Social Network Analysis: Overview Part 2

EPIC - SNA, Columbia University

Zack W Almquist

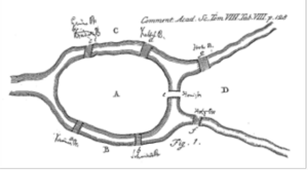
June 11th, 2018

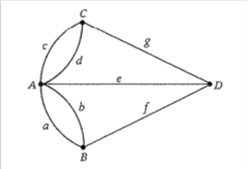
University of Minnesota

Network Vizualization: Theory and Methods

Visualization

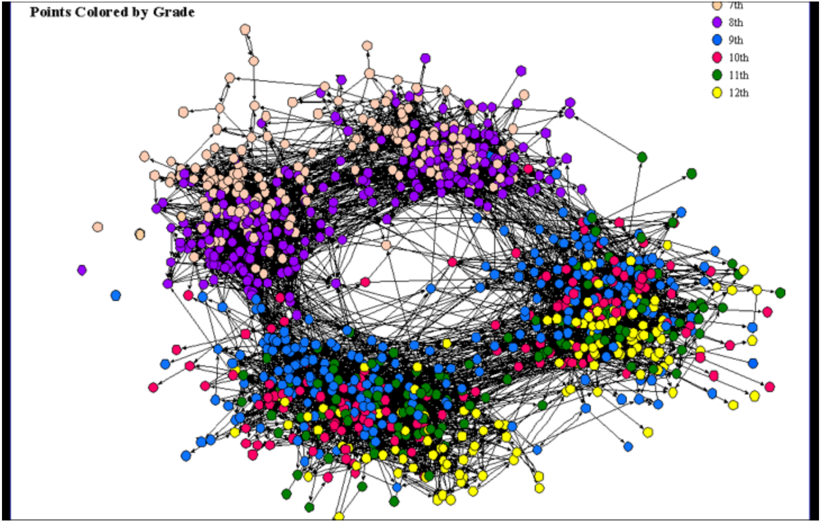
- "The success of a visualization is based on deep knowledge and care about the substance, and the quality, relevance and integrity of the content." (Tufte, 1983)
- Thus a network graph's aim is to clearly communicate something (about patterns of social relationships) that we would have difficulty knowing any other way.





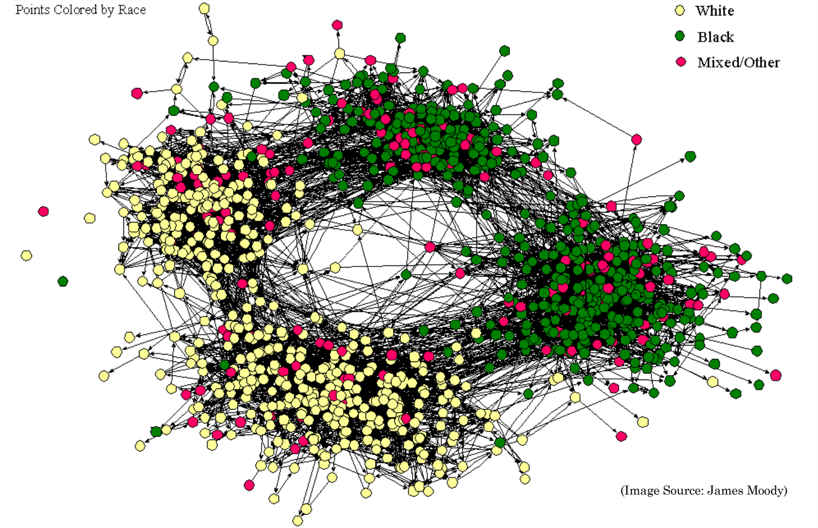
Euler 1741, as reproduced in Freeman LC. Visualizing Social. Journal of Social Structure 2000;1(1).

Visualization: Gaining Some Intuition



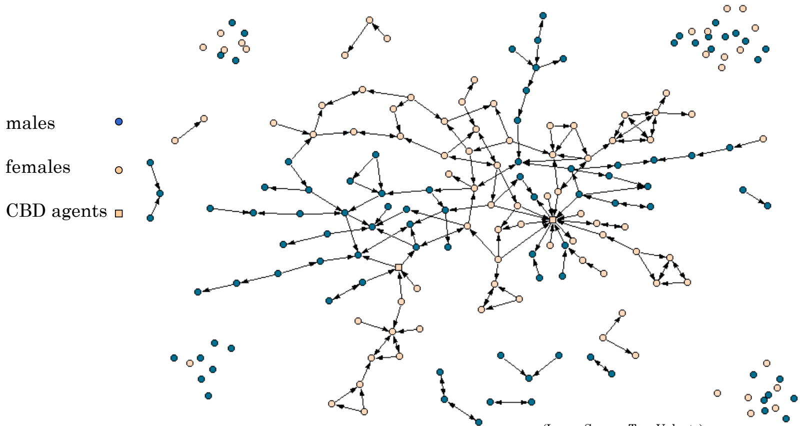
jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin:

Visualization: Gaining Some Intuition



Visualization: Gaining Some Intuition

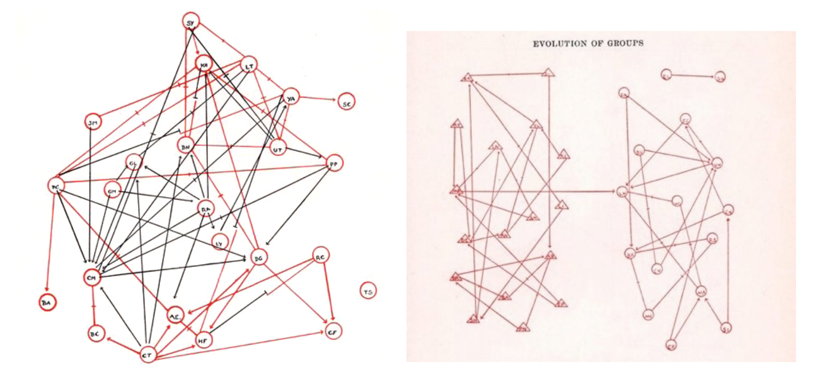
Figure 1: The general advice network of the village of ***Mandrosohasina***, 1999



(Image Source: Tom Valente)

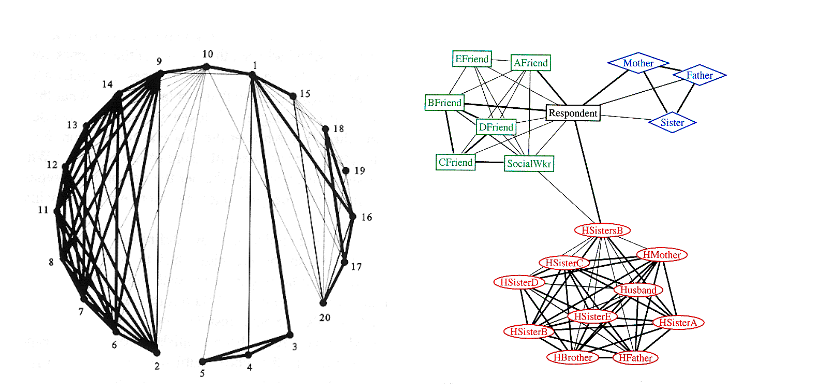
jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin:

Visualization: Gaining Some Intuition



jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin: Moreno 1934, as reproduced in Freeman LC. 2000. Visualizing Social Networks. Journal of Social Structure 1(1).

Visualization: Gaining Some Intuition



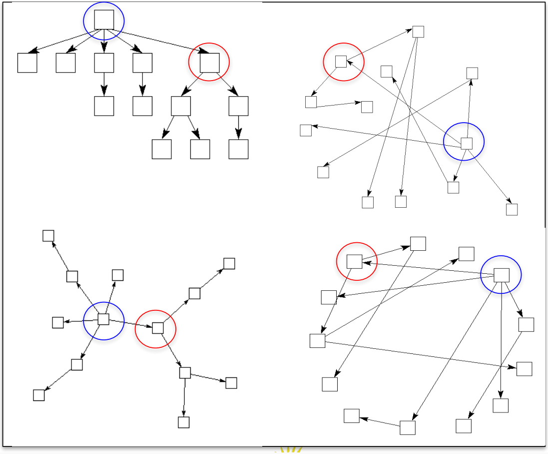
jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin: Mitchell 1994, as reproduced in Freeman LC. 2000. Visualizing Social Networks. Journal of Social Structure 1(1).

Visualization: Gaining Some Intuition

Consider these 4 graphs.

What intuition do you gain from each?

They're the exact same network, laid out differently.



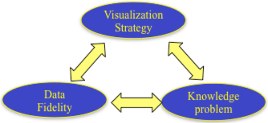
Visualization: Gaining Some Intuition

Information

- Build intuition about the social process generating the network
- Succinctly capture high-dimensional properties of the network
- Maximize Ink to Information ratio
- There is value in beauty (e.g., memory)

Scientifically

- Ability to *replicate* results – same data should produce same picture
- Maximize translation of graph's features
  to quantifiable measurement of the graph
- Theory-relevant

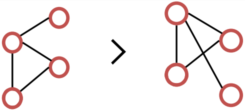


jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin: Moody J, McFarland DA, Bender-DeMoll S.network visualization. American Journal of Sociology 110(4):1206-1241.

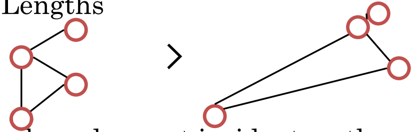
Visualization: Gaining Some Intuition

Practicalities

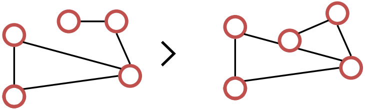
- Minimize edge crossings



- Uniform Edge Lengths

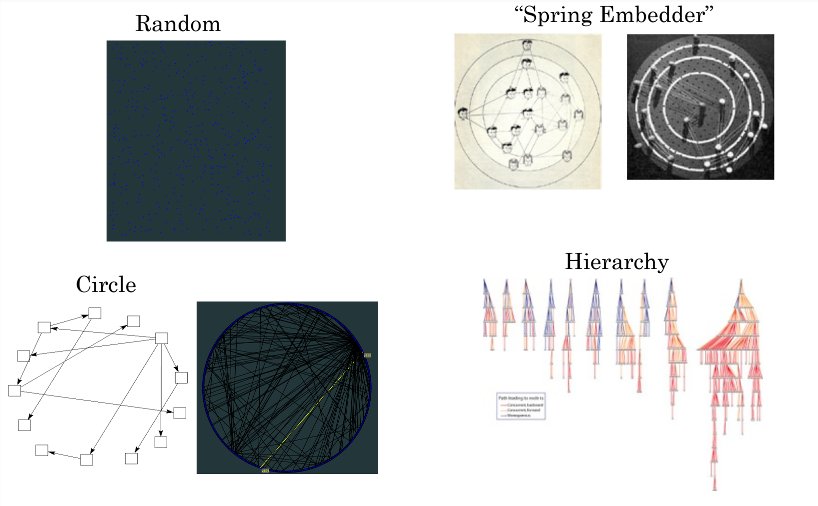


- Nodes don’t overlap edges not incident on them



jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin:

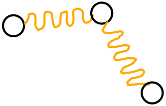
Visualization: Gaining Some Intuition



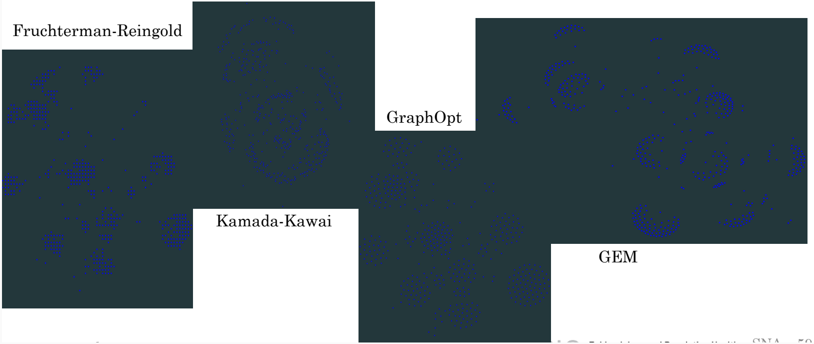
jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin: Lada Adamic

Visualization: Gaining Some Intuition

Spring-Embedder Algorithms

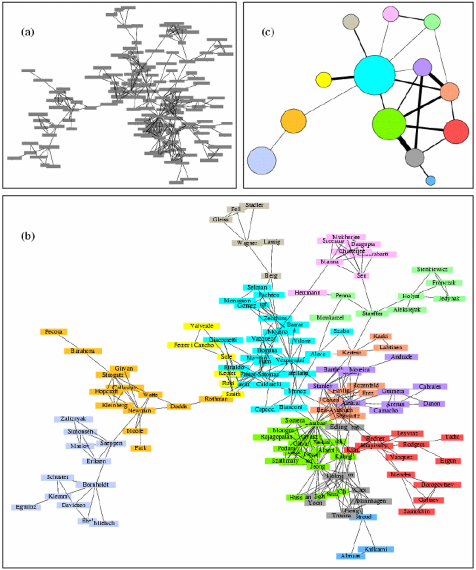


- Connected nodes' springs pull them together
- Disconnected nodes' springs repel one another.
- When balancing these, visualized network's physical space corresponds roughly to social space via network distance/clustering.



jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin:

Visualization: Gaining Some Intuition

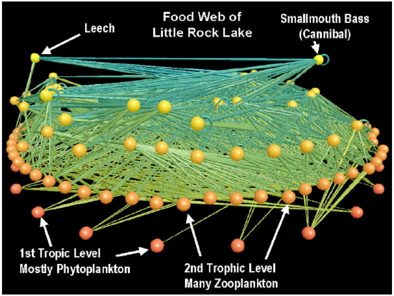


Example of coarsening network structure

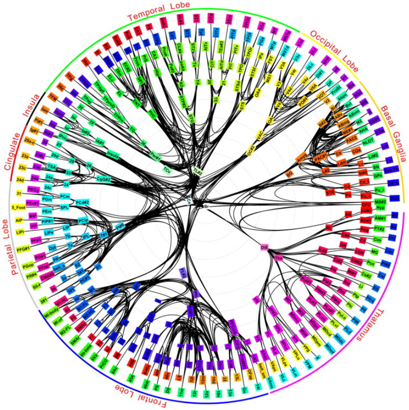
- Newman & Girvan 2004
- co-authorship network of physicists writing papers on networks
- clustering algorithm identifies different subcommunities
- each node is a community – size represents number of authors
- each edge thickness represents the number of co-author pairs between communities

(Image Sources: Lada Adamic)

Visualization: Gaining Some Intuition

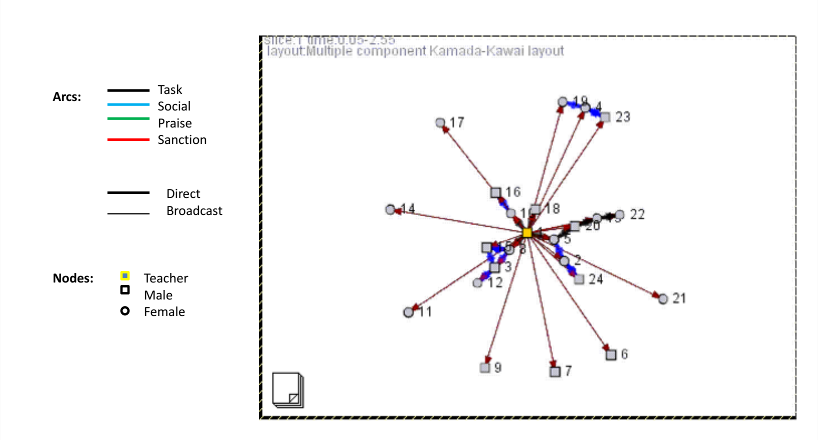


http://news.bbc.co.uk/2/hi/science/nature/2288621.stm



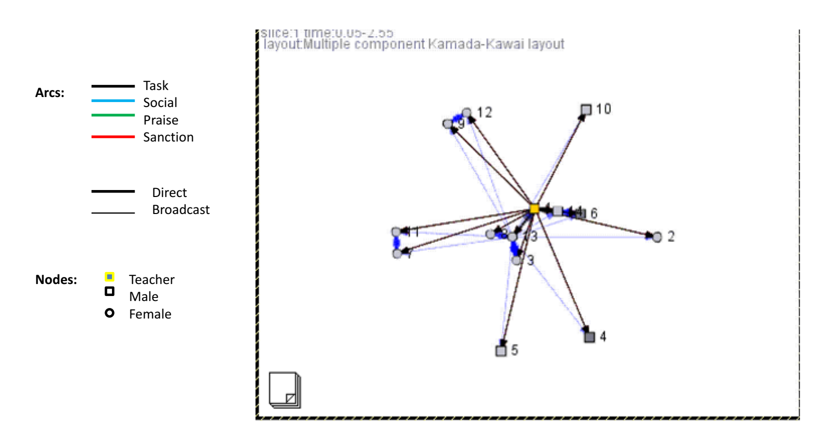
PNAS;107(30):13485-13490.

Visualization: Gaining Some Intuition



jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin: Moody J, McFarland DA, Bender-DeMoll S. 2005. network visualization. American Journal of Sociology 110(4):1206-1241.

Visualization: Gaining Some Intuition



jimi adama. EPIC - SNA 2017. Columbia, Univeristy. Origin: Moody J, McFarland DA, Bender-DeMoll S. 2005. network visualization. American Journal of Sociology 110(4):1206-1241.

Visualization: Gaining Some Intuition

Additional considerations:

- timescale, pace, sampling & representations
  - events vs. relationships
  - discrete vs. continuous time
  - binning (min, max, sum, mean, etc.)
- tie weighting/valuation/direction
- isolates, disconnected components
- initial position
  - random w/ repetition
  - data derived / provided (e.g., chaining)

Network Vizualization: R

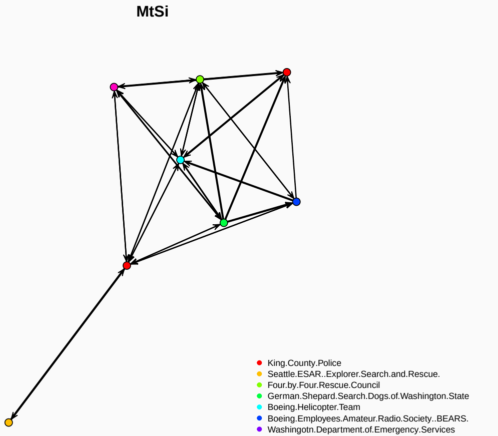
Network Vizualization: R

Why R?

- Open Source
- Large community of developers for SNA
    - STATNET, igraph, ...
- Extensible

Network Data Collection

One-mode data



One-mde data

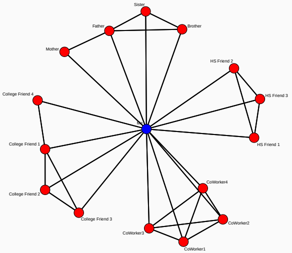
	1	2	3	4	5	6	7	8
1	0.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00
2	1.00	0.00	1.00	1.00	1.00	1.00	0.00	0.00
3	1.00	1.00	0.00	0.00	1.00	1.00	0.00	1.00
4	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00
5	1.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00
6	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00

Table 1: Mt. Si SAR EMON, Confirmed Ties

Special Case: Egocentric network

Egocentric network: focal actor (“ego”) + neighbors (“alters”) + ties among alters

- **What does it tell us?**
  - Number of ties ego has (neighborhood size)
  - Triangles (3-cliques) containing ego
  - Connections among alters
  - Neighborhood composition (if asked)
  - Note: Sometimes called *personal networks*



Two mode data

- Networks with two vertex class
  - Different entity types
  - Membership
  - Matching/containment
- Represented by incidence matrices
  - "Senders" on rows, "receivers" on columns
- Can be used to obtain "dual" representations

Two mode data: Projecting into one mode data

- Let A be an $N \times M$ incidence matrix; the row-projection of A is the $N \times N$ matrix B such that

$$B_{ij} = \sum_{k=1}^{M} A_{ik} A_{jk}$$

- Likewise, the column projection of A is the $M \times M$ matrix C such that

$$C_{ij} = \sum_{k=1}^{N} A_{kj} A_{kj}$$

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 2 \end{pmatrix}, \text{ and } C = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

Two mode data: Projecting into one mode data

What the Projections Mean

- Projections have simple meaning
  - Row: B_{ij} is the number of column elements shared by row elements i and j
  - Column: C_{ij} is the number of row elements shared by column elements i and j
- Ex: Number of shared interests between two faculty; number of faculty having a given interest area in common

Two mode data: Projecting into one mode data

What the Projections Mean

- To analyze network data, we must first collect it!
  - Many approaches exist – some better than others for particular purposes
  - Complex topic overall, but we will at least skim the surface...
- Two important concepts (not always separable):
  - **Instruments**: tools used to elicit information from respondents, assess presence/absence of ties from sensors or archival materials, etc.
  - **Designs**: protocols for determining how information should be elicited, who should be sampled, etc.

Levels of Analysis

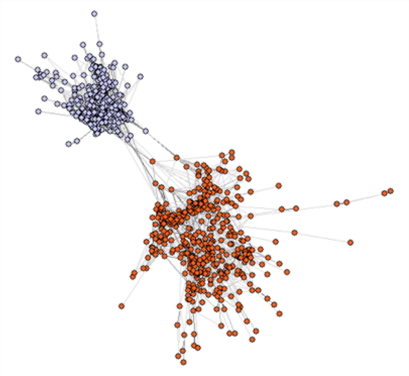
What scope of information do you want?

- Boundary Specification: key is what constitutes the "edge" of the network

	LOCAL	GLOBAL
"Realist" (Boundary from actors' Point of view)	Everyone connected to ego in the relevant manner (all friends, all (past?) sex partners)	All relations relevant to social action ("adolescent peers network" or "Ruling Elite")
Nominalist (Boundary from researchers' point of view)	Relations defined by a name-generator, typically limited in number ("5 closest friends")	Relations within a particular setting ("friends in school" or "votes on the supreme court")

Levels of Analysis

Boundary Specification Probblem



While students were given the option to name friends in the other school, they rarely do. As such, the school likely serves as a strong substantive boundary

Relational Data - Designs

- Own-tie reports
    - Personal ties elicited from each ego
    - Standard instruments: roster and name generator
    - Pros: Easily implemented, most common design
    - Cons: Vulnerable to reporting error
- Egocentric network sampling
    - Personal ties elicited from ego, followed by induced ties
    - Standard instrument: name generator followed by roster
    - Pros: Well-suited to large-scale survey sampling; provides information on ego's neighborhood
    - Cons: Vulnerable to reporting error; false positives/negatives on own ties contaminate sampling of neighbors' ties

Instruments: Name Generators and Rosters

- Name generator: asks respondents to list names
  - E.g. "Think about the persons with whom you have talked in the past week. Please list all such persons in the following space." (followed by space to enter names)
- Pros:
  - Don't have to know name list; can use with large groups or organizations
- Cons:
  - High rate of forgetting; unclear boundary

Roster: asks respondents to choose names from fixed list

- E.g. “For each of the following persons, place a check in the associated blank if you have talked with him her in the past week.” (followed by check list)
  - Pros: More accurate, clear boundary
  - Cons: List may be prohibitively long, can be imposing; alters must be known in advance

Instruments: Complete Ego Ne

- Common way to elicit ego nets: complete instrument followed by roster
  - Asked to name those with whom you discussed important matters
  - Then, asked to fill in same question for all pairs of persons named initially
- Pros:
  - Relatively easy to administer; don't need entire list of possible alters; don't have to ask about all group members
- Cons:
  - Step 1, step 2 questions have different error rates; may need large roster if many alters; hard to use with paper-based surveys

GSS: Important Matters

- Famous Example: General Social Survey

  From time to time, most people discuss important matters with other people. Looking back over the last six months—who are the people with whom you discussed matters important to you? Just tell me their first names or initials.

- IF LESS THAN 5 NAMES MENTIONED, PROBE: Anyone else?

GSS: Important Matters

- Famous Example: General Social Survey

  From time to time, most people discuss important matters with other people. Looking back over the last six months—who are the people with whom you discussed matters important to you? Just tell me their first names or initials.

- Survey
- Let's take a minute and answer this question (you will get lab credit)

GSS: Important Matters

- Famous Example: General Social Survey

  Known as important matters question

- Basic findings

  McPherson, Smith-Lovin, Brashears, "Social Isolation in America: Changes in Core Discussion Networks over Two Decade" ASR 2006

- The number of people saying there is no one with whom they discuss important matters nearly tripled. The mean network size decreases by about a third (one confidant), from 2.94 in 1985 to 2.08 in 2004.
- The modal respondent now reports having no confidant; the modal respondent in 1985 had three confidants

GSS: Important Matters

- Famous Example: General Social Survey

  Known as important matters question

- Basic findings

  Small et al. "How stable is the core discussion network." Social Networks 2015.

- First, we found that when actors enter new institutional environments, their core discussion network changes rather quickly
- Second, our study introduced a new framework to understand the evolution of core discussion networks, or other ego networks, that begins with the two basic processes – addition or subtraction and replacement or non-replacement – while

Designs: Link-tracing

- Personal ties elicited from ego; new ego(s) chosen from alters; process is iterated (possibly many times)
- Standard instruments: multiwave own-report, RDS
- Pros: Allows estimation of network properties for large and/or hard to reach populations; highly scalable; can be robust to poor seed sampling
- Cons: Vulnerable to reporting error; reporting errors can contaminate design (but may be less damaging than ego net case); often difficult to execute

Arc Sampling * Reports on third-party ties elicited from ego; multiple egos may be sampled for each third-party tie

```
* Archival/observer data is a special case Standard instru
* Pros: Very robust to reporting error (via modeling); can
* Cons: Can impose large burden on respondents; can be dif
```

Designs: RDS

- Respondent Driven Sampling (RDS)
  - Combine standard network instrument with recruitment “tickets”
  - Respondents given tickets to give to others; if they volunteer, both get paid
- Pros:
  - Can use with hidden, vulnerable populations
- Cons:
  - Difficult; expensive; complex to analyze; poorly understood

Designs: Link-tracing

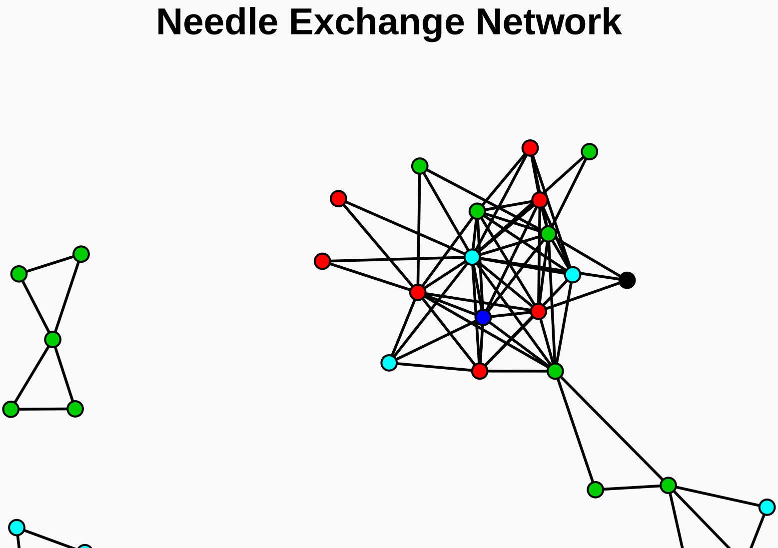
The Data The data was aggregated by Martina Morris (University of Washington) and Richard Rothenberg (Emory University) and put online at ICPSR. The original data can be found here. In this exercise we are goint to investigate four networks derived from the Rural Arizona risk networks in Flagstaff, AZ. These networks were collected from May 1996 to Jan 1998 and originally had 95 respondents interviewed 5 times each. All participants are over 18 years old.

- Name generator
    - Sex, needle, other (illicit) drug contact, social contact in last 6 months
    - Sampling strategy
        - Six seeds chosen at random within same geographic area (Flagstaff) from persons presumed to be at elevated risk for HIV acquisition (through sex and/or drug behaviors)

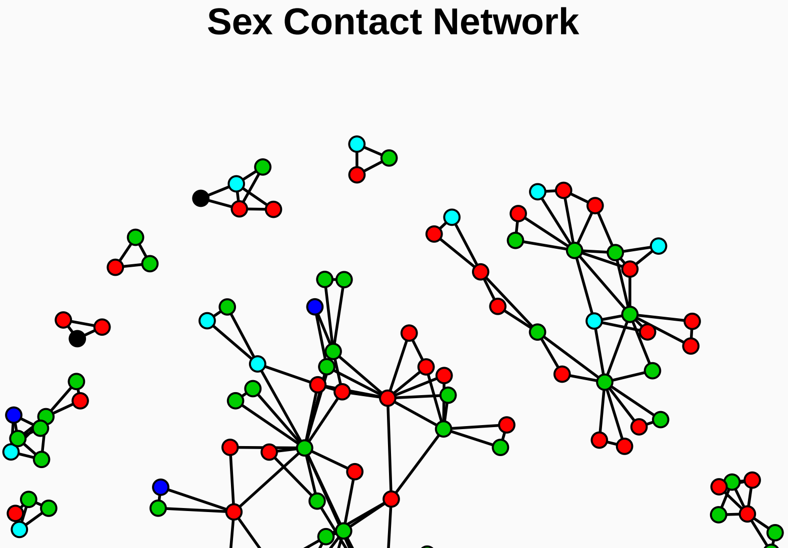
Designs: Link-tracing

```r
library(sna)
library(network)
load(url("https://github.com/zalmquist/ERGM_Lab/raw/master/
```

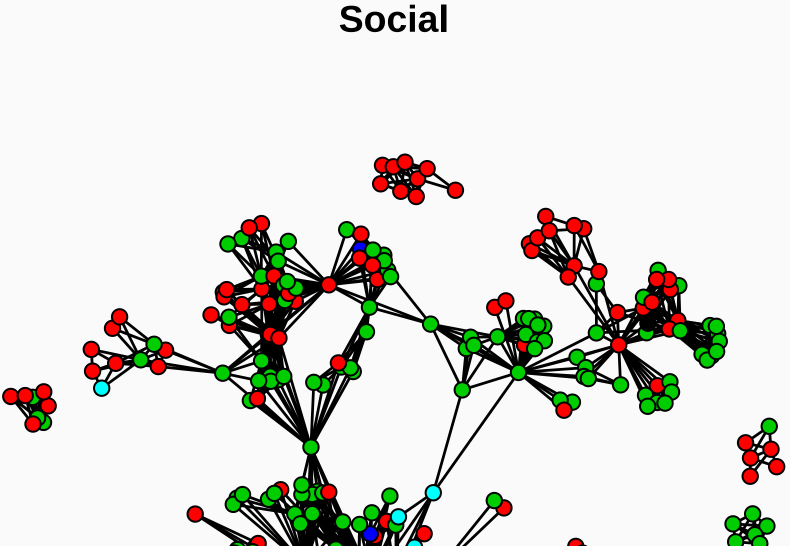
```
plot(flag_needle_net, vertex.col = "race", main = "Needle E
```



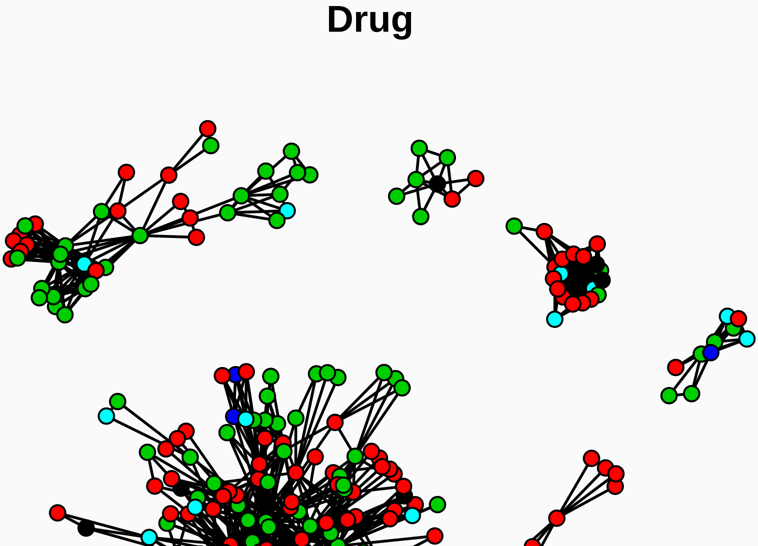
```
plot(flag_sex_net, vertex.col = "race", main = "Sex Contact
```



```
plot(flag_social_net, vertex.col = "race", main = "Social")
```



```
plot(flag_drug_net, vertex.col = "race", main = "Drug")
```



Designs: CSS

- Cognitive Social Structure (CSS)
  - Ask each group member to report on all members‘ ties
  - Ex: “Which of the following persons does Steve go to for help or advice?”
- Pros:
  - Gets information on perception; can be used to get high-accurate estimates
  - Cons:
- Hard to use; requires roster; doesn’t scale well

Designs: CSS

The Data David Krackhardt collected cognitive social structure data from 21 management personnel in a high-tech, machine manufacturing firm to assess the effects of a recent management intervention program. The relation queried was

- "Who does X go to for advice and help with work?" (`krackad`)
- "Who is a friend of X?" (`krackfr`).

Each person indicated not only his or her own advice and friendship relationships, but also the relations he or she perceived among all other managers, generating a full 21 by 21 matrix of adjacency ratings from each person in the group.

Designs: CSS

```r
library(networkdata)
data(krack)
length(krack[[1]])
```

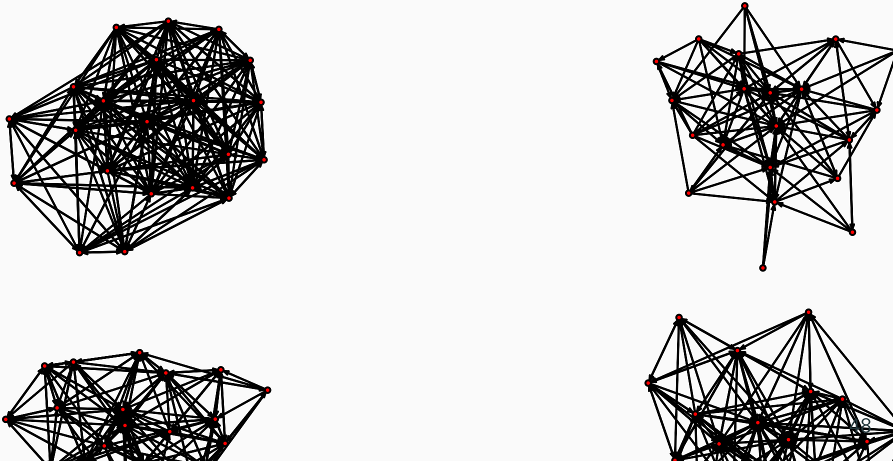
```
[1] 21
```

```r
length(krack[[2]])
```

```
[1] 21
```

Designs: CSS

```r
par(mfrow = c(3, 3), mar = c(0, 0, 0, 0) + 0.1)
for (i in 1:9) plot(krack[[1]][[i]])
```



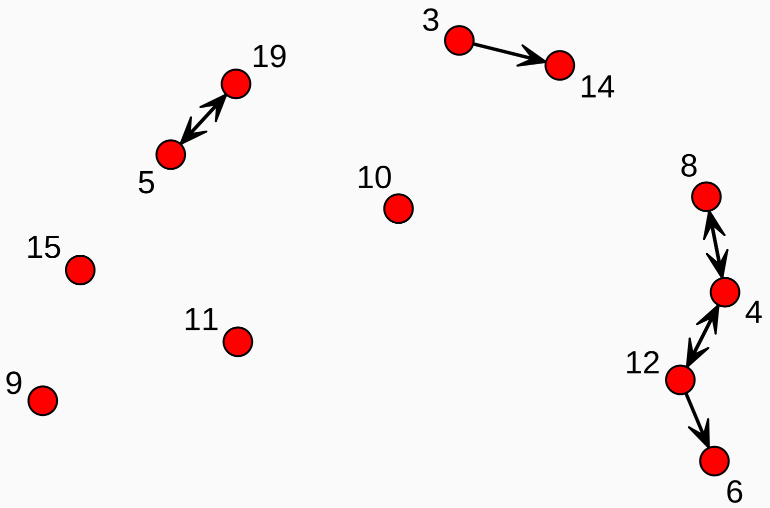
Designs: CSS

```
kfr <- as.sociomatrix.sna(krack$krackfr)
### Baysian network inference Let's start by defining some
### Bayesian network inference model.  We'll use an uninfor
### prior, together with weakly informative (but diffuse ar
### priors on the error rates.  Read the man page ('?bbnam
### information about how the routine works.
np <- matrix(0.5, 21, 21)  # 21 x 21 matrix of Bernoulli pa
emp <- sapply(c(3, 11), rep, 21)  # Beta(3,11) priors for f
epp <- sapply(c(3, 11), rep, 21)  # Beta(3,11) priors for f
# hist(rbeta(100000,3,11)) # This gives you a sense of what
# look like!

# Now, let's take some posterior draws for the friendship n
# using various models (warning: slow)
kfr.post.fixed <- bbnam.fixed(kfr, nprior = np, em = 3/(3 -
```

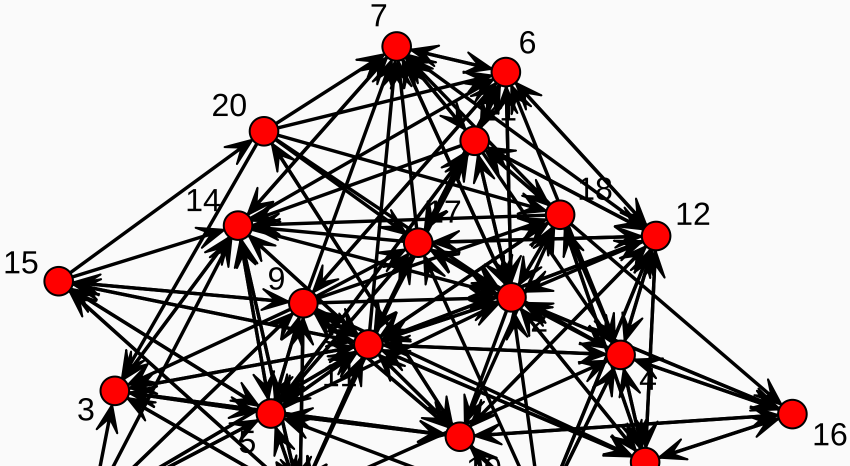
Designs: CSS

```
gplot(apply(kfr.post.fixed$net, c(2, 3), median), displayla
    boxed.lab = FALSE)
```

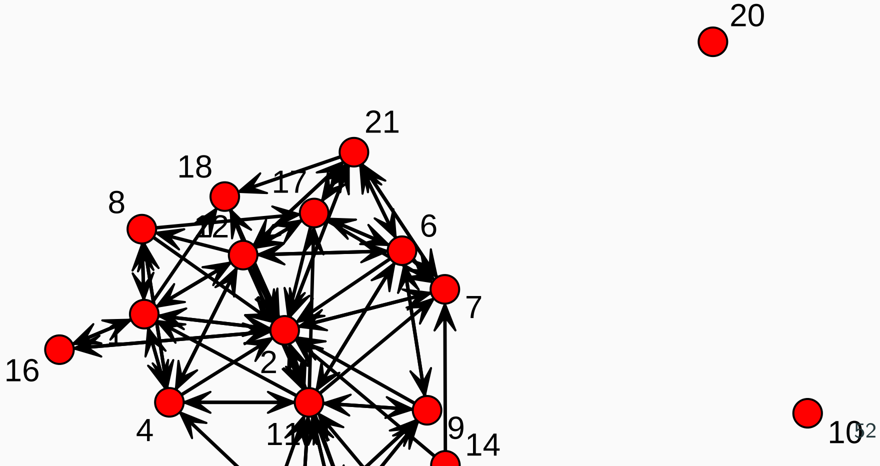


Designs: CSS

```
gplot(apply(kfr.post.pooled$net, c(2, 3), median), displayl
    boxed.lab = FALSE)
```



```
gplot(apply(kfr.post.actor$net, c(2, 3), median), displayla
    boxed.lab = FALSE)
```



Designs: Coding Schemes as "Instruments"

- Can also think of coding schemes for archival materials as "instruments"
- Transcripts
    - Tag each line by sender/receiver - (i, j) tie if i sends to j
- Descriptive lists/tables
    - Common for two-mode data
    - Build entity/property table; fill in (i, j) as 1 if ith row entity has property j
- Video/Audio
    - Determine criterion for interaction
    - Find all interactions, code by sender/receiver
    - (i, j) tie if i sends to j

Designs: Coding Schemes as "Instruments"

- Narrative documents
  - Determine criterion for interactions
  - As before, code by sender/receiver (or just by dyad, if not directed)
  - (i, j) tie if i sends to j, or $\{i, j\}$ tie if i and j interact

Designs: Coding Schemes as “Instruments”

`<img height='100' src='assets/img/gameofthrones.png' />`

Andrew Beveridge and Jie Shan’s Network of Thrones

References and Places for More Information

References and Places for More Information i

Network Vizualization: Theory and Methods

Network Vizualization: R

Network Data Collection

References and Places for More Information