networks

• Network of physicians deduced from “shared patients” is an affiliation network

• Other affiliation networks
  • Network of scientists based on shared publications
  • Network of phenotypes based on shared genotypes (genes linked to multiple phenotypes)
  • Network of genotypes based on shared phenotypes or diseases
  • Network of diseases based on common comorbidities
Comparative Analysis

- Compute network statistics for each hospital or health referral region network
  - Barnett et al (2012) analyzed 526 hospitals
  - Landon et al (2012) sampled and analyzed 51 HRRs
- Model the outcome of interest (cost, intensity of care) in terms of network statistics and other predictors
Network Predictors of Interest

• **Relative PCP Centrality**: Ratio of average betweenness centrality of PCPs to that for non-PCPs:

\[
C_{\text{Rel.Bet}} = \frac{\overline{C}_{\text{Bet}}(\text{PCP})}{\overline{C}_{\text{Bet}}(\text{Non-PCP})}
\]

where \(\overline{C}_{\text{Bet}}\) denotes average betweenness centrality for a set of actors

• **Median adjusted degree (degree per 100 patients)**
  - Adjusted degree for a physician reflects extent to which a physician shares his patients with a broader array of other doctors
  - Take median across physicians for hospital-level measure
Example: Hospital Network Analysis (Barnett et al 2012)

• Predictor of interest: relative centrality of PCPs

- In panel A, PCP relative centrality is 0.35, so PCPs are about a third as central as other physicians in this network
  • Reflects the tight group of medical and surgical specialists at center of network
- In panels B and C, PCPs move more toward the center; PCP relative centrality increases
An elaborate Statistical Model

• Standard regression or hierarchical regression model:
  \[ Y_{ij} = \beta_1 x_{ij} + \beta_2 g_{ij} + \beta_3 x_i + \beta_4 g_i + \theta_i + \epsilon_{ij} \]
  where \( i \) denotes HRR and \( j \) denotes hospital within HRR
  • \( g_{ij} \) is a vector of network statistics for hospital \( j \) in HRR \( i \)
  • \( g_i \) is a vector of network statistics for HRR \( i \)
  • \( x_{ij} \) and \( x_i \) are other predictors at the hospital and HRR-level, respectively
  • Often include just predictors at the hospital or the HRR level, not both

• Random effect for clustering due to hospital:
  \[ \theta_i \sim \text{Normal}(0, \tau^2) \]

• Extend to patient-level (physician-level) if have patient-level (physician-level) outcome (e.g., cost) data
Results (Barnett et al 2012)

FIGURE 3. Adjusted estimates of hospital network structure versus cost and utilization outcomes. Each section represents the estimated effect of increasing a network measure (A), median adjusted degree, (B), PCP relative centrality by 1 standard deviation for the average-sized urban, nonprofit, nonteaching hospital in our sample on 3 different cost, hospital day, and physician visit outcomes. All estimates are adjusted for several hospital characteristics described in the Methods, including hospital size, urban/rural location, and case mix. Error bars show 95% confidence intervals for each estimate. *5 hospitals had missing data for the general medical/surgical and intensive care unit (ICU) hospital days outcomes, but did have data for the total hospital days outcome.
II. Network Structure and Social Selection

• Why identify factors affecting relationship status:
  • Recipe for manipulating the influences to which an individual is exposed
  • Determine factors that reinforce relationships
  • Finding optimal position (actors) at which to intervene on a population
Recall: Key Sociological Constructs

Homophily

$$\text{dist}(x_i, x_j) = \text{small}$$

I ---→ J

Tie more likely

$$\text{dist}(x_i, x_j) = \text{big}$$

I ---→ J

Tie less likely

Triadic Closure (Transitivity)

I K

I ---→ J

I, J tie leads to closure

K

I ---→ J

I, J tie does not lead to closure
Challenges with Sociocentric Data

• Triad = subgraph of three individuals
• Hypothesis: closed triads considered key
  • Must account for triadic effects to isolate other effects
• Triads are not separable units
  • Observed state of one triad restricts state of triads with which it shares two actors
• … Unit of observation is the whole network
  • Either model whole network or use latent variables to account for dependence between dyads

Consider first

Consider second
Random graph model

- Bernoulli or completely-random graph model:
  \[ A_{ij} \sim \text{Bernoulli}(p) \] for all \( i, j \)

implies that the model for the network is:

\[
\Pr(A = a | \mu) = \exp(\mu)^L \left(1 - \exp(\mu)\right)^{N(N-1)-L}
\]

where \( \mu = \log(p) \) and \( L = t_1(A) = \sum_{i \neq j} A_{ij} \) is the lone network statistic

- Erdos-Renyi (1959)

- Model can be generalized to allow heterogeneity in the tie-probabilities

- Generalization: relax assumption that all ties are independent to dyadic independence, allowing dependence within dyads
Dyadic Independence models

- Allow correlations between ties from the same dyad and across dyads due to common senders and receivers of ties
- Leads to the multinomial model:

\[
\begin{align*}
\Pr(A_{ij} = 0, A_{ji} = 0) & \propto 1 \\
\Pr(A_{ij} = 1, A_{ji} = 0) & \propto \exp(\mu_{ij} + \alpha_i + \gamma_j) \\
\Pr(A_{ij} = 0, A_{ji} = 1) & \propto \exp(\mu_{ji} + \alpha_j + \gamma_i) \\
\Pr(A_{ij} = 1, A_{ji} = 1) & \propto \exp(\mu_{ij} + \alpha_i + \gamma_j + \mu_{ji} + \alpha_j + \gamma_i + \rho_{ij})
\end{align*}
\]

where the proportionality constant is the sum of the above four probabilities

- Within dyad independence if \(\rho_{ij} = 0\)
- Actor sender and receiver effects are non-specific ("fixed") effects
Model of Network: Dyadic Independence

- Multiply the dyadic probabilities together to obtain model for the network
- A special case is the $P_1$ model (Holland Leinhardt 1981) obtained by setting $\mu_{ij} = \mu$ and $\rho_{ij} = \rho$:

$$Pr(A = a) \propto \exp\left(\mu t_1(a) + \sum_i \alpha_i t_{2i}(a) + \sum_j \gamma_j t_{3j}(a) + \rho t_4(a)\right)$$

where $t_{2i}(a)$ is the out-degree for actor $i$, $t_{3j}(a)$, is the in-degree for actor $j$, and $t_4(a)$ is the number of mutual dyads

- Four (sets of) network statistics characterize network
  - Richer than the dyad census in terms of isolating effect of mutuality
Stochastic Block Model

- Constrains dyadic independence model by requiring that density and reciprocity effects only vary between actors with different values of an attribute or attributes (e.g., gender, age, ethnicity) (Fineberg and Wasserman 1981, Tanner and Wong 1987, Karrer and Newman 2010)

- Extended to case when block is a latent variable (i.e., attribute variable is unobserved)
  - Alternative to modularity-based community detection (Hoff 2008)

- Further extension to multiple-membership stochastic block models (Airoldi et al 2008; J Mach Learn Res)
Beyond dyadic independence

- Dyadic independence allows model for the network to be generated from the model for the dyad through multiplication of dyad probability distributions.
- Dependence between dyads arises whenever the state of one dyad depends on the state of another dyad over and above actor-specific effects.
  - Triadic dependence: an edge is more (or less) likely to form if its actors have a common third actor.
  - Cannot multiply probability distributions of dyads to generate corresponding model for the network.
Exponential random graph models (ERGMs)

- Let \( t(a) \) denote a vector of network statistics
  - Functions of elements of adjacency matrix (A)
- An exponential random graph (or \( p^* \)) model has the form
  \[
  \Pr(A = a; \theta) = K(\theta)^{-1} \exp \left( \sum_k \theta_k t_k(a) \right)
  \]
- where \( K(\theta) = \sum_{a \in R(a)} \exp(\sum_k \theta_k t_k(a)) \) and \( R(a) \) is the set of all possible realizations for the network
- The dependent variable has \( 4^{N(N-1)/2} \) possible states
  - Observe one realization!
  - Challenge if does not factorize into models for dyads
Implied edge probability

- ERGMs imply the probability of a single edge is given by
  \[
  \text{logit}(\Pr(a_{ij} = 1|a_{ij}^c)) = \sum_k \theta_k \delta_k(a_{ij}^c)
  \]
- where \(a_{ij}^c\) denotes all ties other than \(a_{ij}\), \(\delta_k(a_{ij}^c) = t_k(a_{ij}^c \cup a_{ij} = 1) - t_k(a_{ij}^c \cup a_{ij} = 0)\) is the difference in the network statistics when \(a_{ij} = 1\) to when \(a_{ij} = 0\)
- \(\theta_k\) is the log-odds of \(a_{ij} = 1\) to \(a_{ij} = 0\) if the presence of \(a_{ij}\) would lead to a one-unit increase in \(t_k(a)\), conditioned on the rest of the network
Common network statistics, $t_k(a)$

- Density: $\sum_{i<j} a_{ij}$
- k-out stars: $\sum_i \left( \binom{a_{i+}}{k} \right)$
- k-in stars: $\sum_i \left( \binom{a_i+}{k} \right)$
- Reciprocity: $\sum_{i<j} a_{ij} a_{ji}$
- Transitive triad: $\sum_{i<j<k} a_{ij} a_{ik} a_{jk}$
- 3-cycle: $\sum_{i<j<k} a_{ij} a_{jk} a_{ki}$
- Nodematch or homophily covariate: $\sum_{i<j} a_{ij} I(x_i = x_j)$
- Preferential out-degree attachment: $\sum_{i<j} a_{ij} f(a_{i+} a_{j+})$
- For more extensive list see (e.g.,) Snijders et al (2007, SIENA manual)
Estimation of ERGMs

• Markov Chain Monte Carlo (MCMC) methods allow inferences to be based on the true likelihood function
  • MLE via MCMC integration (Geyer and Thompson 1992)
  • Fully Bayesian estimation
• Statnet R package (Handcock et al. 2003) can fit models on networks of up to thousands of actors
• Obtaining convergence can be difficult because the likelihood surface often has a highly irregular shape
  • Traps at local maxima, failure to converge, or converges to inappropriate “degenerate” solutions (Handcock 2003)
  • Degenerate solution: Simulated draws of the network under the model or estimated model yield the empty or complete network
Example: Use of StatNet

• pnet <- network(physnetwork, directed=TRUE, matrixtype="adjacency",
               vertex.attr=nodecov,
               vertex.attrnames = c("male", "whexpert",
                                    "pctwom", "numsess", "practice", "bcma", "bima",
                                    "bpp", "wnhlth", "numcat", "pctcat"))

• plot(pnet, mode = "fruchtermanreingold", displaylabels=T)
Statnet cont.

- `model1d <- ergm(pnet~edges + mutual + intransitive + nodeocov("whexpert") + nodeocov("pctwom") + nodeocov("numsess"))`
- `> summary(model1d)`

==========
Summary of model fit
==========

Formula:  `pnet ~ edges + mutual + intransitive + nodeocov("whexpert") + nodeocov("pctwom") + nodeocov("numsess")`

Iterations: 4 out of 20

Monte Carlo MLE Results:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>MCMC %</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>edges</td>
<td>-0.846548</td>
<td>0.300668</td>
<td>0</td>
<td>0.00496 **</td>
</tr>
<tr>
<td>mutual</td>
<td>1.734905</td>
<td>0.284658</td>
<td>0</td>
<td>&lt; 1e-04 ***</td>
</tr>
<tr>
<td>intransitive</td>
<td>-0.248473</td>
<td>0.030284</td>
<td>0</td>
<td>&lt; 1e-04 ***</td>
</tr>
<tr>
<td>nodeocov.whexpert</td>
<td>-0.116626</td>
<td>0.196923</td>
<td>0</td>
<td>0.55382</td>
</tr>
<tr>
<td>nodeocov.pctwom</td>
<td>0.002841</td>
<td>0.002970</td>
<td>0</td>
<td>0.33901</td>
</tr>
<tr>
<td>nodeocov.numsess</td>
<td>0.030477</td>
<td>0.026134</td>
<td>0</td>
<td>0.24381</td>
</tr>
</tbody>
</table>

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null Deviance: 1463.9  on 1056  degrees of freedom
Residual Deviance: 832.2  on 1050  degrees of freedom

AIC: 844.2  BIC: 874  (Smaller is better.)
Pros and Cons of ERGMs

- Model implied through specification of sufficient statistics of network; usually models specified first, then seek sufficient statistics
  - Pro: Allows tests for specific dependencies
  - Con: If network predictors include network configurations involving 3 or more actors, solutions susceptible to degeneracy (Handcock et al, 2003; Goldenberg et al, 2009)
  - Con: Computational barriers when N is ``large``
  - Con: Not generative

- Actor oriented (choice) models (Snijders, 2005, 2006) in SIENA are an alternative family of models that may be less vulnerable to problematic computational issues
Conditional Dyadic Independent Models

• Instead of using network statistics to account for dyadic dependence, condition on common or correlated latent variables
• Motivation: A random (or latent) effect extension of the $P_1$ model is the $P_2$ model (van Duijn et al, 2004):

$$\Pr(A_{ij} = a_{ij}, A_{ji} = a_{ji}) \propto \exp(\mu_{ij}a_{ij} + \mu_{ji}a_{ji} + \rho_{ij}a_{ij}a_{ji})$$

where

$$\mu_{ij} = \mu + \alpha_i + \gamma_j + \beta_1^T x_{1ij}$$

$$\alpha_i = a_i + \beta_2^T x_{2i}$$

$$\gamma_i = b_i + \beta_3^T x_{3i}$$

$$\rho_{ij} = \rho + \beta_4^T x_{4ij}$$

and $x_{1ij}, x_{2i}, x_{3i},$ and $x_{4ij}$ are covariates impacting density, propensity to extend ties, propensity to receive ties, and propensity for mutual ties

• $(a_i, b_i) \sim N(0, \Sigma)$ is a bivariate latent variable capturing unexplained variation and correlation in providers sending and receiving of ties
Latent Space Models

- $P_2$ model is a dyadic independence model that can be considered a foundation for models that use latent variables to account for between dyad dependence

- These more advanced models include:
  - Latent class models (Nowicki and Snijders, 2001; Airoldi et al 2008, 2010; Choi et al 2010)
  - Longitudinal data (O’Malley and Christakis 2011, Paul and O’Malley 2013, O’Malley and Paul 2013)
Latent space models cont.

• Adapt $P_2$ model by using additional latent variables to account for dependence between dyads
• Peter Hoff and colleagues have led the development of such models (Hoff 2002, 2005, 2008)
• Add $\xi(z_i, z_j)$ to $\mu_{ij}$ where

$$\xi(z_i, z_j) = \begin{cases} 
\lambda_{z_i, z_j} \text{ where } z_i, z_j \in \{1, \ldots, K\} \text{ and } \lambda_{z_i, z_j} = \lambda_{z_j, z_i} \\
-|z_i - z_j|^c \text{ where } c > 0 \text{ and } z_i, z_j \text{ have } K \text{ elements} \\
z_i^T U z_j \text{ where } z_i \sim N(0, \Sigma_z) \text{ and } U \text{ is a } K\text{-dimensional diagonal matrix}
\end{cases}$$

• These correspond to a categorical latent variable (latent class), a distance in a continuous space, and a latent eigenmodel, respectively.
Latent space models: interpretation

- Case 1 (latent class model) captures latent homophily by allowing individuals with the same value of the latent variable to have a greater probability of sharing an edge.
- Cases 2 and 3 portray similarity as a distance measure to capture structural equivalence. The dimension of $Z$ is arbitrary but generally is low.
  - Case 2 model accounts for transitivity by requiring that the latent distances between actors obey the triangle inequality.
  - Case 3 model accounts for both structural equivalence and latent homophily.
- Some latent space models can be estimated using the latentnet package in R.
- Hoff (2015) contributed amen package in R.
Example: Health Referral Region networks

- Applied model to multiple HRR networks (51 in total)
- Report results here for the binary-valued network of physicians in the North Dakota HRR where a tie represents that the sending physician shares 10% or more of their patients with the receiving physician
  - Network is directional!
Example of analysis with a latent space analysis

From Paul et al (2014)

2.3. Extended p2-model with latent effects (EP2 model)

We extend the traditional $p_2$ model in (1) to account for between-dyad dependence by incorporating the bilinear term from Section 2.2, using for illustration the inner-product form for $f$. We assume that the conditional probability for the status of the $ij^{th}$ dyad satisfies:

$$
\Pr((Y_{ij}, Y_{ji})|\theta_{ij}, \rho_{ij}) = g^{-1}(\theta_{ij}y_{ij} + \theta_{ji}y_{ji} + \rho_{ij}y_{ij}y_{ji}),
$$

$$
\theta_{ij} = \mu_{ij} + \alpha_i + \beta_j + f(v_i, v_j)
$$

$$
\rho_{ij} = \rho + \tau f(v_i, v_j)
$$

where \( \{v_i : (v_{i1}, \ldots, v_{ik})\}_{i=1,\ldots,n} \) are independent random variables from a $k$-component multivariate normal distribution. Therefore,
Results (log-odds-ratios)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>P2 Posterior mean</th>
<th>95% CI</th>
<th>EP2 (full) Posterior mean</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density (edges)</td>
<td>-6.80</td>
<td>(-7.22, -6.42)</td>
<td>-7.50</td>
<td>(-8.37, -6.99)</td>
</tr>
<tr>
<td>Dyadic covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same gender</td>
<td>0.06</td>
<td>(-0.11, 0.24)</td>
<td>0.10</td>
<td>(-0.01, 0.29)</td>
</tr>
<tr>
<td>Same specialty</td>
<td>0.05</td>
<td>(-0.07, 0.17)</td>
<td>-0.04</td>
<td>(-0.19, 0.10)</td>
</tr>
<tr>
<td>Same practice locality</td>
<td>4.03</td>
<td>(3.77, 4.32)</td>
<td>3.37</td>
<td>(2.99, 3.98)</td>
</tr>
<tr>
<td>Reciprocity (mutuality)</td>
<td>3.20</td>
<td>(2.90, 3.45)</td>
<td>2.43</td>
<td>(1.95, 2.93)</td>
</tr>
<tr>
<td>Latent space modification of reciprocity</td>
<td>n/a</td>
<td>(n/a, n/a)</td>
<td>0.24</td>
<td>(0.00, 0.52)</td>
</tr>
<tr>
<td>Random effects Covariance parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sender variance</td>
<td>2.24</td>
<td>(1.69, 2.86)</td>
<td>1.47</td>
<td>(1.00, 2.33)</td>
</tr>
<tr>
<td>Receiver variance</td>
<td>4.10</td>
<td>(3.11, 5.51)</td>
<td>6.91</td>
<td>(4.89, 9.83)</td>
</tr>
<tr>
<td>Correlation (sender, receiver)</td>
<td>-0.51</td>
<td>(-0.63, -0.34)</td>
<td>-0.71</td>
<td>(-0.84, -0.51)</td>
</tr>
<tr>
<td>Variance of latent effects</td>
<td>n/a</td>
<td>(n/a, n/a)</td>
<td>1.20</td>
<td>(1.00, 1.45)</td>
</tr>
</tbody>
</table>

The closer two physicians are in the latent space the greater the effect of reciprocity on the status of their dyad

Model was estimated using self-written R code (and C code)
Improvement of extended model over $P_2$ model and assessment of model fit

**Fig. 5.** (a–b) Posterior predictive distributions of overall triadic clustering in the physician patient-sharing generated network (example 2). The distributions are computed under our EP2 model (solid line) and the traditional $p_2$ model (dashed line). The solid dot represents the observed value of the statistic. (c) Posterior predictive distributions of physicians’ in-degree distribution. The gray lines are degree distributions computed under the full EP2 model. The black line denotes the observed degree distribution of the nodes in the patient-sharing generated physician network.
Longitudinal Analysis of Relationships

- Example from Framingham Heart Study: O’Malley and Christakis (2011); Paul and O’Malley (2013)
- $A_{ijt}$ denotes tie (1 if friend, 0 otherwise) from individual $i$ to $j$ at time $t$
- Dyad status $(A_{ijt}, A_{jiti}) = \text{unit of analysis}$
  - Four-component multinomial distribution
  - Simultaneous modeling of tie-dissolution and tie-formation
- Assume Markov model
  - No dependence beyond prior time (e.g. no multi-period lags)
  - Contingency table(s) of dyadic transitions form “sufficient statistics”
Triadic Terms in $X_{ijt}$: Test key sociological constructs in lagged time

- Count over all individuals $k \neq i, j$
- Dashed line present (closed structure)
- Dashed line absent (open structure)
- Tie-formation and tie-dissolution effects may differ
- Represent as lagged predictors in model to avoid imposing constraints on the network and encountering degeneracy etc.
## Base Terms: Homogeneous Markov model

<table>
<thead>
<tr>
<th>((A_{ij}, A_{ji}))</th>
<th>Time = (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((0, 0))</td>
</tr>
<tr>
<td>((0, 0))</td>
<td>(\pi_1)</td>
</tr>
<tr>
<td>((1, 0))</td>
<td>(\pi_4)</td>
</tr>
<tr>
<td>((0, 1))</td>
<td>(\pi_4)</td>
</tr>
<tr>
<td>((1, 1))</td>
<td>(\pi_8)</td>
</tr>
</tbody>
</table>

- Transition probabilities invariant to sender and receiver:
  - Probabilities along each row sum to 1 (row stochastic)
  - \(10 - 4 + 1 = 7\) “fundamental” parameters
Full Longitudinal Model of Relationship Status

• Paul and O'Malley (2013)
• Ensure sensible solutions by requiring independence of relationships between dyads at current time
  • $A_{ijt}$ must be independent of $A_{hkt}$ if $h \neq i$ or $k \neq j$
  • But $A_{ijt}$ may depend on $A_{jit}$ (“reciprocity”)
• Allow dependence of relationships over time
  • $A_{ijt}$ can depend on $A_{hk(t-1)}$ for any $(h, k)$
Predictors of Change in Relationship Status

• $X_{ijt}$: other covariates relevant to $A_{ijt}$. Examples:
  • Difference in BMI: $|BMI_{it} - BMI_{jt}|$
  • Difference in smoking status: $(\text{smoke}_{it} \neq \text{smoke}_{jt})$
  • Lagged terms involving other individuals in network

• $(A_{ij(t-1)}, A_{ji(t-1)})$: past status of dyad
  • Density effects: prevalence of ties
  • Reciprocity effects: modification for mutual ties
III. Social Influence with longitudinal data ("Peer Effects", "Induction", "Contagion")

The Spread of Obesity in a Large Social Network Over 32 Years

Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.
Peer Effects

• One person(s) influence on another
• Also referred to as “social influence or contagion”

History: Modeling diffusion of innovations

• Coleman et al (1957): Social network/social structure related to diffusion of information among physicians
• Coleman et al (1966): Diffusion of tetracycline, a new medical drug, among doctors (more interpersonal connections = faster adoption)
• Cosma Shalizi implied that can’t get causal effects outside of randomized studies (here, last year)
Why do peer effects matter in Medicine?

1. Justify behavioral interventions
   - Peer-support and group interventions
   - Targeted ("seeded") interventions

2. Evaluate full effect of an intervention ("collateral effects")
   - Intervene on the untreated
   - Account for spillover effects (Sobel 2006)
Dyadic Influence: Longitudinal case

• Focus on a single dyad for now (individuals $i, j$)
• $Y_{it}$ = Outcome for individual $i$ at time $t$
• $X_{it}$ = Other characteristics of individual $i$ at $t$
• Example Model:
  $$Y_{it} = \beta_1 Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + \beta_3 X_{i(t-1)} + \theta_i + \epsilon_{it}$$
  - $\beta_1 > 0$: positive effect of $j$’s outcome on subsequent outcome for $i$
  - $\theta_i, \epsilon_{it}$ are latent variables, error terms
  - May use $X_{j(t-1)}$ as additional predictor

• Roles of $i$ and $j$ reversed in another observation!
  - Reflection problem (Manski, 1993)

• Estimate lagged effects to distinguish effect of $i$ on $j$ from $j$ on $i$

Other concerns addressed later!
Beyond Dyads: Network Influence

Digraph

Influence Matrix, $W$

Row stochastic matrix
Network Influence Models

• Assumes influence from multiple peers operates through the mean of their focal trait

• Let $\tilde{Y}_{-it} = [W_t Y_t]_i$ denote alters average trait at $t$, where $W_t$ is the influence matrix at $t$ and $Y_t$ is the vector of the trait of interest

• Simple network influence model:

$$Y_{it} = \beta_1 \tilde{Y}_{-i(t-1)} + \beta_2 Y_{i(t-1)} + \beta_3 X_{i(t-1)} + \theta_i + \epsilon_{it}$$

• Lots of other specifications!
Analogy of Network Influence to Spatial Statistics Model?

• Notion of Adjacency matrix; distance
  • Spatial analysis of areal data
  • Notion of distance with point referenced data

• Subtle issues in network data due to interference between dyads in sociocentric data, unlike the mutually exclusive dyads case!
  • No longer have “isolated experiments”
  • Stable Unit Treatment Value Assumption (SUTVA) violated!
Estimation

- Ordinary least squares (OLS)
  - Parametric assumptions key with cross-sectional data
- Instrumental variables/Two-stage least squares: (Anselin, 1988; Land and Deane, 1992)
  - Isolate variation in $\tilde{Y}_{i(t-1)}$ unrelated to unmeasured variables shared between actors
Causality concerns

• Homophily (U): “Birds of a feather flock together”
  • Seek individuals with similar habits (e.g., over-eating, smoking); then become friends
  • Tie-dissolution due to diverging traits

• Unmeasured common cause (C)
  • Propaganda about healthy living
  • New fad diet
  • New drug
  • Unknown peers in common
Causal Model for Peer Effects (Dyadic case)

1. Assume lag-1 peer effects (no simultaneity) and times (0, 1, 2)
2. Multiple-periods (t denotes time-periods since friendship formed; below)
3. (Latent) homophily bias at all times (U_k affects A and Y_k in all periods)
4. Unobserved confounding at all times (C affects Y_i and Y_j in all periods)
5. Observed individual level confounders, X_k, e.g., race, age, education

![Diagram of causal model]
Causal Model Is Not Identified

• The causal effect $Y_{j(t-1)} \rightarrow Y_{it}$ is not identified
• The association between $Y_{j(t-1)}$ and $Y_{it}$ includes:
  1. The causal effect: $Y_{j(t-1)} \rightarrow Y_{it}$
  2. Homophily bias: $Y_{j(t-1)} \leftarrow U_j \rightarrow [A_{ij}] \leftarrow U_i \rightarrow Y_{it}$
  3. Confounding bias: $Y_{j(t-1)} \leftarrow C \rightarrow Y_{it}$
  4. Other paths depending on what observables $Z \in \{X_i, X_j, Y_{i(t-1)}, Y_{i(t-2)}, Y_{j(t-2)} \ldots\}$ are conditioned on
How to identify causal effect?

• Longitudinal data helps:
  • Especially if U, C have cross-sectional or short-term effects

• In general, can use a joint modeling approach:
  • Impose parametric distribution assumptions; for example, joint multivariate normality
  • Steglich et al. (2010)

• For causal inference, do not want to rely on joint parametric distribution assumptions
  • Assumptions are not conclusively testable
Instrumental Variables (IVs)

• An IV is a variable that, under appropriate causal assumptions, permits
  1. **Nonparametric test** of the Null of no causal effect
  2. Estimation of the total causal effect via Wald estimator or two-stage least squares (2SLS) assuming linearity & homogeneity but not a parametric distribution
  3. Helpful in other situations (e.g., nonlinear models)
• Estimation performed using two-stage least squares or related method
Genetic Alleles: IVs for peer effects?

- Genetic alleles (denoted G) fixed at conception
  - Measurable any time
  - Endure over lifetime

- Genes linked to obesity and BMI:
  - Fat mass and obesity gene (FTO)
  - Melanocortin-4 receptor gene (MC4R)

- Homozygous (risk and non-risk) and heterozygous states are not directly observable

- Seems like a good idea but turns out (O’Malley et al 2014) that IV needs to be time-varying
New Idea: Use gene-expression as the IV

- Gene expression may vary over one's lifetime?
- So use age is a likely modifier of gene expression
  - Hypothesis: Relative influence of genetic state varies with age
  - Allow change in effects of genetic states with age
  - Let $G_{X_{jt}} = G_j * Age_{jt}$
- Helpful as creates instrument with temporary effects
Time-Varying Gene Expression as IV

Theorem: $G X_{j(t-1)}$ is a valid IV for peer effect $Y_{j(t-1)} \rightarrow Y_{i(t)}$ conditional on $G_j, X_j, A_{ij}$. 

Diagram: 

- $G_j$ 
- $X_j$ 
- $U_j$ 
- $Y_{j(t-2)}$ 
- $Y_{j(t-1)}$ 
- $G X_{j(t-2)}$ 
- $G X_{j(t-1)}$ 
- $Y_{j(t)}$ 
- $A_{ij}$ 
- $C$ 
- $U_i$ 
- $Y_{i(t-2)}$ 
- $Y_{i(t-1)}$ 
- $G X_{i(t-2)}$ 
- $G_i(t-1)$ 
- $G X_{i(t)}$ 
- $Y_{i(t)}$ 
- $Y_{it}$ 
- $G_i(t)$ 
- $X_i$ 
- $G_i$
Estimation: Two-stage least squares (2SLS) procedure with $G X_{j(t-1)}$ as IV

Let $Z = \text{conditioning set}; \text{ e.g., } Z \subseteq \{X, G_i, Y_{t-2}\}$

Stage I: $Y_{j(t-1)} = G X_{j(t-1)}^T \theta_1 + Z_t^T \theta_2 + \delta_{it}$

Compute fitted values: $\hat{Y}_{j(t-1)}$

Stage II: $Y_{it} = \hat{Y}_{j(t-1)} \beta_1 + Z_t^T \beta_2 + \hat{\epsilon}_{it}$

where $\hat{\epsilon}_{it} = \epsilon_{it} + (Y_{j(t-1)} - \hat{Y}_{j(t-1)}) \beta_1$
Example: Framingham heart study (FHS) network analysis of peer effects of BMI

  - Slightly modified dataset to what Christakis and Fowler (2007), and subsequent papers, analyzed
  - Non-directed relationships
- Conduct separate analyses for two types of relationships:
  - A “close friend” (named friend at exam)
  - Spouse
- Genes related to obesity (based on work of others)
  - FTO (A = risk allele, T = protective allele)
  - MC4R
First Stage: GX → BMI (non-parallel = good)
Table 2

Dyadic Peer Effect Analysis of Lag Alter BMI using Time-varying gene-age expression as an instrument

<table>
<thead>
<tr>
<th>Discretionary $Z_{(t)}$ Terms</th>
<th>IV Regression (2SLS) $^a$</th>
<th>Regression (OLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$GX_{2(t-2)}$, $Y_{1(t-1)}$</td>
<td>$F_5^b$</td>
<td>Estimate</td>
</tr>
<tr>
<td>Exclude, Exclude</td>
<td>2.150</td>
<td>0.888</td>
</tr>
<tr>
<td>Exclude, Covariate</td>
<td>1.731</td>
<td>0.874</td>
</tr>
<tr>
<td>Covariate, Exclude</td>
<td>1.181</td>
<td>0.133</td>
</tr>
<tr>
<td>Covariate, Covariate</td>
<td>1.144</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

Spouse

| Exclude, Exclude | 4.064 | 0.099 | -0.324 | 0.522 | 0.066 | 0.039 | 0.094 |
| Exclude, Covariate | 4.351 | 0.101 | -0.287 | 0.488 | 0.032 | 0.008 | 0.055 |
| Covariate, Exclude | 0.268 | -0.102 | -1.855 | 1.652 | 0.050 | 0.017 | 0.082 |
| Covariate, Covariate | 0.181 | 0.906 | -1.832 | 3.643 | 0.023 | -0.006 | 0.051 |

$^a$ $Z_{(t)} = (GX_{1(t)}, X_{1(t)}, X_{2(t)})$ are exogenous covariates and $GX_{2(t-1)}$ is an IV in all IV analyses. The elements of $X_{k(t)}$, $k = 1, 2$, are: gender, age, gender-age interaction, birth era, birth year, smoking status, number of siblings, and (for $k = 1$ only) the geographic distance between residential locations of ego and alter at tie-formation. All models include dyad fixed effects. $GX_{2(t-2)}$ and $Y_{1(t-1)}$ are added to $Z_{(t)}$ as indicated in the two left-most columns.

$^b$ The F-statistic is for the overall effect of the IV, $GX_{2(t-1)}$, in the first-stage equation. The critical value of the Cragg-Donald F-statistic, which quantifies the power of an IV, at the 20% level ranges from 6.71 to 6.77 across the models.
Conclusion

• Social network analysis is a broad field
  • Many disciplines involved now, but not historically

• Comparison of networks of different organizations, analysis of social selection via models of network relationships, and social influence ("peer effects") are structurally different problems!

• Many methods for the analysis of social network data differ from standard methods in biostatistics:
  • One person’s "treatment" may affect another’s "outcome"
  • Interference of observations!

• New and growing area in statistics; lots still to be done!!
References: Descriptive network measures

- Freeman, L. (1979), Centrality in Social Networks, Conceptual Clarification, Social Networks, 1, 215-239
- Faust, K. (1997), Centrality in Affiliation Networks, Social Networks, 19, 157-191
- O'Malley, A. J. and Marsden, P. V. (2008), The Analysis of Social Networks, Health Services & Outcomes Research Methodology, 8, 222-269
References: Bipartite Networks and Comparison of Multiple Networks

References: Basic Analyses and Models

References: cross-sectional models of relationships

References: cross-sectional models of relationships cont.

References: dynamic network models of relationships

- O'Malley AJ, Christakis NA. (2011). Longitudinal Analysis of Large Social Networks: estimating the Effect of Health Traits on changes in Friendship Ties. Statistics in Medicine, 30, 9, 950-964
References: dynamic network models of relationships cont.

References: social influence analyses

- O'Malley AJ, Marsden PV. (2008). The Analysis of Social Networks Health Services and Outcomes Research Methodology, 8, 222--269
References: social influence analyses cont.

- Coleman, J., Katz, E., and et al. (1966), Medical Innovation: A Diffusion Study, Bobbs-Merrill
References: network-influence related models


