Ego Network Analysis

Jeffrey A. Smith
University of Nebraska-Lincoln
Department of Sociology
Social Networks & Health Workshop
May 17, 2016

Structure of the Class

- Preliminaries
  - What are ego network data?
- Analysis
  - What can we do with ego network data?
- Ego networks and health
  - What health-related questions can we answer?
- Example of ego network analysis in R

Goals

- By the end you should have gained:
  - Familiarity with ego network data
  - Background on measurement and applications to health outcomes
  - Tools for analyzing ego network data in R

Definitions

- Ego network (personal network)
- Ego: focal node/respondent
- Alter: actors ego has ties with
- Ties between alters

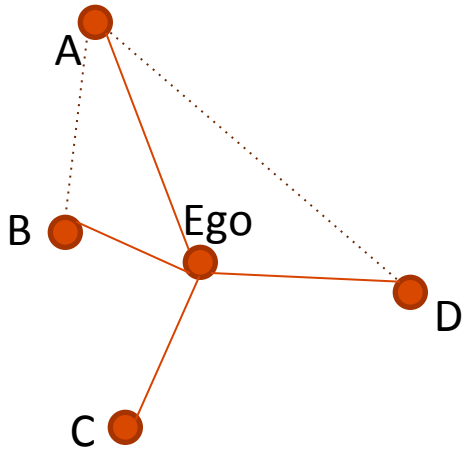
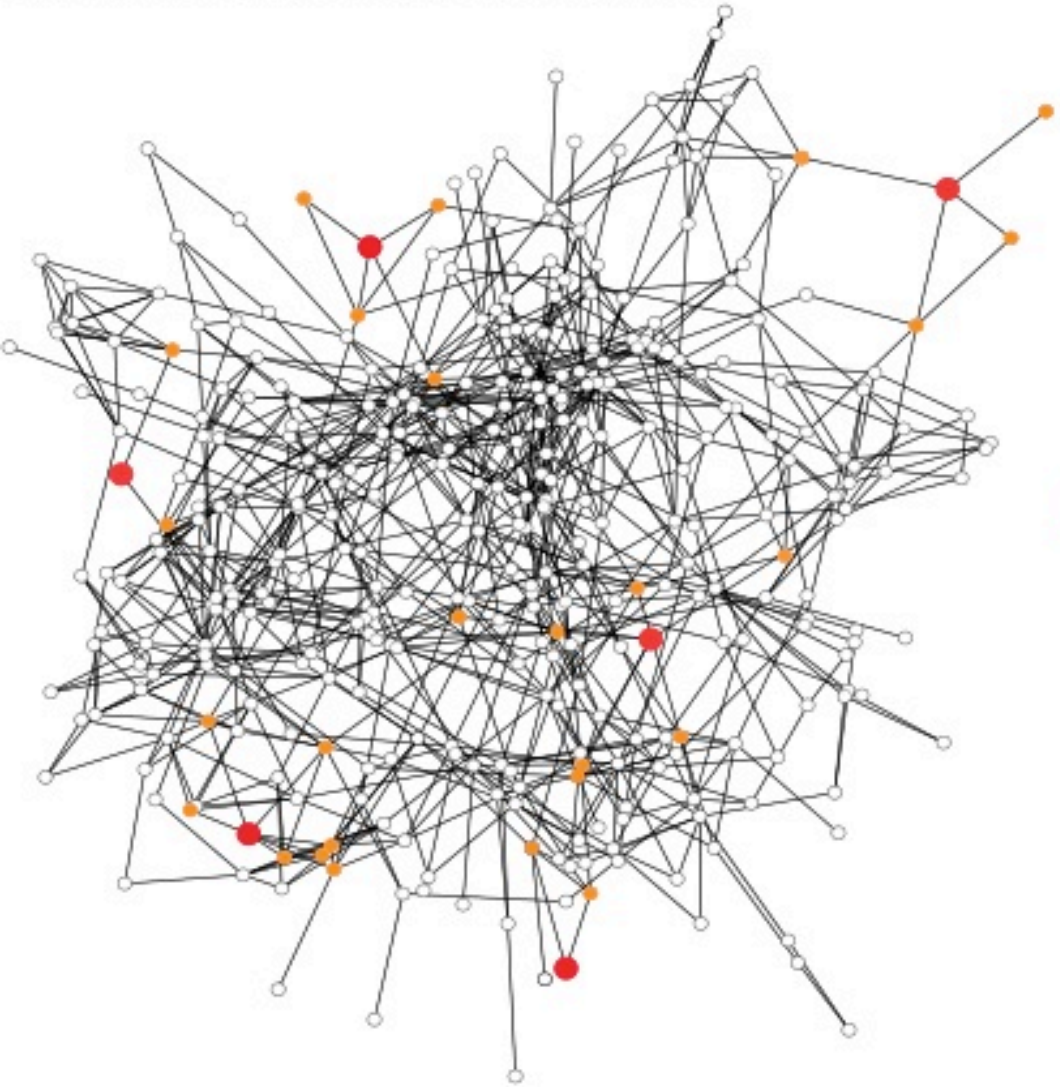
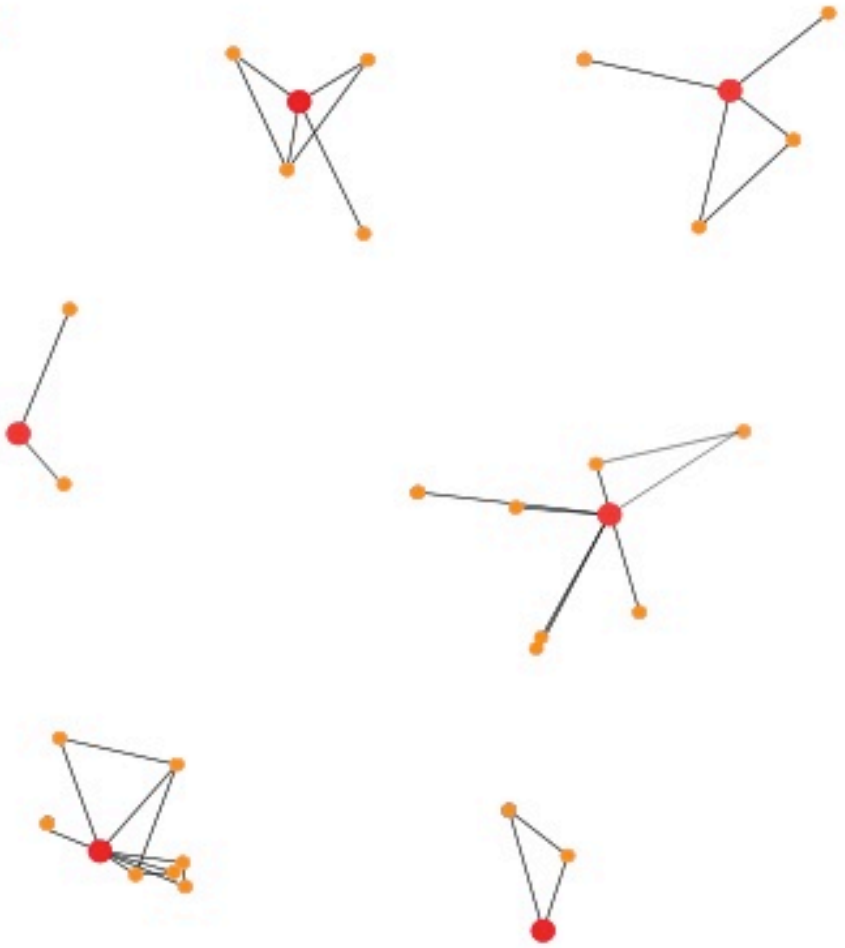


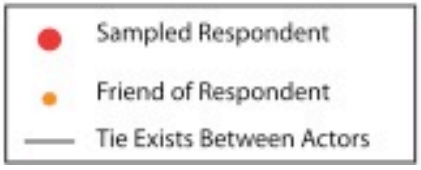
Figure 1. Example Network and Ego Network Sample



Panel 1: Full Network with Highlighted Ego Networks



Panel 2: Isolated Ego Networks



Ego Networks and Traditional Survey Data

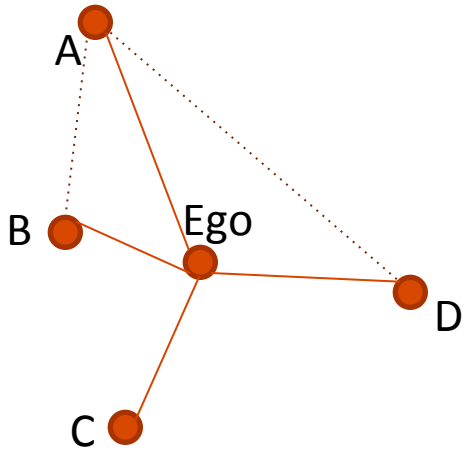
- Ego networks combine aspects of traditional survey data with network data
- Traditional survey
  - Independently sampled respondents
  - Get characteristics of those respondents through survey questions

Respondent	Age	Educ	Smokes	Drinks
Ego	40	16	No	Yes

Ego Networks and Traditional Survey Data

- Ego networks combine aspects of traditional survey data with network data

Respondent	Age	Educ	Smokes	Drinks
Ego	40	16	No	Yes



	Ego	A	B	C	D
Ego		1	1	1	1
A	1		1	0	1
B	1	1		0	0
C	1	0	0		0
D	1	1	0	0	

Why use ego network data?

- From ego's perspective, personal network is important for:
  - Social support
  - Access to resources
  - Influence/normative pressure

Why use ego network data?

- From a more global perspective, ego network data are useful for:
  - Studying mixing patterns between groups
  - Potential for diffusion
    - Disease propagation
    - Adoption of innovation: new product or health practice

Why use ego network data?

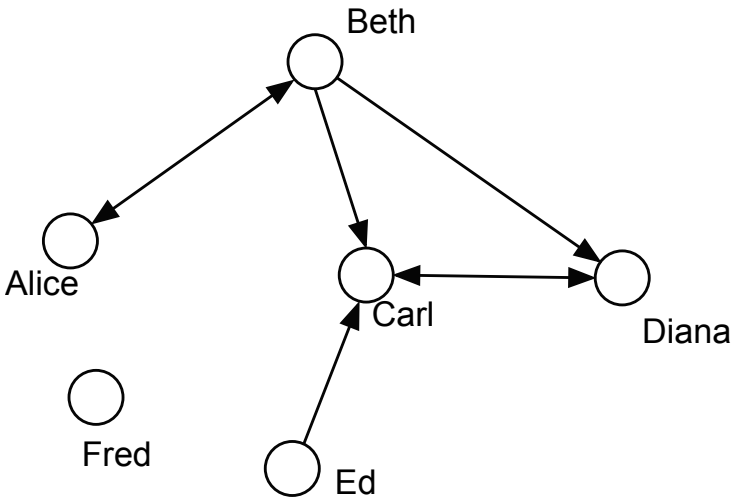
- Lots can be had from ego network data!
  - Composition of individual's local social world
    - Demographic characteristics of alters
    - Shared health behaviors
  - Structural features
    - Size
    - Density
  - Nature of the ties
    - Frequency, duration, closeness
    - Specific exchanges

Cost/Benefits to Ego Network Data

- Benefits:
  - Lots of information for cheap
  - Easy to collect as part of a traditional survey
  - Often only interested in personal networks
  - Can use local information to describe global network properties
- Costs:
  - Rely (typically) on self reports of ties
  - Egos are treated as independent
  - Will generally miss larger structure of network

Collecting Ego Network Data

- 1) Could collect full network data
  - extract ego networks
  - analyze as independent



Alice	Alice	Beth
Alice		1
Beth	1	

Beth	Beth	Alice	Carl	Diana
Beth		1	1	1
Alice	1		0	0
Carl	0	0		1
Diana	0	0	1	

- 2) More typical is to collect independently sampled ego network data
  - Random sample of individuals
    - Ask standard survey questions (age, education)
  - Ask each person to report on
    - alters
    - alter characteristics
    - ties between alters

Step 1

- Each respondent is asked to list set of contacts
  - No need for actual names or ids
  - Often truncated but best not to
  - Generally a good idea to ask multiple questions

Step 1

- Ask open ended questions to elicit social contacts
  - General: Looking back over the last six months who are the people with whom you discussed an important personal matter?
  - Behavioral: Who have you slept with in the last 6 months? Who have you shot up with in the last 6 months?
  - Support: If you were sick, who would be willing to accompany you to the hospital? Who do you go to for advice on health-related matters?

Step 2

- Ask respondents about each named alter
  - Characteristics
    - age, education, gender…
  - Nature of tie with alter
    - Frequency of contact
    - Kin/non-kin, type of relationship
    - Closeness to alter
    - Duration of relationship

	Age	Educ	Age Alter1	Age Alter2	Educ Alter1	Educ Alter2	Freq Alter1	Freq Alter 2
Ego	40	16	38	35	16	12	Weekly	Monthly

Step 3

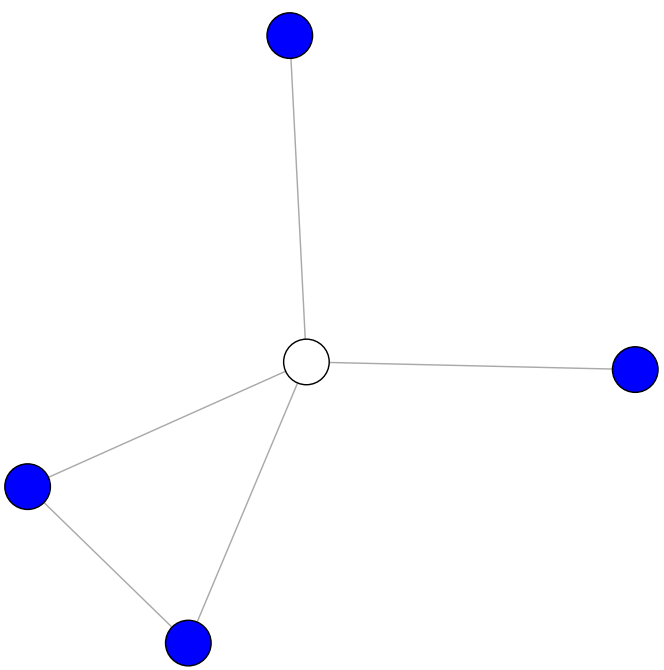
- Ask respondent about ties between alters (if possible)
  - Think about the relationship between <alter1> and <alter2>. Would you say that they are strangers, just friends, or especially close?

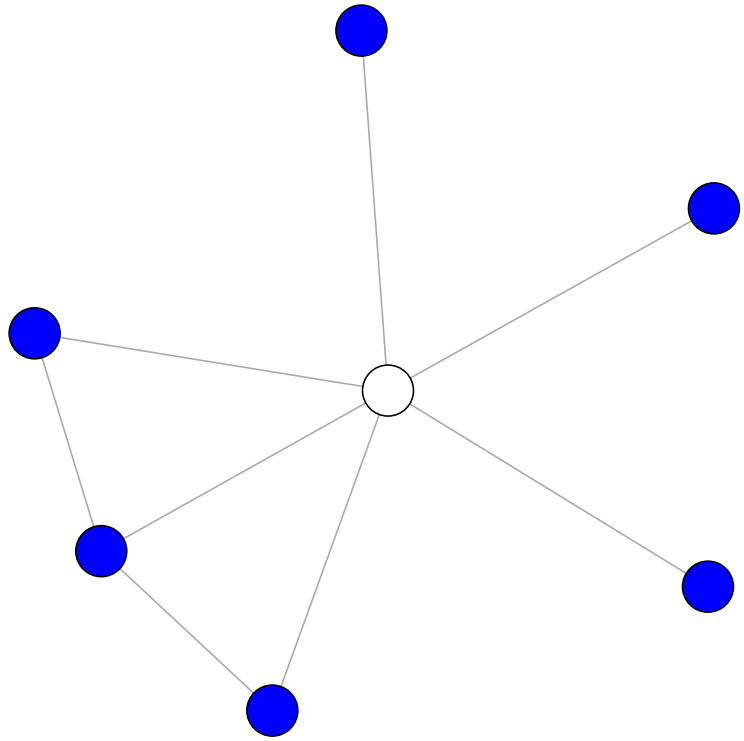
	Age	Educ	Age Alter1	Age Alter2	Educ Alter1	Educ Alter2	Freq Alter1	Freq Alter 2	Tie: A1-A2?
Ego	40	16	38	35	16	12	Weekly	Monthly	Close

Analyzing Ego Network Data

- Different kinds of questions/analyses than with full network data
- Often measure property of ego network to use as predictor in typical statistical model

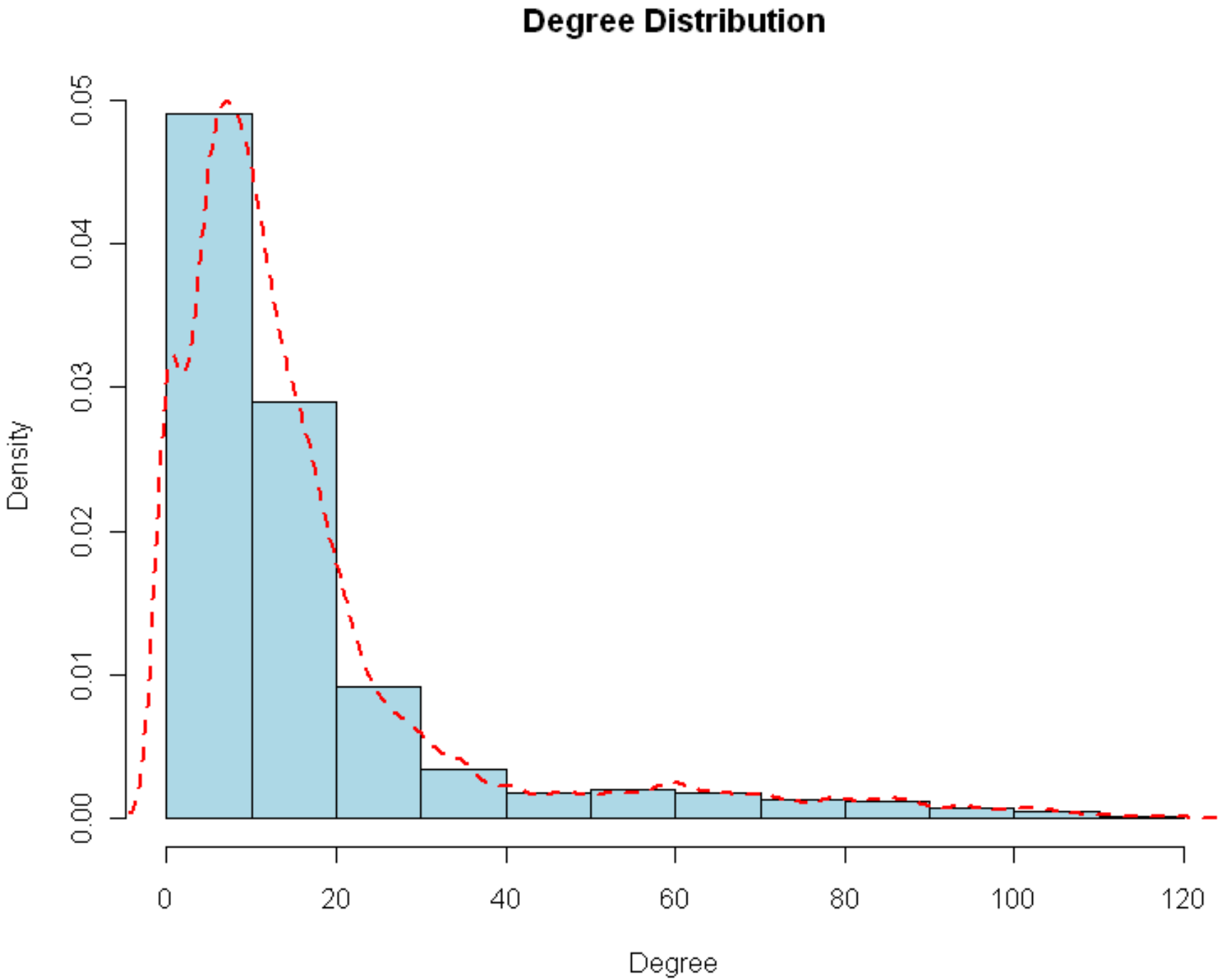
Network Size (Degree)





Network size: Local and Global Measures

- Local: personal network size
  - Number of alters (social support) predicting health outcomes
  - Number of drug partners predicting future risky behavior
- Global: degree distribution by aggregating over all cases
  - Distribution of ties per person


Degree Distribution
Density
0.00
0.01
0.02
0.03
0.04
0.05
0
20
40
60
80
100
120
Degree

Composition

- Distribution of types of people and resources in ego network
  - Demographic characteristics
  - Types of relationships (kin/non-kin)
  - Resources available to ego
  - Risks to ego

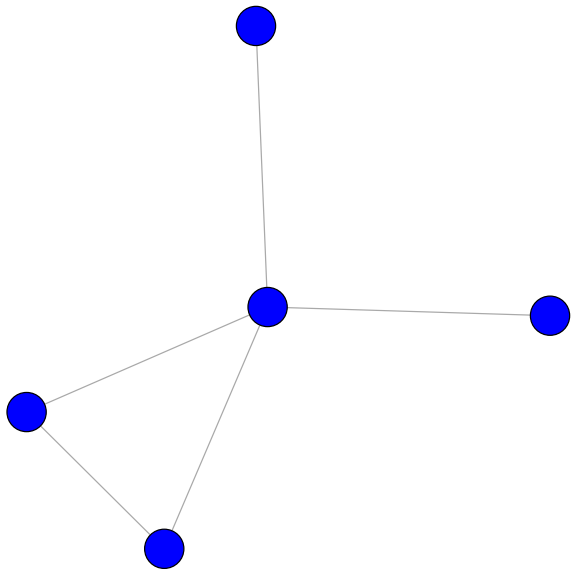
Demographic characteristics

- Homophily tends to prevail
  - Ego networks are more homogenous than population at-large
  - Largely due to structural constraints
- Local: How diverse/homogenous is an individual's social world?
- Global: How much contact is there between demographic groups?

- Local: How diverse/homogenous is an individual's social world?
  - What proportion of ego's friends are white? Female? What proportion of ego's friends are of different gender than themselves?
- Global: How much contact is there between demographic groups?
  - What proportion of ties are between/within racial groups?

Homophily

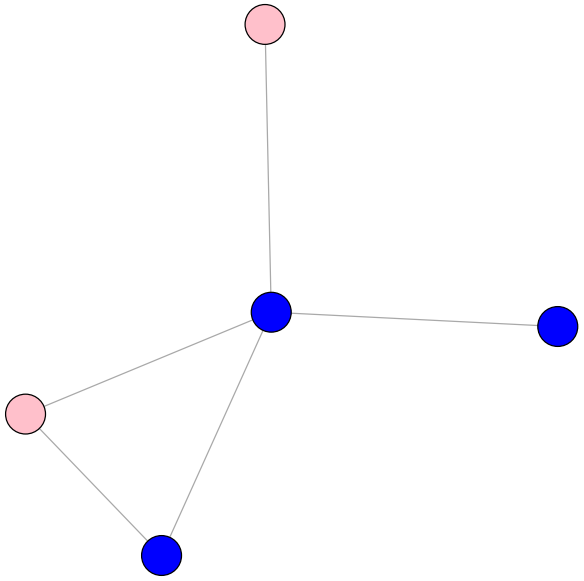
- How similar is ego to their alters?
- Two simple measures (just focus on ego-alter pairs):
  - proportion homophilous
  - E-I index: $\frac{E-I}{E+I}$
  - Where E= number of ties to different groups
  - And I=number of ties to same group
  - Ranges from -1 (homophily) to 1



E=0
I=4
E-I: (0-4)/4=-1

$$\frac{E-I}{E+I}$$

Proportion homophilous:
Proportion same/degree
4/4=1



E=2
I=2
E-I: (2-2)/4=0

$$\frac{E-I}{E+I}$$

Proportion homophilous:
Proportion same/degree
2/4=.5

Heterogeneity

- Could also measure heterogeneity amongst alters (ignore ego here)
  - e.g., diversity of social support or job contacts may be advantageous
- IQV as one possible measure:

$$\mathrm{IQV} := \frac{K}{K-1}\left(1 - \sum_{i=1}^{K} p_i^2\right)$$

- Where k is the number of categories and p is the proportion in category i

Resources and Risks

- We can use the same measures on other characteristics of interest, such as:
  - Distribution of drugs users in network (risks)
  - Distribution of income/education, kin, etc. in network (resources)
  - Distribution of health statuses (chronic illness) in network (i.e., stressors)

Global Measures

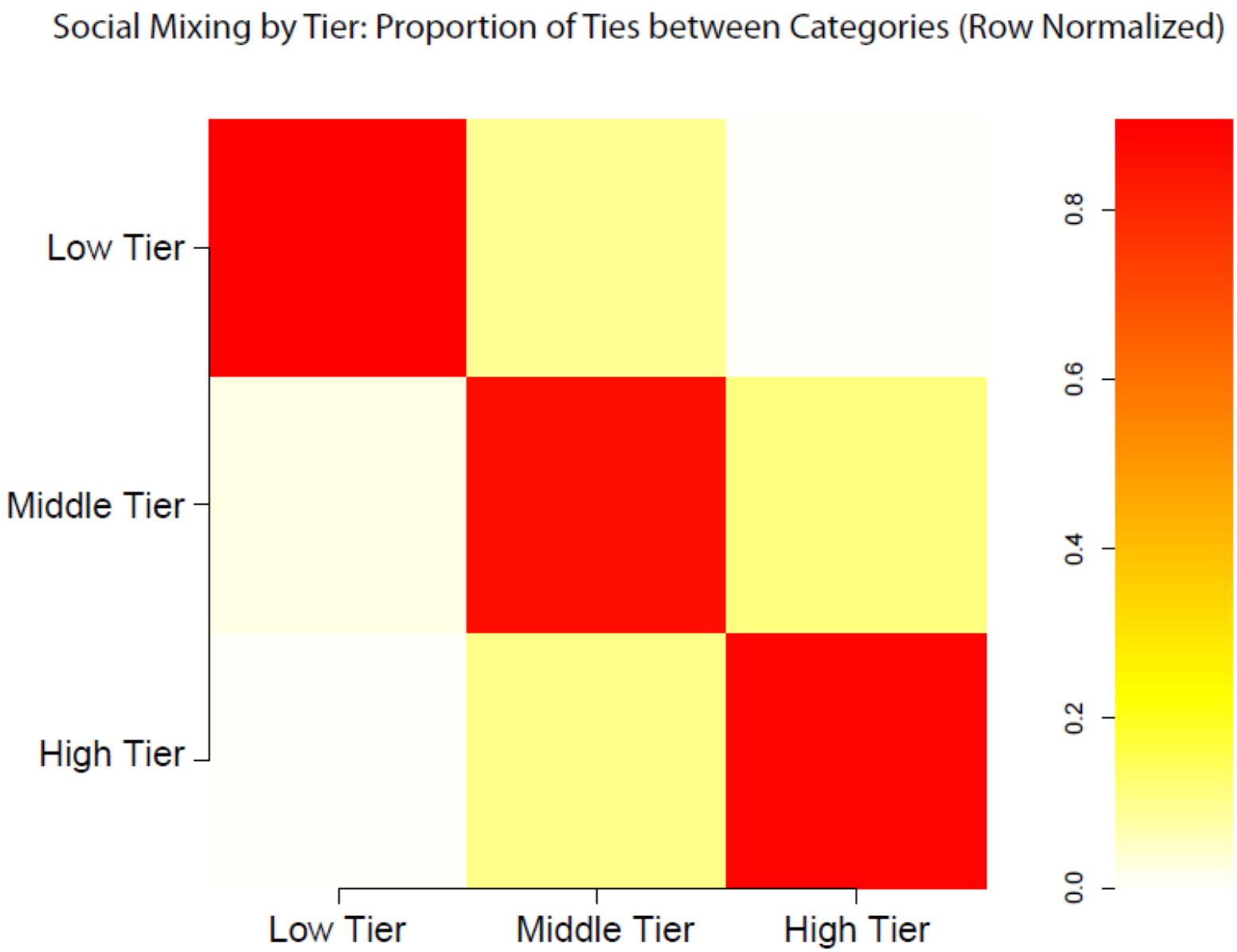


Table 1. Summary Statistics

	1985		2004	
	Mean	SE	Mean	SE
Race				
Racial Mismatch between Respondent and Confidant***	.047	.006	.098	.010
Racial Mismatch Expected by Chance***	.276	.015	.387	.020
Religion				
Religious Mismatch between Respondent and Confidant**	.241	.010	.290	.014
Religious Mismatch Expected by Chance***	.535	.011	.658	.013
Sex				
Sex Mismatch between Respondent and Confidant*	.403	.008	.433	.011
Sex Mismatch Expected by Chance	.498	.003	.492	.005
Age				
Absolute Age Difference between Respondent and Confidant	11.792	.234	11.150	.283
Absolute Age Difference Expected by Chance**	19.839	.287	18.584	.354
Education				
Absolute Education Difference between Respondent and Confidant	2.115	.049	2.047	.058
Absolute Education Difference Expected by Chance	3.317	.084	3.120	.079

Note: The table includes significance tests comparing the level of homophily in 1985 to the level in 2004. The level of significance is placed next to the name of the statistic. Standard errors are calculated from bootstrap samples for the observed level of homophily, and using complex survey design for the level expected by chance.

*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests)

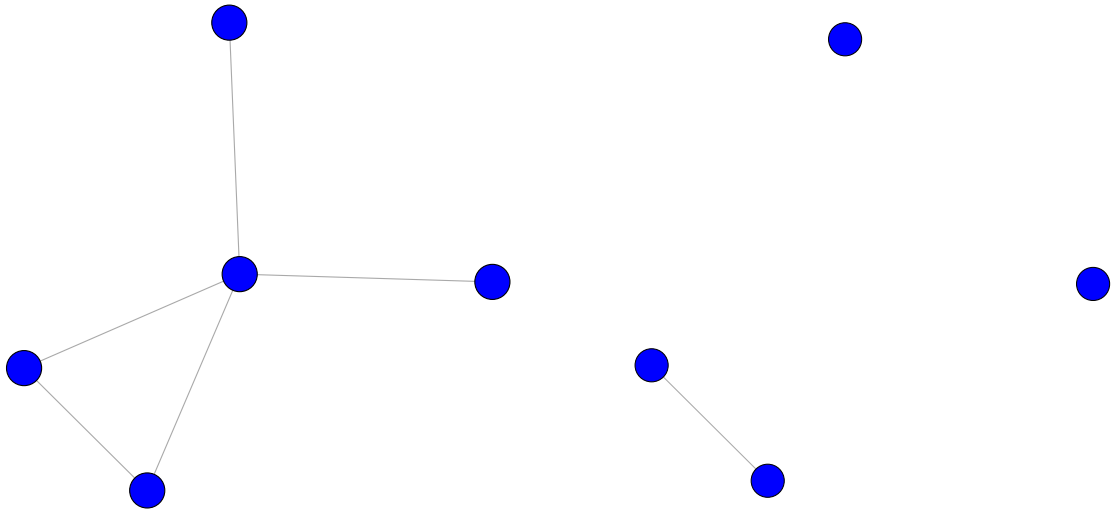
Smith et al. 2014

Structural Measures

- Measuring structural features of ego network
  - Use alter-alter ties+network size
- Different network environments lead to different outcomes
  - Norms easier to establish/maintain if all friends know each other (stronger social closure)

Density

- Like normal density but ignore ego-alter ties
- Proportion of ties between alters compared to number possible



Total Ties=1
Possible=4*3/2=6
Density=1/6

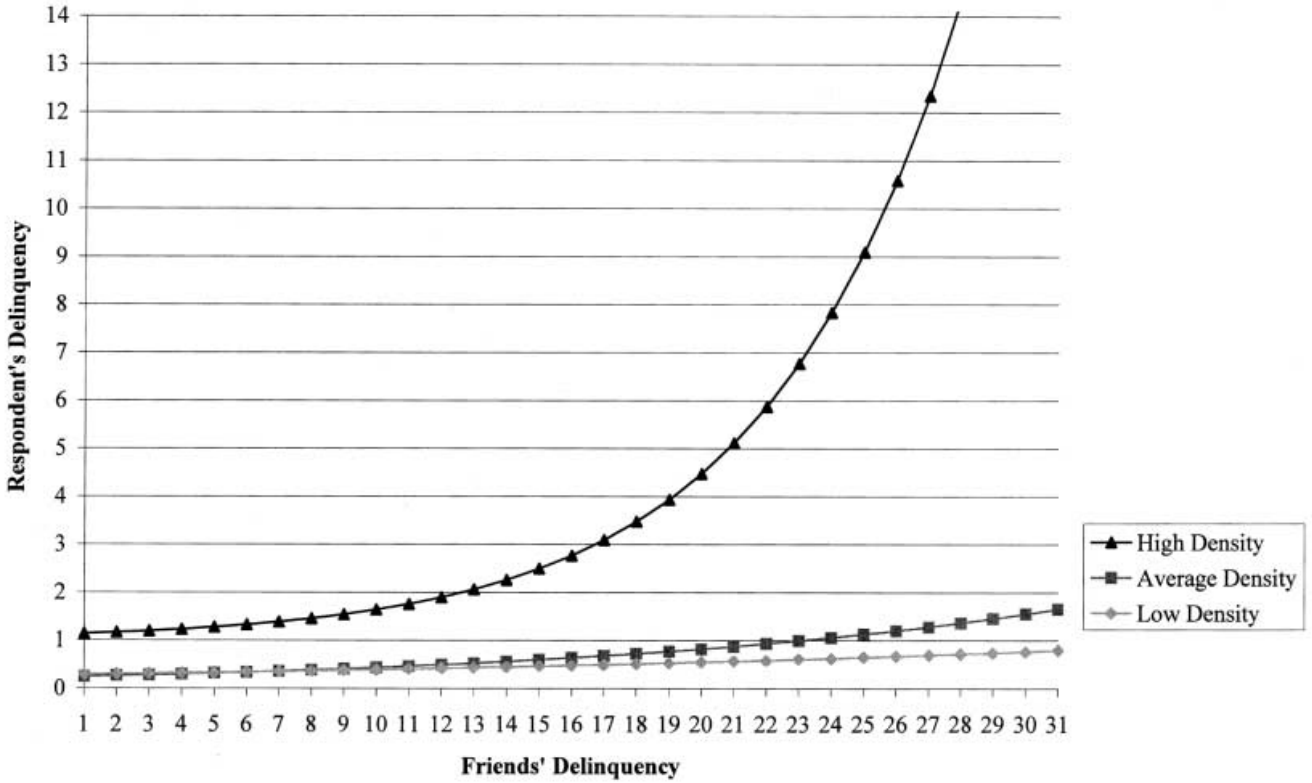
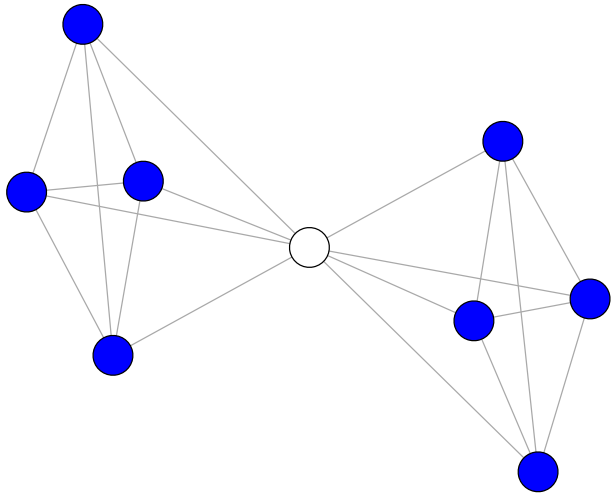


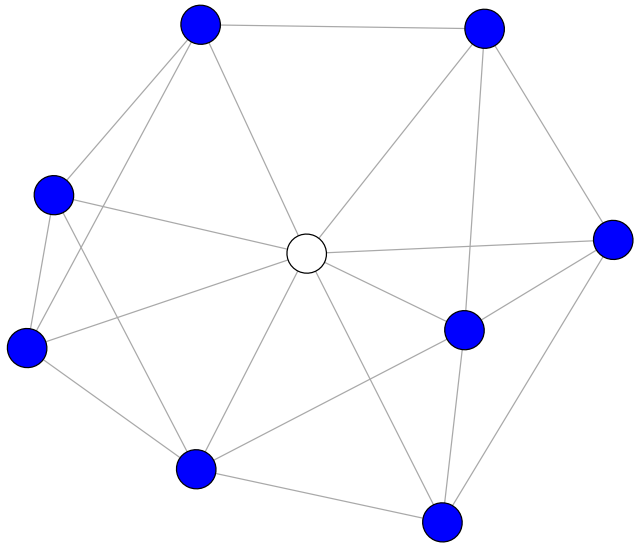
FIG. 1.—Interaction: density by friends' delinquency

Haynie 2010

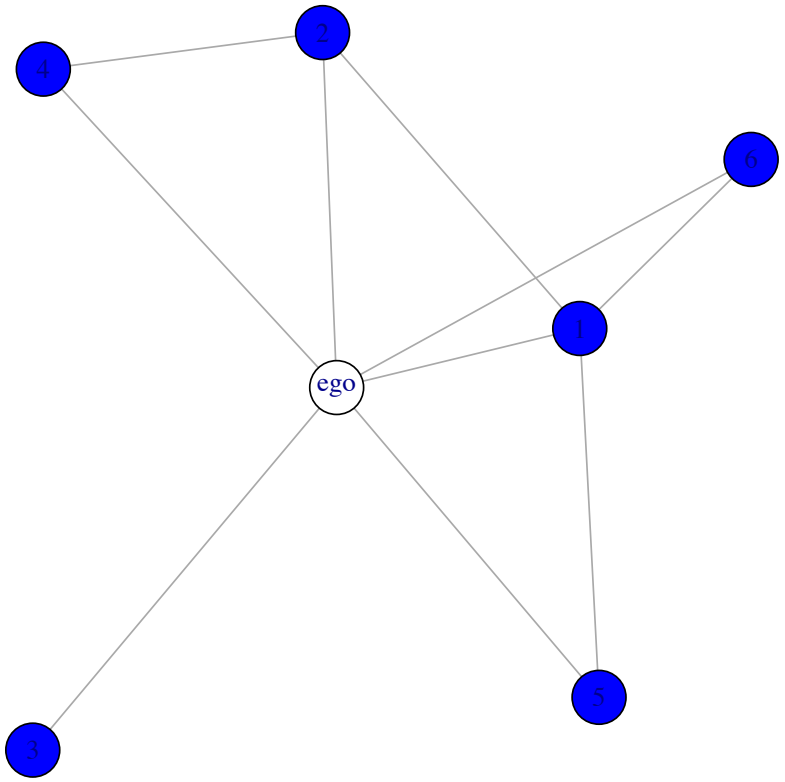
Structural Holes (Burt)

- Basic idea is that ego may be in an advantageous position if they have lots of non-redundant ties
  - More diverse information
  - Broker between two groups





Effective Size



Effective size=number of alters-'redundancy' with alters

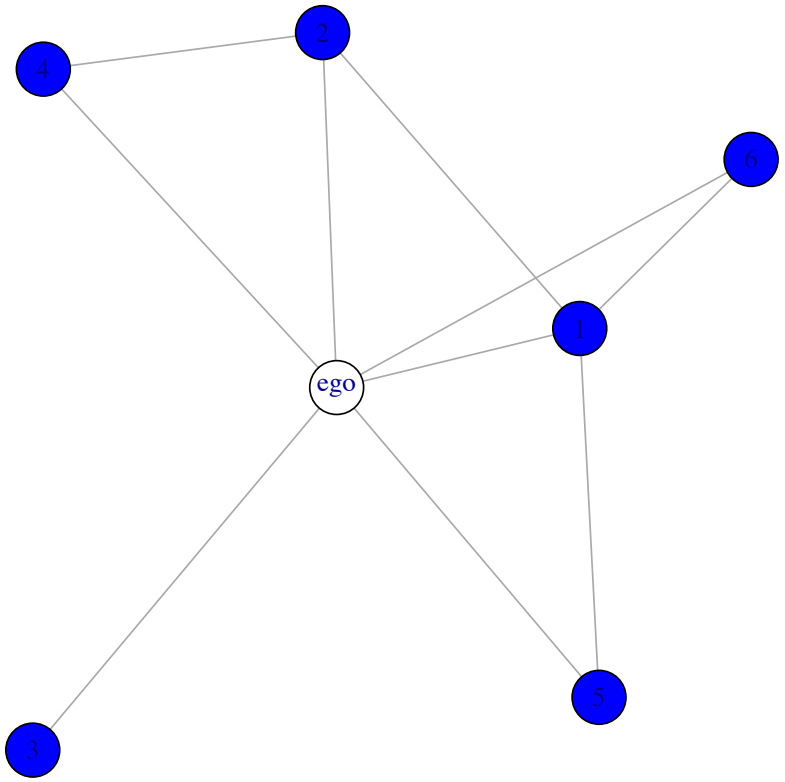
n-2t/n

Where n=number of alters

t=number of ties between alters

6-(2*4)/6=4.67

Efficiency



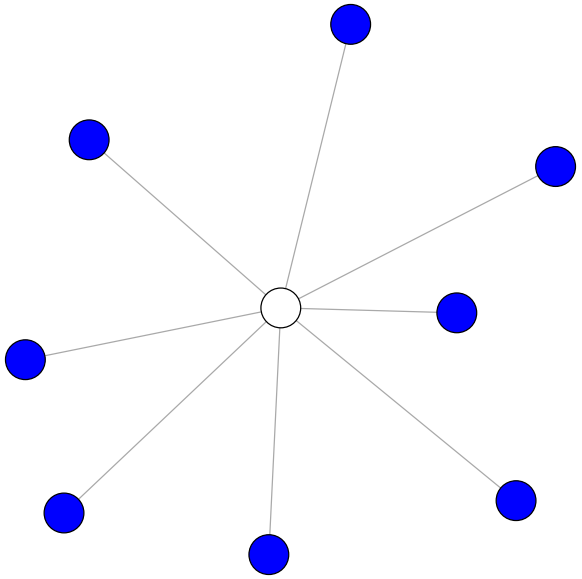
Efficiency =
Effective Size/actual size

4.67/6=.78

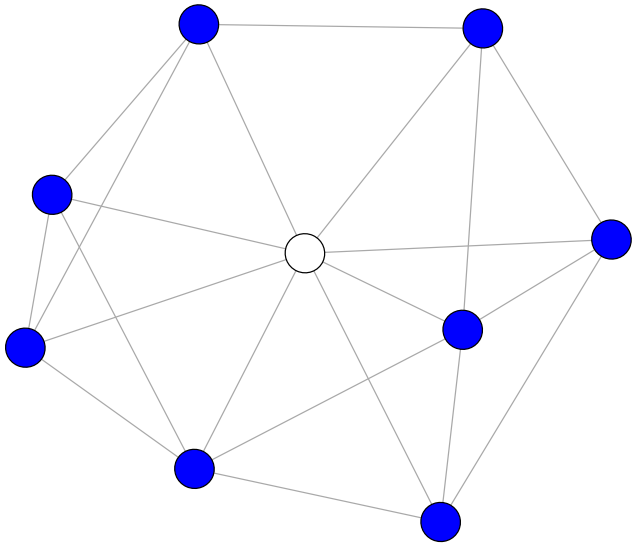
Constraint

- How much room to exploit structural holes
- To what extent is ego tied to people who are also connected to each other $c_{ij} = \left(p_{ij} + \sum_q p_{iq} p_{qj}\right)^2$
  - Less room to navigate if everyone tied to everyone

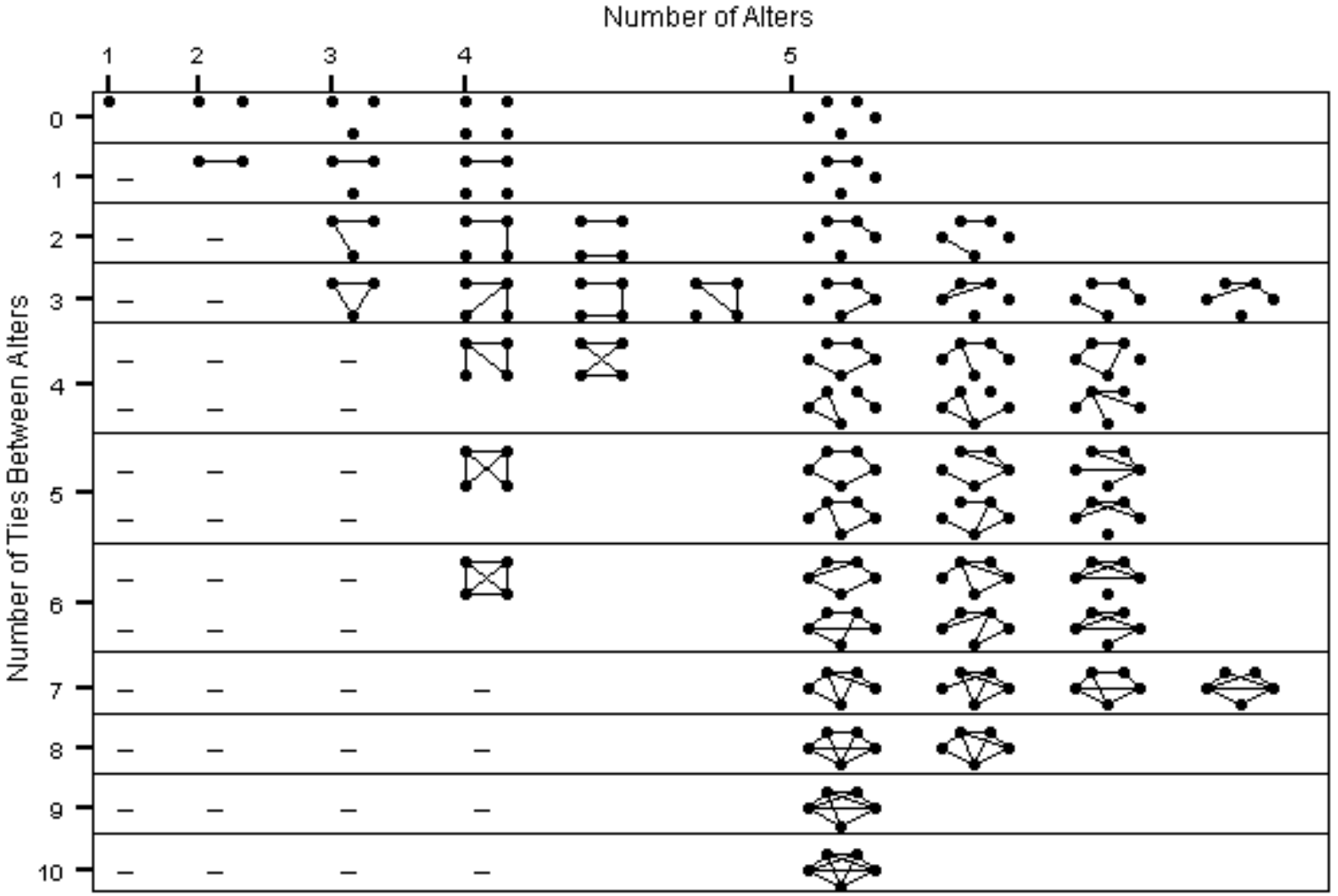
Low Constraint



Higher Constraint

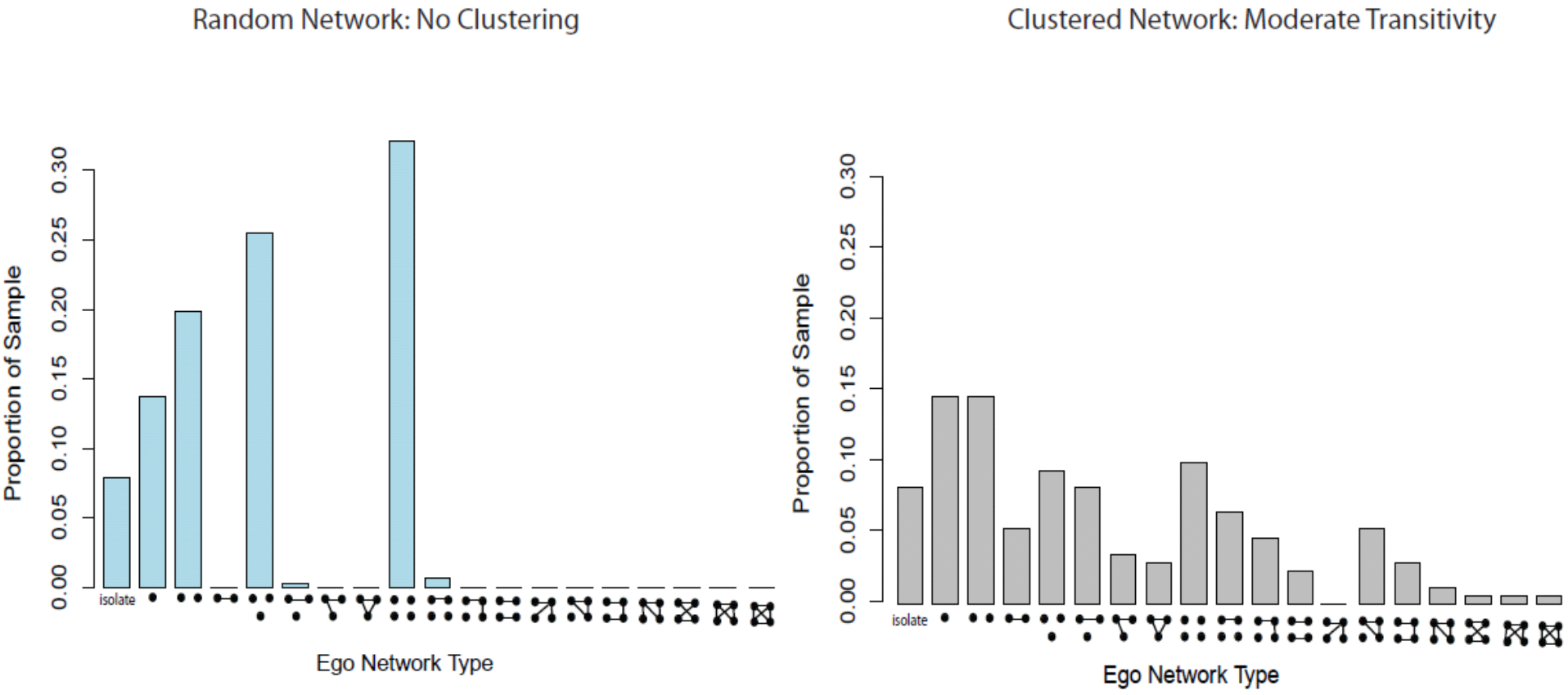


A Configurational Approach

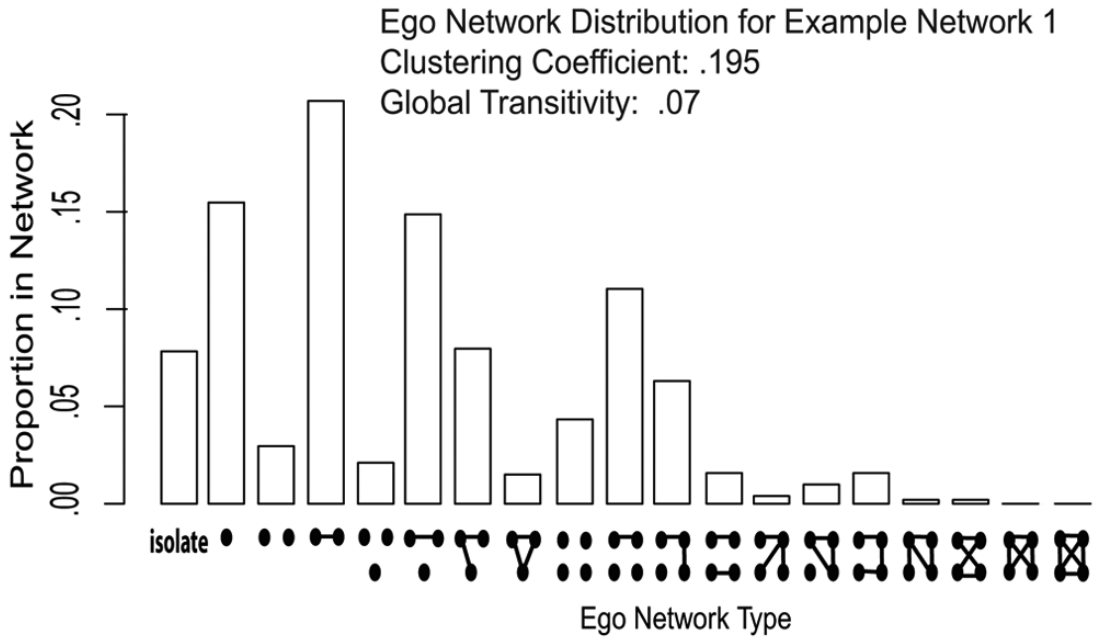


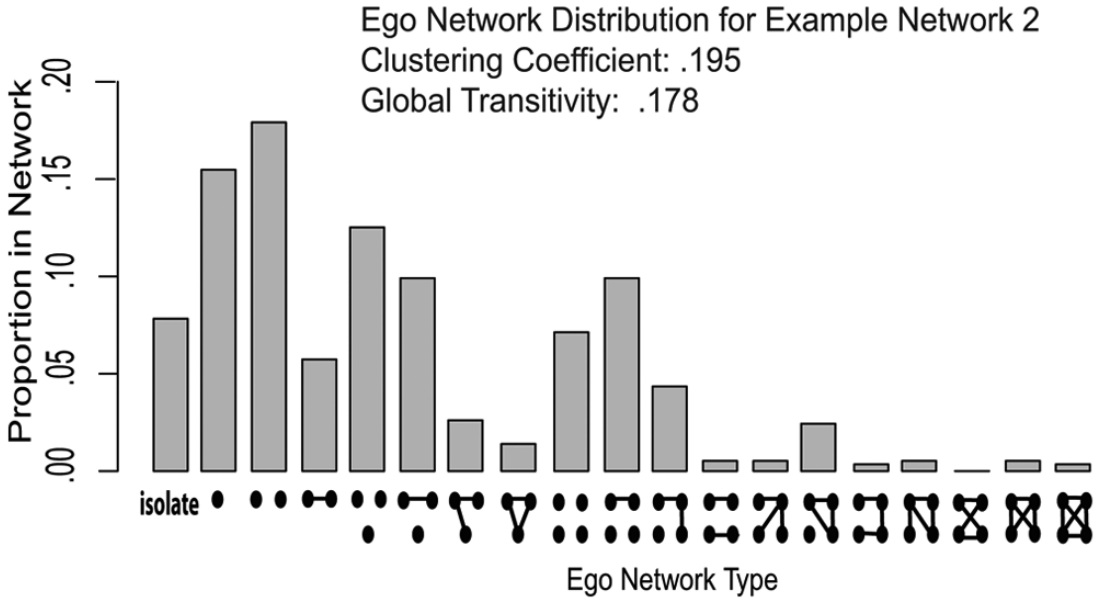
Smith 2012

Figure 2. Example Ego Network Configuration Distributions



Notes: ego is not shown in the configurations. I only use categories with four alters or below for space considerations. The random network was generated from a simple simulation while the clustered network was taken from one Add Health network.


Ego Network Distribution for Example Network 1
Clustering Coefficient: .195
Global Transitivity: .07
Proportion in Network
.00
.05
.10
.15
.20
isolate
Ego Network Type


Ego Network Distribution for Example Network 2
Clustering Coefficient: .195
Global Transitivity: .178
Proportion in Network
.00
.05
.10
.15
.20
isolate
Ego Network Type

Analysis Part II: getting global network features from ego network data

- Take ego network data and measure:
  - Degree distribution, mixing between groups,, differential degree and (possibly) ego network configuration distribution
- Simulate full networks of the right size (using ERGM-more on this Thursday!) where ego networks have same properties as observed ego networks

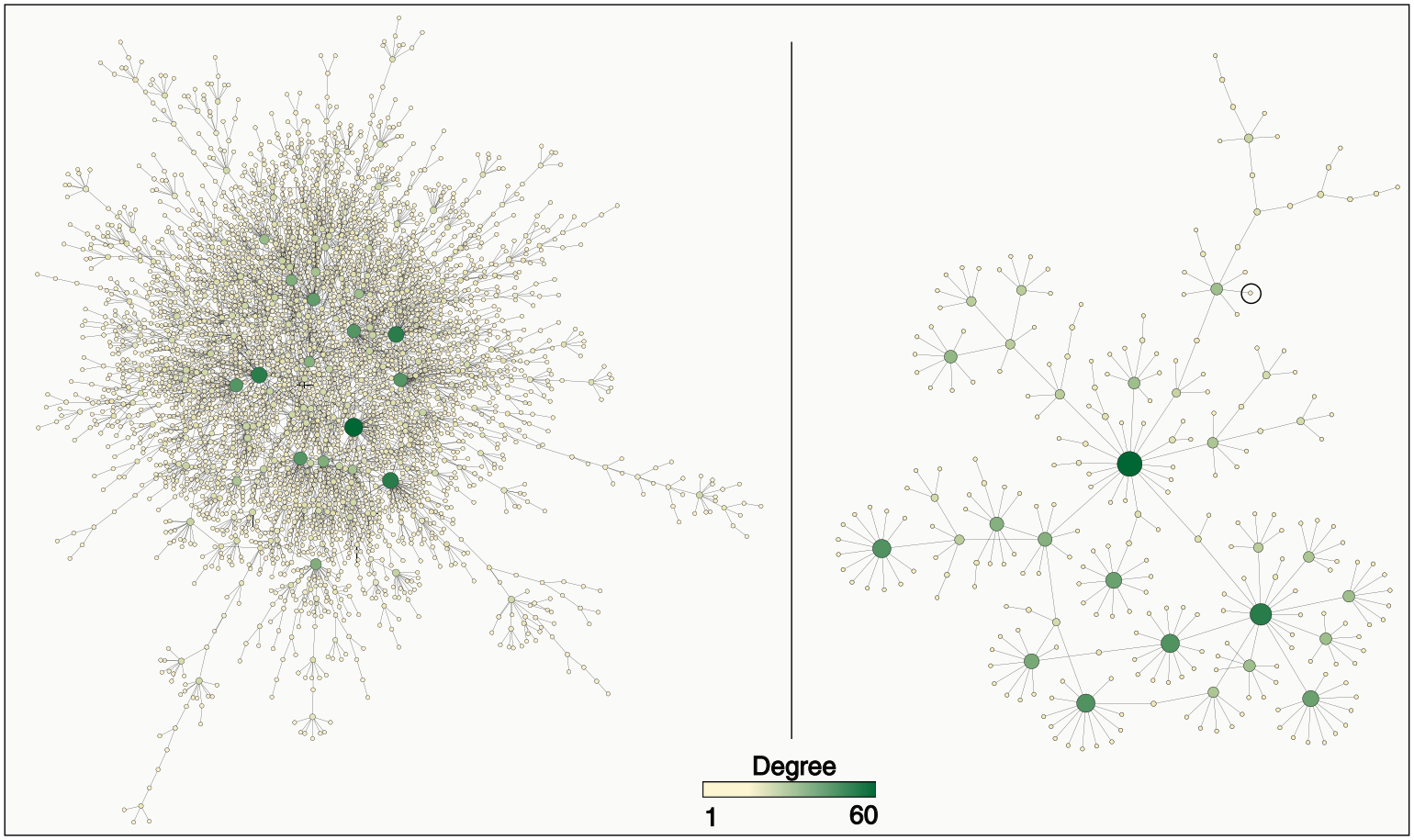
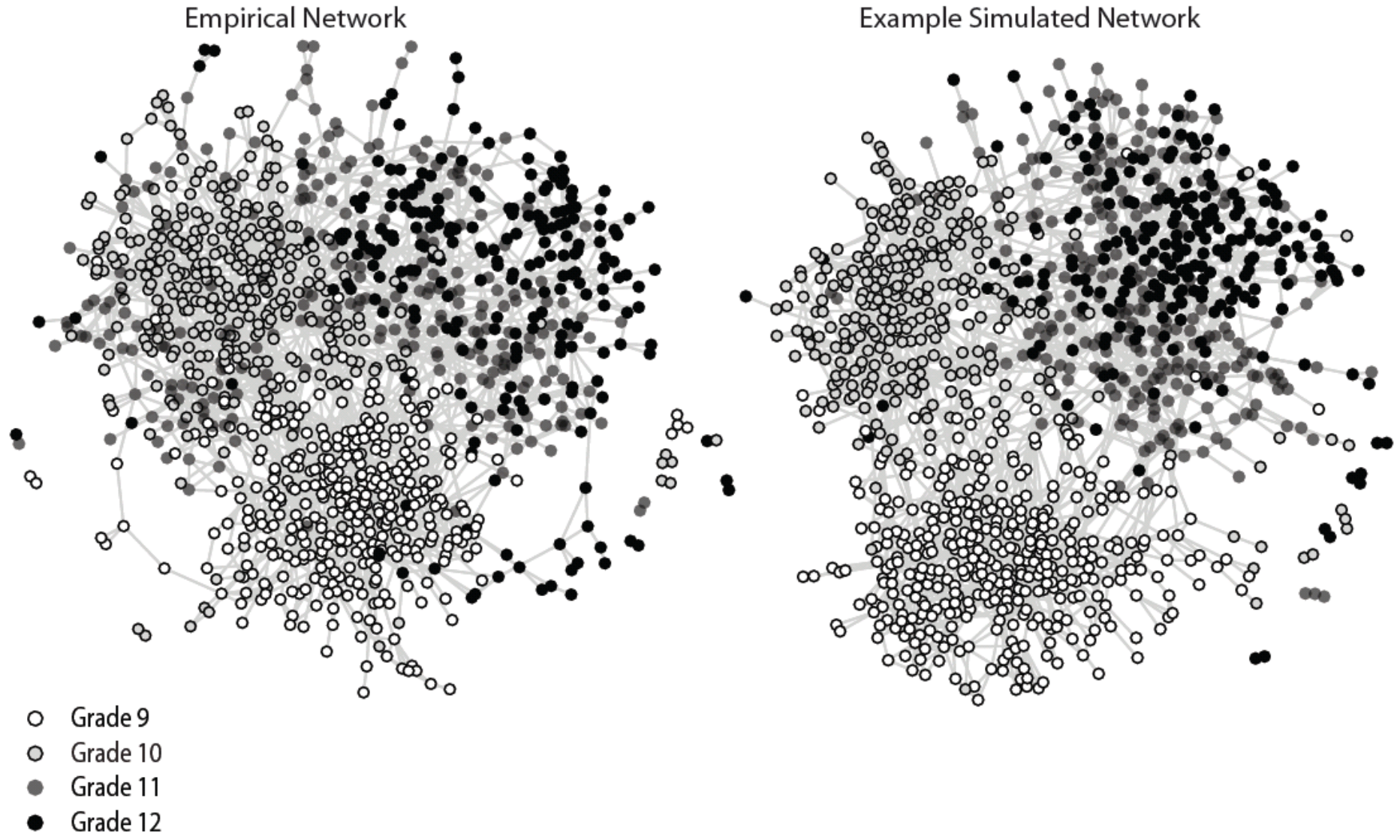


Fig. 6 Single run of largest component (left) and one eight-step walk from a randomly chosen node (right)

Merli et al. 2015

Figure 4. Comparing Add Health Network #6 with Example Simulated Network

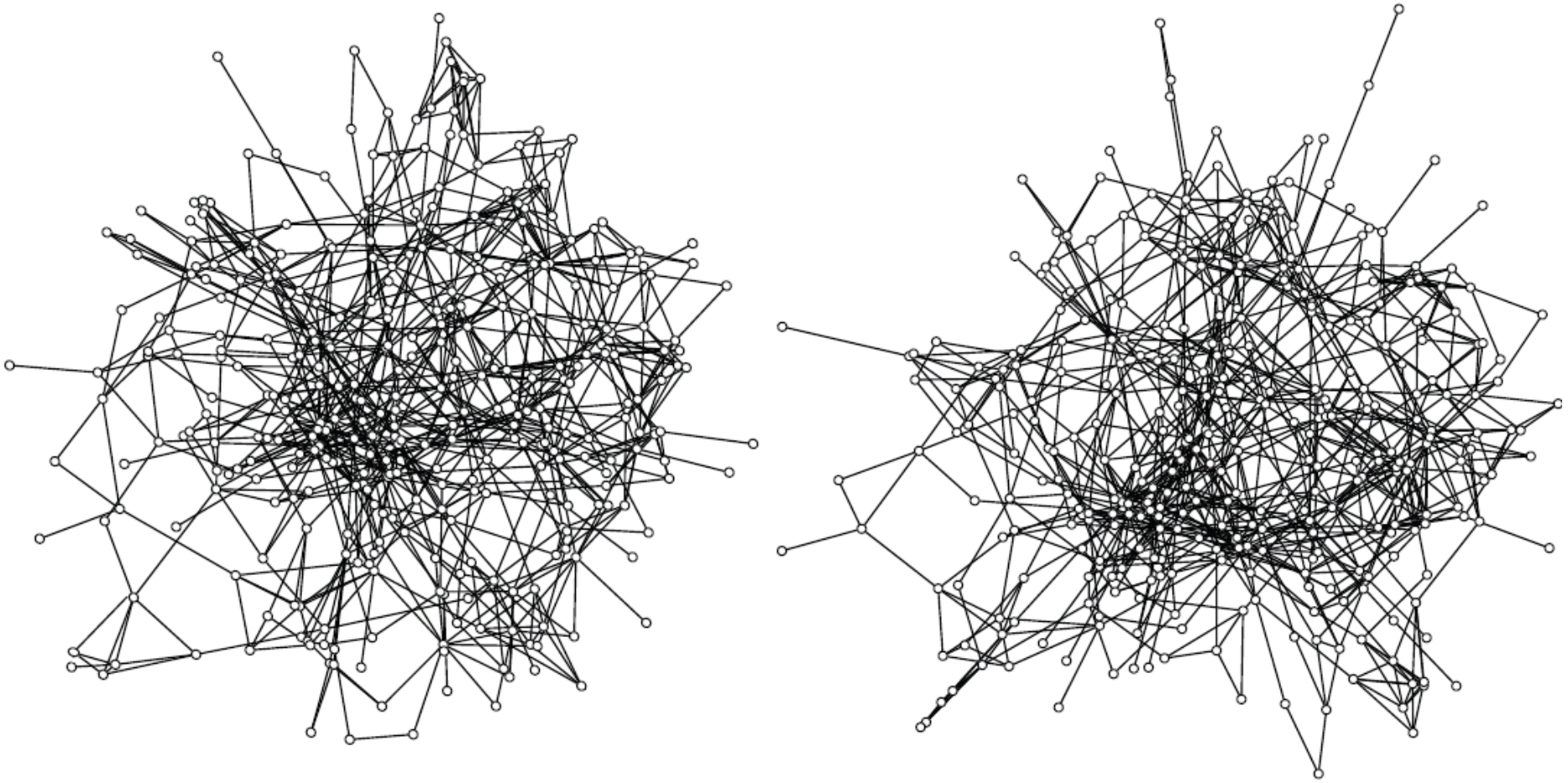


Smith 2012

Figure 5. Comparing Add Health Network #6 with Example Simulated Network: Grade 9 Only

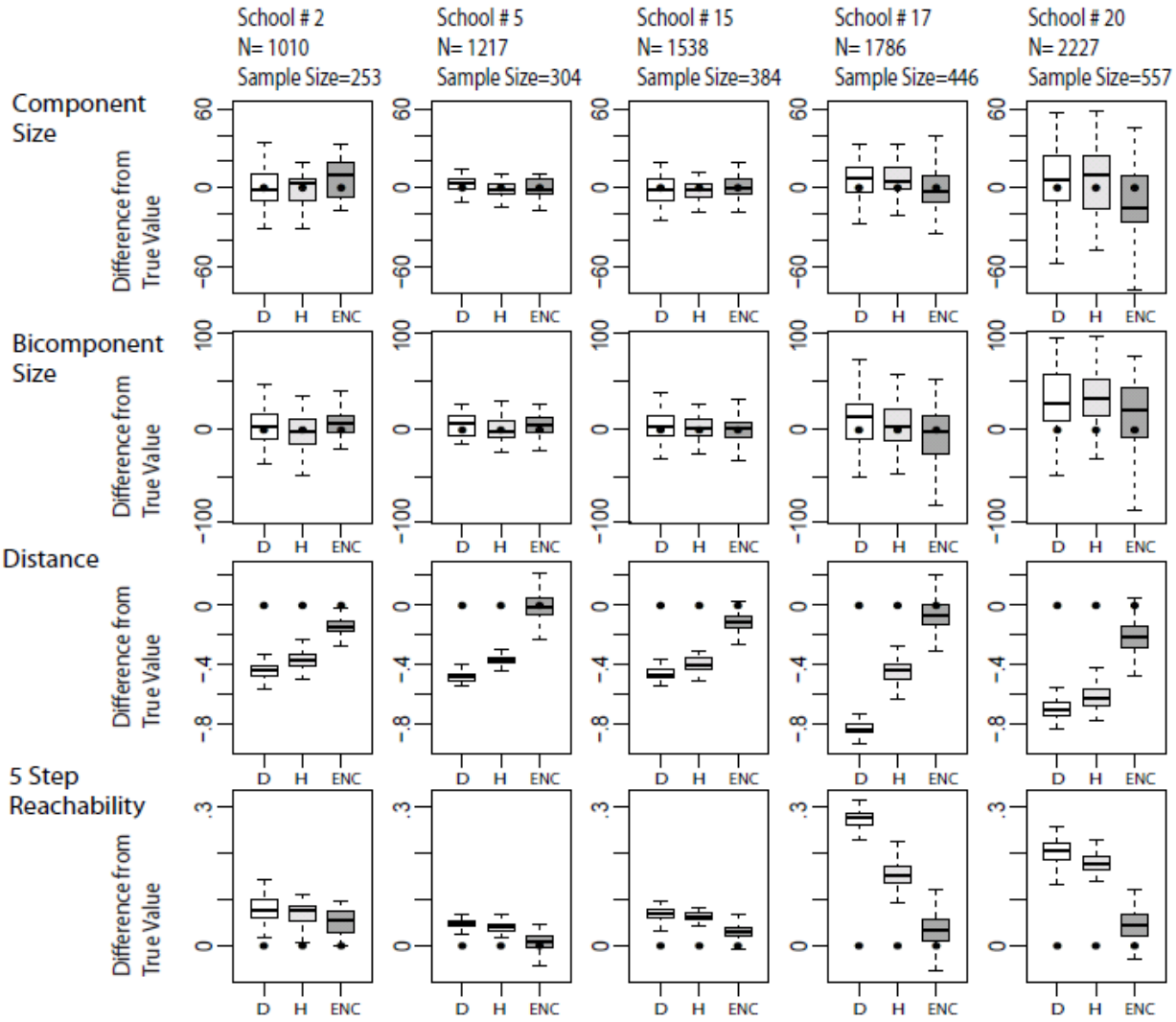
Empirical Network among Grade 9 Students

Example Simulated Network among Grade 9 Students



The plotted networks only include nodes in grade 9 in Add Health Network #6. The figure increases the resolution from Figure 4 in order to offer a more detailed look at network structure.

Figure 6. Comparison between True and Estimated Values for 5 Illustrative Add Health Schools: Connectivity Measures, 25% Ego Network Samples



This figure plots the difference between the estimated values and the true value for each method. The black circles note the zero point on the graph, where the estimates equal the true value. The boxplots are based on 30 iterations for each method/school; the black line indicate the median over all of the simulated networks and the edges of the box indicate the interquatile range.

Summary

- Ego network data are easy to collect and (potentially) useful for studying health
- Offers measures of social support (related to better health outcomes)
- Offers measures of risk/norms that an individual faces
- Offers information that can be used for inferring full network structure
  - Useful in understanding risk of disease spread