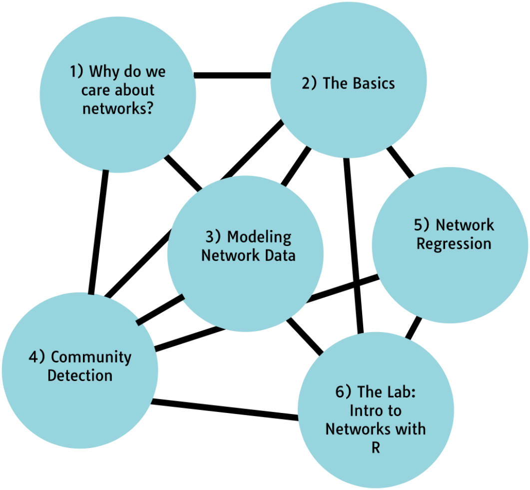
Introduction to Network Analysis

Heather Mathews

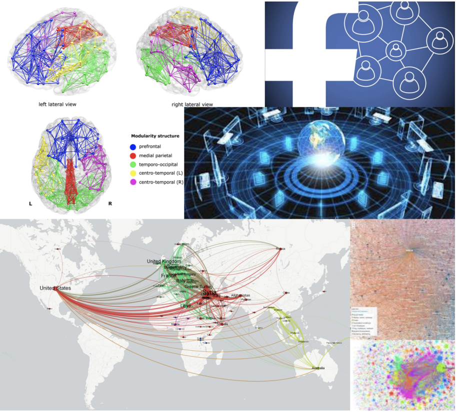
Duke University

November 7, 2019

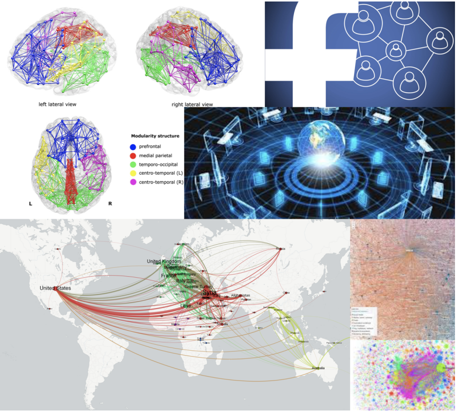
Outline


1) Why do we care about networks?
2) The Basics
3) Modeling Network Data
4) Community Detection
5) Network Regression
6) The Lab: Intro to Networks with R

Who uses networks?

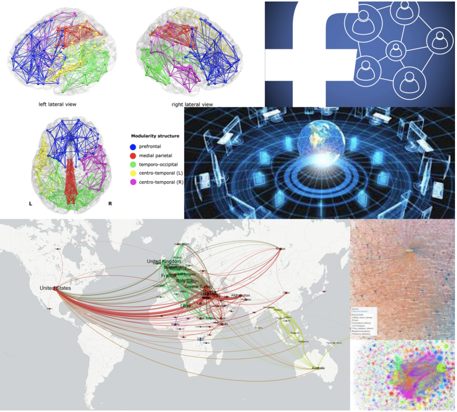


Who uses networks?



- Almost everyone! From economists to sociologists to biologists

Who uses networks?



- Almost everyone! From economists to sociologists to biologists
- Brain networks, social networks, computer networks, traffic networks, trade networks...

Motivating Example: National Longitudinal Study of Adolescent to Adult Health (AddHealth)

Motivating Example: National Longitudinal Study of Adolescent to Adult Health (AddHealth)

- Study was conducted due to a Congressional mandate to study factors influencing health behaviors of adolescents

Motivating Example: National Longitudinal Study of Adolescent to Adult Health (AddHealth)

- Study was conducted due to a Congressional mandate to study factors influencing health behaviors of adolescents
- Collected data on approximately 20,000 high school students across the United States during the 1994-1995 school year

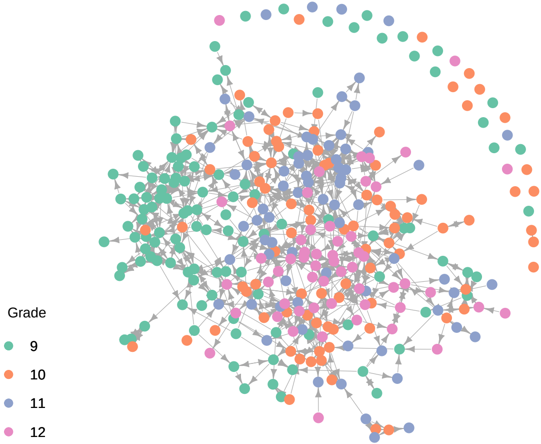
Motivating Example: National Longitudinal Study of Adolescent to Adult Health (AddHealth)

- Study was conducted due to a Congressional mandate to study factors influencing health behaviors of adolescents
- Collected data on approximately 20,000 high school students across the United States during the 1994-1995 school year
- Network Data: In each high school, each student was asked to nominate their top 5 male and female friends

Motivating Example: National Longitudinal Study of Adolescent to Adult Health (AddHealth)

- Study was conducted due to a Congressional mandate to study factors influencing health behaviors of adolescents
- Collected data on approximately 20,000 high school students across the United States during the 1994-1995 school year
- Network Data: In each high school, each student was asked to nominate their top 5 male and female friends
- Covariate information on students was collected including grade, smoking status, drinking habits, club involvement, GPA, and sport involvement

AddHealth: Visualizing our Network

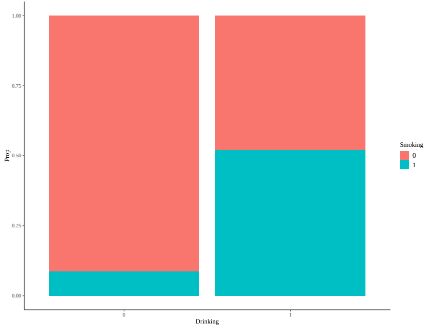


AddHealth

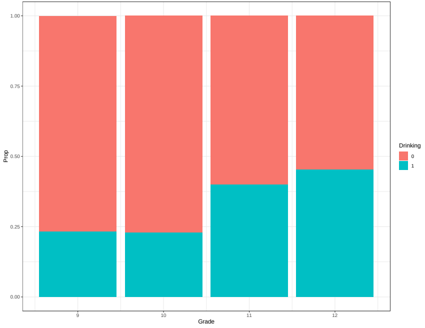
- Fairly uniform grade distribution
- Approximately 84% of male students are white
- About 22% of students are smokers, 30% drink alcohol
- On average, 4 individuals live in a household

AddHealth

- Fairly uniform grade distribution
- Approximately 84% of male students are white
- About 22% of students are smokers, 30% drink alcohol
- On average, 4 individuals live in a household



Smoking and drinking



Drinking by grade

Goals of AddHealth

- Investigate which covariates might influence the probability of a friendship

Goals of AddHealth

- Investigate which covariates might influence the probability of a friendship
- Identify possible clustering within our network. Do latent communities exist?

Goals of AddHealth

- Investigate which covariates might influence the probability of a friendship
- Identify possible clustering within our network. Do latent communities exist?
- Connect covariate estimation and latent communities

Goals of AddHealth

- Investigate which covariates might influence the probability of a friendship
- Identify possible clustering within our network. Do latent communities exist?
- Connect covariate estimation and latent communities
  - Higher GPA might increase popularity within the 'nerdy' clique but decrease popularity within the 'popular' clique

and I got that

red lip

CLASSIC

and I got that
red lip
CLASSIC

Karate Club Data

Network Scientists with Karate Trophies

5 MONTHS AGO
#NETWORKSCIENCE
#KARATECLUB
#TROPHY

The first scientist at any conference on networks who uses Zachary's karate club as an example is inducted into the Zachary Karate Club Club, and awarded a prize. This tumblr records those moments.

RSS
ARCHIVE

Motivating Example: Karate Club

- Over three years (1970-1972), Zachary studied 34 individuals who once belonged to one karate club

Motivating Example: Karate Club

- Over three years (1970-1972), Zachary studied 34 individuals who once belonged to one karate club
- During that time, there was a conflict between John A. (adminstrator) and Mr. Hi (instructor) that resulted in the one club splitting into 2

Motivating Example: Karate Club

- Over three years (1970-1972), Zachary studied 34 individuals who once belonged to one karate club
- During that time, there was a conflict between John A. (adminstrator) and Mr. Hi (instructor) that resulted in the one club splitting into 2
- Interactions outside of the karate clubs were observed by Zachary

Motivating Example: Karate Club

- Over three years (1970-1972), Zachary studied 34 individuals who once belonged to one karate club
- During that time, there was a conflict between John A. (adminstrator) and Mr. Hi (instructor) that resulted in the one club splitting into 2
- Interactions outside of the karate clubs were observed by Zachary
- He was able to classify 33/34 of the members into either John's or Mr. Hi's club

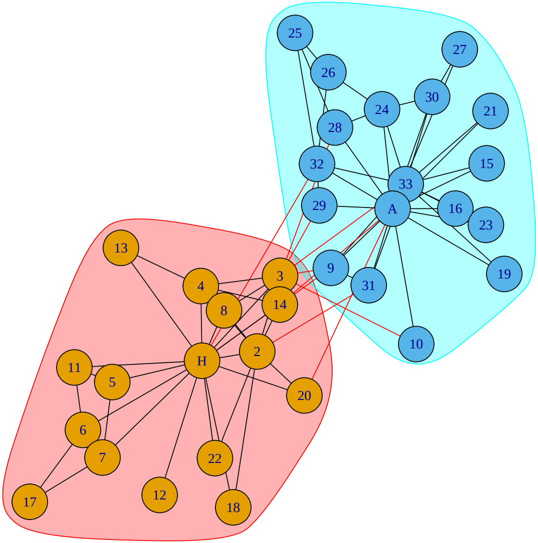
Motivating Example: Karate Club

- Over three years (1970-1972), Zachary studied 34 individuals who once belonged to one karate club
- During that time, there was a conflict between John A. (adminstrator) and Mr. Hi (instructor) that resulted in the one club splitting into 2
- Interactions outside of the karate clubs were observed by Zachary
- He was able to classify 33/34 of the members into either John's or Mr. Hi's club
- Goal: Investigating possible fission in small community setting and interested in how information flows between the 2 clubs

Motivating Example: Karate Club

- Over three years (1970-1972), Zachary studied 34 individuals who once belonged to one karate club
- During that time, there was a conflict between John A. (adminstrator) and Mr. Hi (instructor) that resulted in the one club splitting into 2
- Interactions outside of the karate clubs were observed by Zachary
- He was able to classify 33/34 of the members into either John's or Mr. Hi's club
- Goal: Investigating possible fission in small community setting and interested in how information flows between the 2 clubs
- This gives us ground truth for communities! Which has made it very popular for testing community detection algorithms

Motivating Example: Karate Club



Motivating Example: Sports Analytics

Motivating Example: Sports Analytics

What is a network?

- A network consists of **relational data** (data that describes relationships between actors)

What is a network?

- A network consists of **relational data** (data that describes relationships between actors)
- **Sociomatrix**: $n \times n$ matrix, A, to represent relationships between nodes. If binary entries, this is called an **adjacency matrix**

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

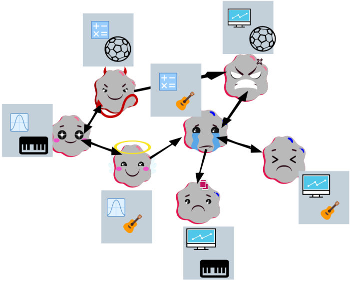
Adjacency matrix, A

What is a network?

- A network consists of **relational data** (data that describes relationships between actors)
- **Sociomatrix**: $n \times n$ matrix, A, to represent relationships between nodes. If binary entries, this is called an **adjacency matrix**

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Adjacency matrix, A



Graph of A

Formally Defining a Network

- **Graph**: $G = (V, E)$ with $V = \{1, ..., n\}$ and $\mathcal{E} = \{(i, j) \; \forall 1 \leq j \neq i \leq n\}$. $E \subset \mathcal{E}$ (if $E = \mathcal{E}$, fully connected graph)
  - Number of nodes: $n = |V|$
  - Number of edges: $m = |E|$

From	To
1	2
1	3
2	1
2	3
3	1

Example Edgelist

Directed vs Undirected

- Sometime edges between individuals are reciprocated (that is, if i is friends with j, j is friends with i)
- If all edges in our graph are reciprocated, then we have an **undirected network**

Directed vs Undirected

- Sometime edges between individuals are reciprocated (that is, if i is friends with j, j is friends with i)
- If all edges in our graph are reciprocated, then we have an **undirected network**

Directed vs Undirected

- Sometime edges between individuals are reciprocated (that is, if i is friends with j, j is friends with i)
- If all edges in our graph are reciprocated, then we have an **undirected network**

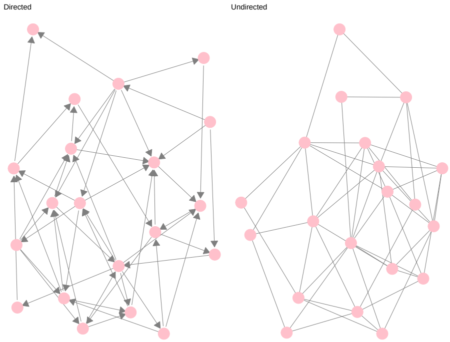
- However, if i is connected to j and j does not reciprocate that connection, then we have a directed edge from i to j. This leads to a **directed network**

Directed vs Undirected

- Sometime edges between individuals are reciprocated (that is, if i is friends with j, j is friends with i)
- If all edges in our graph are reciprocated, then we have an **undirected network**

- However, if i is connected to j and j does not reciprocate that connection, then we have a directed edge from i to j. This leads to a **directed network**

Directed vs. Undirected



Directed: Asymmetric adjacency matrix, Undirected: Symmetric adjacency matrix

Directed vs. Undirected

Degree: Number of nodes a person is connected to

Density: Proportion of edges in graph over maximum possible number of edges

Directed vs. Undirected

Degree: Number of nodes a person is connected to

Density: Proportion of edges in graph over maximum possible number of edges

	Directed	**Undirected**
Max Possible # of Edges	$n^2 - n = n(n-1)$	$n(n-1)/2$
Degree	$d_i^{out} = \sum_{j:i \neq j} A_{i,j}$ (out) $d_i^{in} = \sum_{j:i \neq j} A_{j,i}$ (in)	$d_i = \sum_{j:i \neq j} A_{i,j}$
Density	$m/(n(n-1))$	$2m/(n(n-1))$

Directed vs. Undirected

Degree: Number of nodes a person is connected to

Density: Proportion of edges in graph over maximum possible number of edges

	Directed	**Undirected**
Max Possible # of Edges	$n^2 - n = n(n-1)$	$n(n-1)/2$
Degree	$d_i^{out} = \sum_{j:i \neq j} A_{i,j}$ (out) $d_i^{in} = \sum_{j:i \neq j} A_{j,i}$ (in)	$d_i = \sum_{j:i \neq j} A_{i,j}$
Density	$m/(n(n-1))$	$2m/(n(n-1))$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

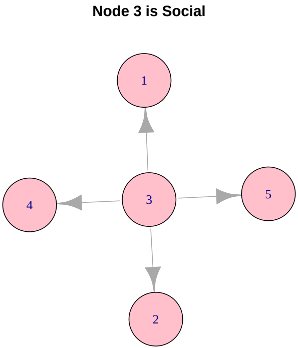
Directed matrix, A

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

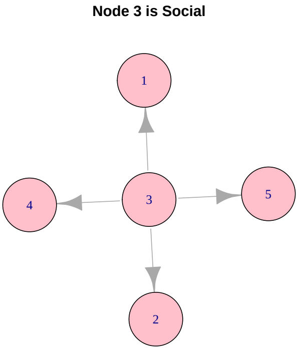
Undirected matrix, A

Sociability vs Popularity

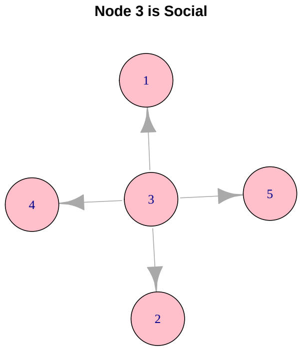
Sociability vs Popularity

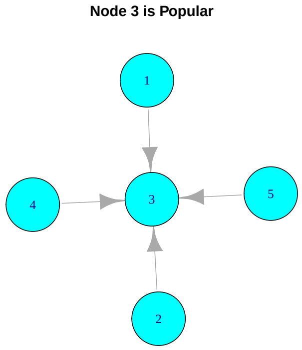


Sociability vs Popularity



Sociability vs Popularity





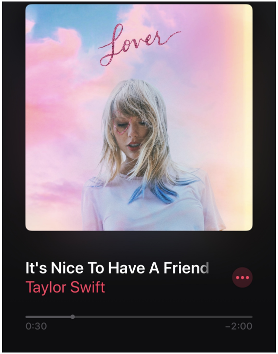
Some Relevant Definitions

Some Relevant Definitions

- **Reciprocity**: If person 1 is connected to person 2, person 2 is connected to person 1

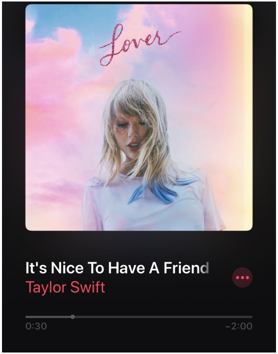
Some Relevant Definitions

- **Reciprocity**: If person 1 is connected to person 2, person 2 is connected to person 1



Some Relevant Definitions

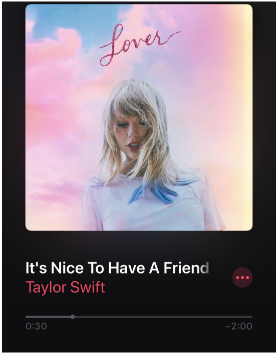
- **Reciprocity**: If person 1 is connected to person 2, person 2 is connected to person 1



- **Homophily**: More likely to connect with people who are similar to you

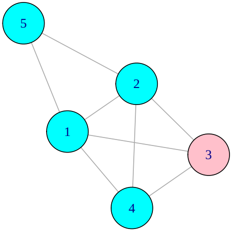
Some Relevant Definitions

- **Reciprocity**: If person 1 is connected to person 2, person 2 is connected to person 1

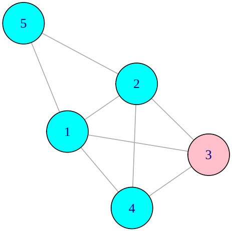


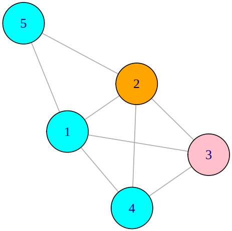
- **Homophily**: More likely to connect with people who are similar to you
- **Transitivity**: Friends of friends have higher probability of being friends

Sorry to say, but your friends have more friends than you...

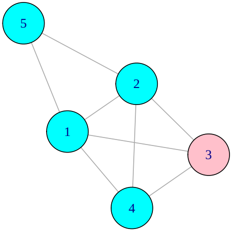


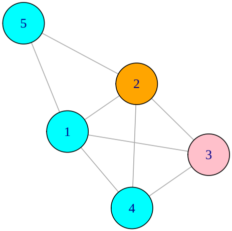
Sorry to say, but your friends have more friends than you...





Sorry to say, but your friends have more friends than you...





	1	2	3	4	5
Degree	4.00	4.00	3.00	3.00	2.00
Avg. Deg. of Friends	3.00	3.00	3.67	3.67	4.00

Notions of Centrality

Notions of Centrality

- **Betweenness Centrality:** Looks at how many times a node is part of the shortest path between other nodes

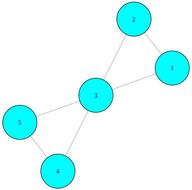
$$g_{j,k} = \# \text{ of shortest paths to get from } j \text{ to } k$$
$$g_{j,k}(i) = \# \text{ of shortest paths from } j \text{ to } k \text{ that go through i}$$
$$c_i = \textstyle\sum_{j<k} g_{j,k}(i)/g_{j,k}$$

Notions of Centrality

- **Betweenness Centrality:** Looks at how many times a node is part of the shortest path between other nodes

$$g_{j,k} = \# \text{ of shortest paths to get from } j \text{ to } k$$
$$g_{j,k}(i) = \# \text{ of shortest paths from } j \text{ to } k \text{ that go through i}$$
$$c_i = \sum_{j<k} g_{j,k}(i)/g_{j,k}$$

  - This is useful for finding indiviudals that are like bridges (flow of info)



Betweenness Centrality: 0 0 4 0 0

Notions of Centrality

- **Closeness Centrality:** Measures how close one node is to all other nodes in the network.Define $d_{i,j}$ as the minimum path length from i to j.

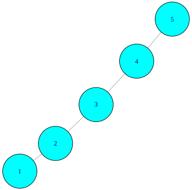
$$c_i = \frac{1}{\sum_{j:j\neq i} d_{i,j}}$$

Notions of Centrality

- **Closeness Centrality:** Measures how close one node is to all other nodes in the network.Define $d_{i,j}$ as the minimum path length from i to j.

$$c_i = \frac{1}{\sum_{j:j \neq i} d_{i,j}}$$

  - Sum up the shortest paths between all nodes (good for looking at who influences spread of info)



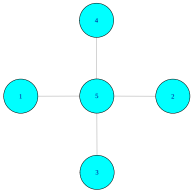
Closeness Centrality: 0.40 0.57 0.67 0.57 0.40

Notions of Centrality

- **Degree Centrality:** Importance based on number of connections a node has

$$c_i = \sum_{j:j \neq i} A_{i,j}$$

  - Useful for revealing direct connections and locating popular nodes



Degree Centrality 1 1 1 1 4

Notions of Centrality

Notions of Centrality

- **EigenCentrality:** Centrality of each node is proportional to the sum of its neighbor's centralities

$$c_i = \frac{1}{\lambda} \sum_{j:j \neq i} A_{i,j} c_j$$

Notions of Centrality

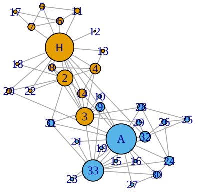
- **EigenCentrality:** Centrality of each node is proportional to the sum of its neighbor's centralities

$$c_i = \frac{1}{\lambda} \sum_{j:j \neq i} A_{i,j} c_j$$

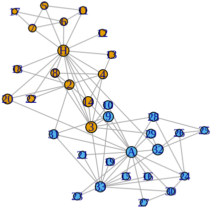
  - λ corresponds to the greatest eigenvalue of A and c corresponds to the top eigenvector. The i^{th} component of c gives the relative centrality score of vertex i
  - Central nodes are connected to other central nodes (very similar to degree centrality)
  - Basis for Google's PageRank
  - For graph on previous slide, eigen centrality is: 0.5 0.5 0.5 0.5 1.0

Example: Karate Club

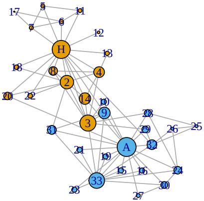
Degree



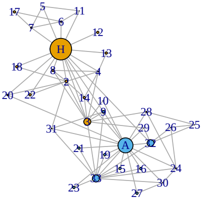
Closeness



Eigen



Betweenness



Now off to the lab...

Now off to the lab...

Modeling and Generating Networks

- Erdos-Renyi Graph (ER)
- Exponential Random Graph Models (ERGMS)
- Stochastic Block Models (SBM)
- Additive and Multiplicative Effects Network (AMEN)

Erdos-Renyi Graph

Generating graphs:

- **G(n,p)**: ER graph where edges are drawn independently with probability, p

$$P(G_0) = P(G = G_0) = p^m(1-p)^{(N-m)}$$
$$E(\# \text{ of Edges}) = \binom{n}{2}p$$

Relating ER to the Adj. Matrix

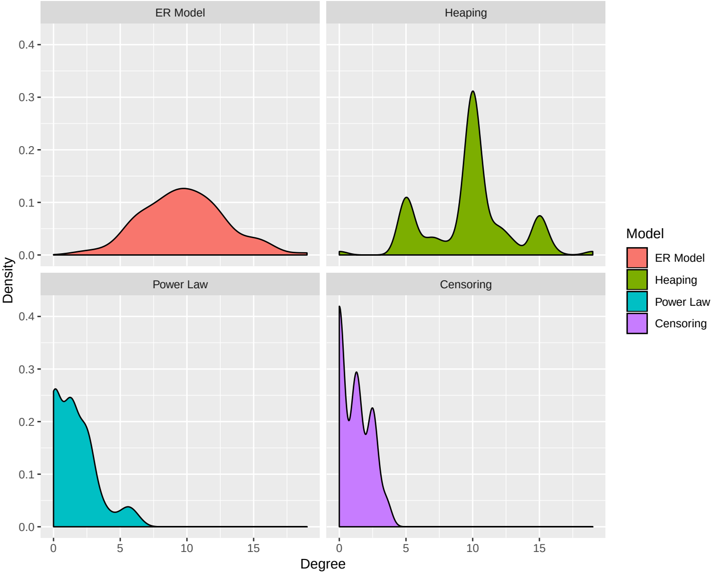
How do the models on the previous slide relate to the adjacency matrix, A?

Relating ER to the Adj. Matrix

How do the models on the previous slide relate to the adjacency matrix, A?

$$A_{i,j} | p \sim Bern(p)$$

Degree Distribution... and why we care



Degree Distribution

- What is the degree distribution of an ER model?

Degree Distribution

- What is the degree distribution of an ER model?
- ER(p) has (n-1) possible friends

Degree Distribution

- What is the degree distribution of an ER model?
- ER(p) has (n-1) possible friends
- Let person i have k friends, thus $\binom{n-1}{k}$ possibilities and probability of a friend is p:

$$\binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Degree Distribution

- What is the degree distribution of an ER model?
- ER(p) has (n-1) possible friends
- Let person i have k friends, thus $\binom{n-1}{k}$ possibilities and probability of a friend is p:
$$\binom{n-1}{k} p^k (1-p)^{n-1-k}$$
- Bin(n-1,p)

Back to the lab...

Back to the lab...

Exponential Random Graph Models: ERGM

Basic generative model for networks that is based off of sufficient statistics

$$P_\theta(G_n) = exp(\theta^T t_n(G_n) - \Psi_n(\theta))$$

where

- θ: Parameter (or vector of parameters) we want to estimate
- t_n: Sufficient statistic
- Ψ_n: Normalizing constant

Examples of ERGM

The ER Model:

- In this model, each edge is sampled iid Bernoulli with some probability p. For this model, $t_n = \sum_{i<j} A_{i,j}$

$$P_\theta(A_n) \propto exp\{\theta \sum_{i<j} A_{ij}\}$$

ER as an ERGM

We have seen the likelihood for the $ER(p)$:

$$\begin{aligned} p^m(1-p)^{N-m} &= exp\{log(p^m(1-p)^{N-m})\} \\ &= exp\{m \times log(p) + (N-m) \times log(1-p)\} \\ &= exp\{m \times log(\frac{p}{1-p}) + N \times log(1-p)\} \end{aligned}$$

ER as an ERGM

We have seen the likelihood for the $ER(p)$:

$$\begin{aligned} p^m(1-p)^{N-m} &= exp\{log(p^m(1-p)^{N-m})\} \\ &= exp\{m \times log(p) + (N-m) \times log(1-p)\} \\ &= exp\{m \times log(\frac{p}{1-p}) + N \times log(1-p)\} \end{aligned}$$

Recall the ERGM

ER as an ERGM

We have seen the likelihood for the $ER(p)$:

$$
\begin{aligned}
p^m(1-p)^{N-m} &= exp\{log(p^m(1-p)^{N-m})\} \\
&= exp\{m \times log(p) + (N-m) \times log(1-p)\} \\
&= exp\{m \times log(\frac{p}{1-p}) + N \times log(1-p)\}
\end{aligned}
$$

Recall the ERGM

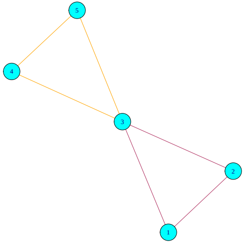
$$P_\theta(A_n) \propto exp(\theta t_n(A_n)) \tag{1}$$

where

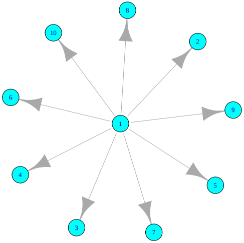
- $\theta = \log(\frac{p}{1-p})$
- $t_n = m = \sum_{i<j} A_{i,j}$

ERGM Cont.

We can also consider some other sufficient statistics to include in our model such as # of stars, # of triangles, # of edges, etc.



Triangles



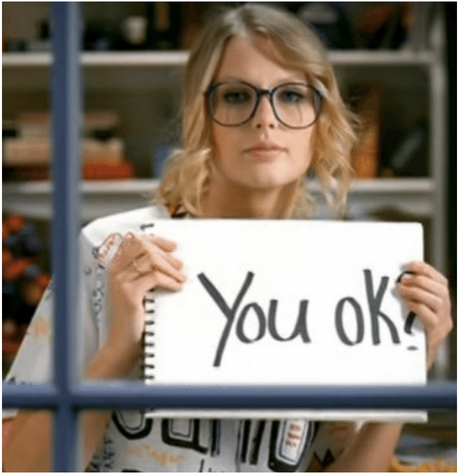
Star

What kinds of questions can ERGMs help answer?

- We can look at expectation of degree, edges, degree distribution
- We can test what model fits a new network best (were 2 graphs generated from the same model?)
- HOWEVER, not always consistent estimators :/

Back to the lab...

Back to the lab...

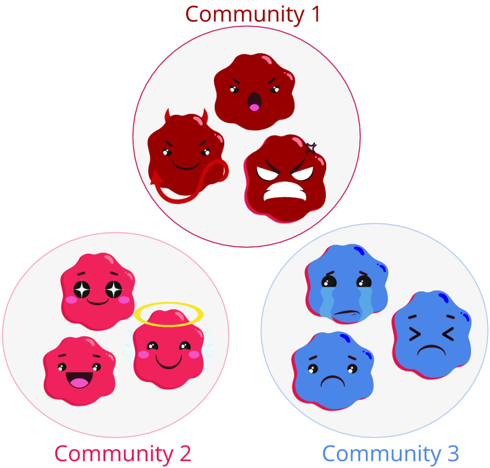


The Stochastic Block Model

Now we consider if nodes come from **communities** (Holland et al, 1983)

The Stochastic Block Model

Now we consider if nodes come from **communities** (Holland et al, 1983)



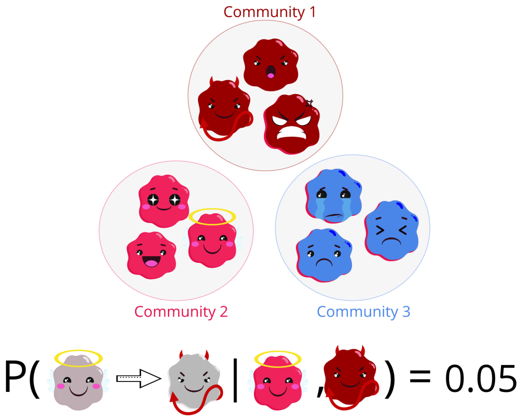
$$\theta = \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Understanding the SBM Visually

$$B = \begin{pmatrix} 0.4 & 0.05 & 0.3 \\ 0.05 & 0.4 & 0.1 \\ 0.3 & 0.1 & 0.4 \end{pmatrix}$$

Understanding the SBM Visually

$$B = \begin{pmatrix} 0.4 & 0.05 & 0.3 \\ 0.05 & 0.4 & 0.1 \\ 0.3 & 0.1 & 0.4 \end{pmatrix}$$



Another Generative Model: The Stochastic Block Model

- $\theta = [\theta_1, \theta_2, .., \theta_K]$ is a vector containing probabilities that a person belongs to a particular community, $k \in \{1, ..., K\}$

Another Generative Model: The Stochastic Block Model

- $\theta = [\theta_1, \theta_2, .., \theta_K]$ is a vector containing probabilities that a person belongs to a particular community, $k \in \{1, ..., K\}$
- $B \in \mathbb{R}^{K \times K}$ is a preference matrix that describes the probability of connection of nodes based soley on an individual's membership

Another Generative Model: The Stochastic Block Model

- $\theta = [\theta_1, \theta_2, .., \theta_K]$ is a vector containing probabilities that a person belongs to a particular community, $k \in \{1, ..., K\}$
- $B \in \mathbb{R}^{K \times K}$ is a preference matrix that describes the probability of connection of nodes based soley on an individual's membership
- Z indicates which community a person belongs to

$$P(Z_i = k) = \theta_k$$
$$A_{i,j} | Z_i, Z_j \sim Bern(B_{z_i,z_j})$$
$$P(A, Z, \theta, B) = \prod_K \theta_k^{\sum 1_{z_i=k}} \prod_{i,j} B_{Z_i,Z_j}^{A_{ij}} (1 - B_{Z_i,Z_j})^{1-A_{i,j}}$$

Balanced Multi-Label Propagation for Overlapping Community Detection in Social Networks

Authors Authors and affiliations

Zhi-Hao Wu, You-Fang Lin, Steve Gregory, Huai-Yu Wan, Sheng-Feng Tian

Mutually Enhancing Community Detection and Sentiment Analysis on Twitter Networks

William Deitrick, Wei Hu*

Bayesian Inference and Testing of Group Differences in Brain Networks

Daniele Durante* and David B. Dunson†

Listen

Listen to this page using ReadSpeaker webReader

Identifying functional urban regions within traffic flow

Ed Manley

Pages 40-42 | Received 30 Jan 2014, Accepted 03 Feb 2014, Published online: 12 Mar 2014

Promoting Small and Medium Enterprises with a Clustering Approach: A Policy Experience from Indonesia

by Tulus Tambunan

Community Detection in General Stochastic Block models: Fundamental Limits and Efficient Algorithms for Recovery

Publisher: IEEE

2 Author(s) Emmanuel Abbe ; Colin Sandon View All Authors

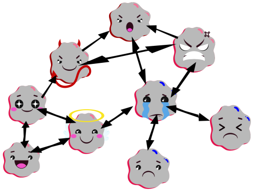
Social selection and peer influence in an online social network

Community Detection

- Maybe we believe our network, A, came from an SBM, and we care about finding community labels

Community Detection

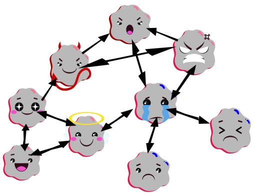
- Maybe we believe our network, A, came from an SBM, and we care about finding community labels



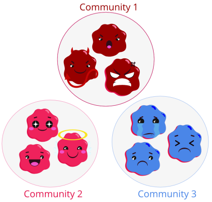
Observed Network

Community Detection

- Maybe we believe our network, A, came from an SBM, and we care about finding community labels



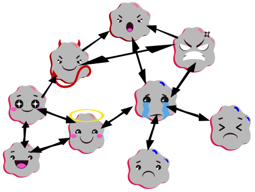
Observed Network



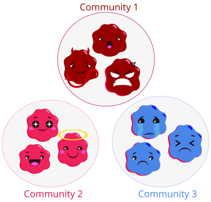
True latent communities

Community Detection

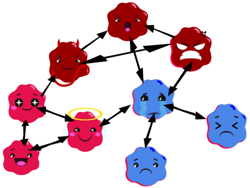
- Maybe we believe our network, A, came from an SBM, and we care about finding community labels



Observed Network



True latent communities



Goal

Methods for Community Detection

- Spectral methods
- Gaussian Mixture Models
- Centrality based approaches

Spectral Clustering

- Consider taking a *spectral decomposition* of A (eigendecomposition, singular value decomposition)

Spectral Clustering

- Consider taking a *spectral decomposition* of A (eigendecomposition, singular value decomposition)
- If we take the eigendecomposition,

$$A = V\Lambda V^T$$

where $\Lambda = Diag(\lambda_1, \ldots, \lambda_n)$ contains the eigenvalues in decreasing order (in magnitude), and $V = (V_1, \ldots, V_n)^T$ is a matrix containing the orthonormal columns corresponding to the eigenvectors

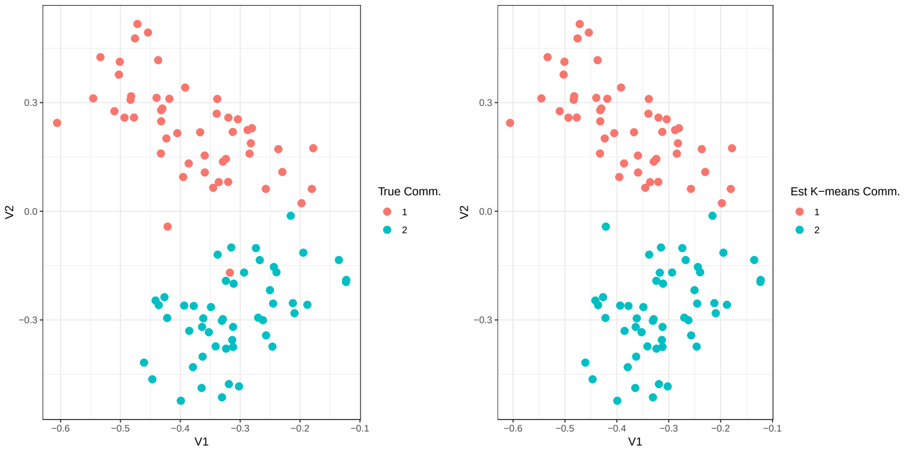
Spectral Clustering

- Consider taking a *spectral decomposition* of A (eigendecomposition, singular value decomposition)
- If we take the eigendecomposition,

$$A = V \Lambda V^T$$

  where $\Lambda = Diag(\lambda_1, \ldots, \lambda_n)$ contains the eigenvalues in decreasing order (in magnitude), and $V = (V_1, \ldots, V_n)^T$ is a matrix containing the orthonormal columns corresponding to the eigenvectors
- If community structure exists, it should presumably show up in a lower dimensional representation of A
- Consider taking the top K eigenvectors of A where K is the number of hypothesized communities that exist

Looking at our eigenvectors



The Ultimate Question: How do we pick the number of communities?

- The answer to this is not super clear

The Ultimate Question: How do we pick the number of communities?

- The answer to this is not super clear
- Elbow plots, look for eigen values that escape the bulk

The Ultimate Question: How do we pick the number of communities?

- The answer to this is not super clear
- Elbow plots, look for eigen values that escape the bulk
- Prior information

The Ultimate Question: How do we pick the number of communities?

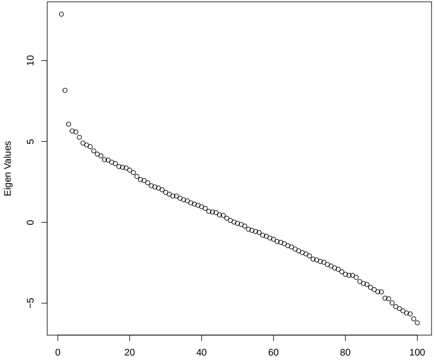
- The answer to this is not super clear
- Elbow plots, look for eigen values that escape the bulk
- Prior information
- Try a few different options

The Ultimate Question: How do we pick the number of communities?

- The answer to this is not super clear
- Elbow plots, look for eigen values that escape the bulk
- Prior information
- Try a few different options

The Ultimate Question: How do we pick the number of communities?

- The answer to this is not super clear
- Elbow plots, look for eigen values that escape the bulk
- Prior information
- Try a few different options

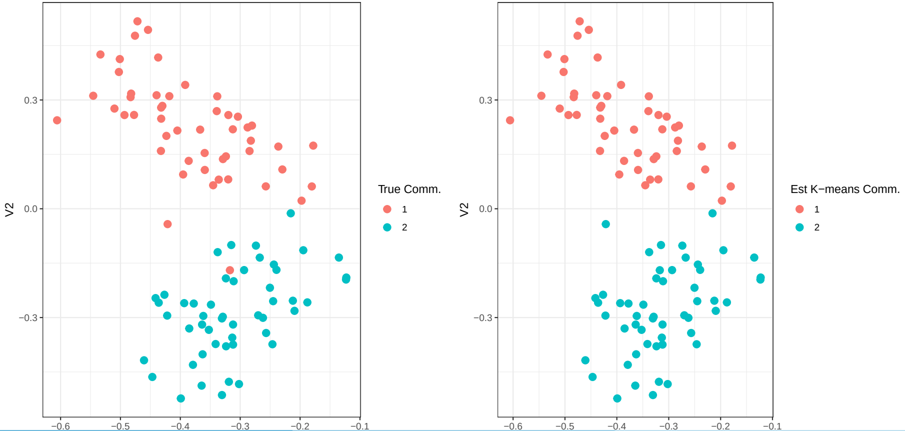


Once we have our number of communities... [Rohe et al., 2012]

- Run a clustering algorithm on the top K eigenvectors corresponding to the top K eigenvalues (in magnitude), $V_{,1:K}$
- Can also cluster on $V_{,1:K}\Lambda^{1/2}_{1:K,1:K}$

Once we have our number of communities... [Rohe et al., 2012]

- Run a clustering algorithm on the top K eigenvectors corresponding to the top K eigenvalues (in magnitude), $V_{,1:K}$
- Can also cluster on $V_{,1:K}\Lambda^{1/2}_{1:K,1:K}$



Gaussian Mixture Models

Gaussian Mixture Models

- Consider the latent positions that we can obtain from spectral decompositions of A

Gaussian Mixture Models

- Consider the latent positions that we can obtain from spectral decompositions of A
- It can actually be shown that these latent positions can be modeled using Gaussian Mixture Models:

Gaussian Mixture Models

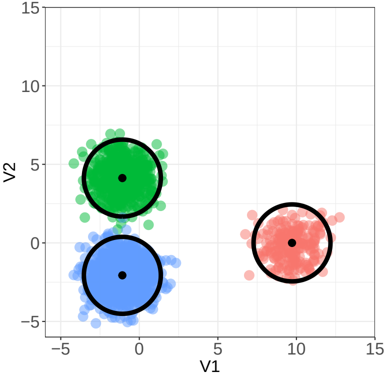
- Consider the latent positions that we can obtain from spectral decompositions of A
- It can actually be shown that these latent positions can be modeled using Gaussian Mixture Models:

$$P(V_i = v_i) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k N(v_i; \mu_k, \sigma_k^2)$$

Gaussian Mixture Models

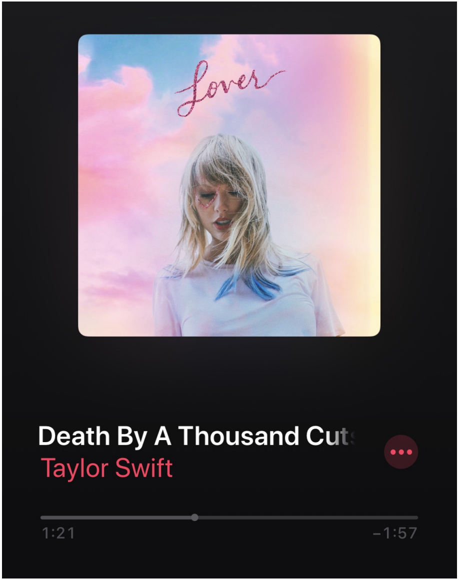
- Consider the latent positions that we can obtain from spectral decompositions of A
- It can actually be shown that these latent positions can be modeled using Gaussian Mixture Models:

$$P(V_i = v_i) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k N(v_i; \mu_k, \sigma_k^2)$$



Moving to Centrality Based Example...

Moving to Centrality Based Example...

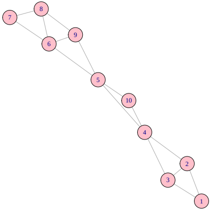


Edge Betweenness Algorithm

- Basic Idea:
  - Find the edge with maximum edge betweenness centrality and delete it

Edge Betweenness Algorithm

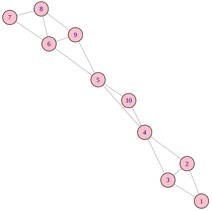
- Basic Idea:
  - Find the edge with maximum edge betweenness centrality and delete it



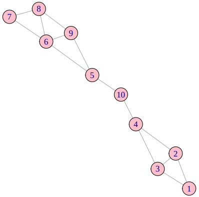
Step 1

Edge Betweenness Algorithm

- Basic Idea:
  - Find the edge with maximum edge betweenness centrality and delete it



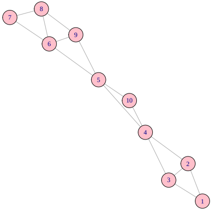
Step 1



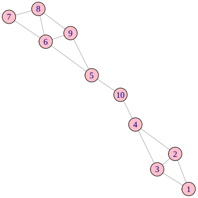
Step 2

Edge Betweenness Algorithm

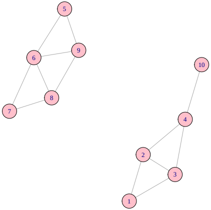
- Basic Idea:
  - Find the edge with maximum edge betweenness centrality and delete it



Step 1



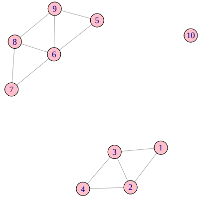
Step 2



Step 3

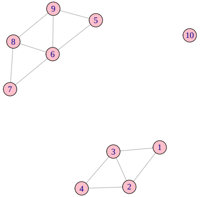
Edge Betweenness Algorithm

Edge Betweenness Algorithm

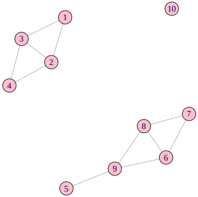


Step 4

Edge Betweenness Algorithm



Step 4



Step 5

Back to the lab...

Back to the lab...

Moving to Regression: Connecting to things we know

Moving to Regression: Connecting to things we know

- Consider a typical probit regression where we observe a binary response and covariates

Moving to Regression: Connecting to things we know

- Consider a typical probit regression where we observe a binary response and covariates
- We want to estimate β so we consider probit regression where ϵ are assumed to be iid

$$A = \mathbb{1}_{Z>0}$$
$$Z = X\beta + \epsilon$$

Connecting to things we know

A is a type of response variable, but it is a matrix. To get something more familiar, we could vectorize A such that:

Connecting to things we know

A is a type of response variable, but it is a matrix. To get something more familiar, we could vectorize A such that:

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \rightarrow vec(A) = \begin{pmatrix} A_{,1} \\ A_{,2} \\ A_{,3} \\ A_{,4} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \\ 0 \end{pmatrix}$$

Connecting to things we know

A is a type of response variable, but it is a matrix. To get something more familiar, we could vectorize A such that:

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \rightarrow vec(A) = \begin{pmatrix} A_{,1} \\ A_{,2} \\ A_{,3} \\ A_{,4} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \\ 0 \end{pmatrix}$$

Remember that dyadic dependence...

Remember that dyadic dependence...

However, with a network, it is not reasonable to assume that all of these entries are independent

Social Relations Model [Warner et al., 1979]

Social Relations Model [Warner et al., 1979]

- We are interested in modeling the variability in A

Social Relations Model [Warner et al., 1979]

- We are interested in modeling the variability in A

$$
\begin{aligned}
a_{i,j} &= \mathbb{1}_{z_{i,j}>0} \\
z_{i,j} &= \beta_0 + c_i + d_j + \epsilon_{i,j}
\end{aligned}
$$

Social Relations Model [Warner et al., 1979]

- We are interested in modeling the variability in A

$$
\begin{aligned}
a_{i,j} &= \mathbb{1}_{z_{i,j}>0} \\
z_{i,j} &= \beta_0 + c_i + d_j + \epsilon_{i,j}
\end{aligned}
$$

- β_0: Overall global mean
- c: Individual row (sociability/sender behavior) random effects
- d: Individual column (popularity/reciever behavior) random effects

$$
\begin{pmatrix} c_i \\ d_i \end{pmatrix} \overset{i.i.d}{\sim} N(0, \Sigma_{cd}) \text{ where } \Sigma_{cd} = \begin{bmatrix} \sigma_c^2 & \sigma_{cd} \\ \sigma_{cd} & \sigma_d^2 \end{bmatrix}
$$

$$
\begin{pmatrix} \epsilon_{i,j} \\ \epsilon_{j,i} \end{pmatrix} \overset{i.i.d}{\sim} N(0, \Sigma_e) \text{ where } \Sigma_e = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}
$$

Adding in covariate information

Adding in covariate information

$$a_{i,j} = \mathbb{1}_{z_{i,j}>0}$$

$$z_{i,j} =$$

$$\beta_0 + \sum_{p=1}^{P_r}(x_{r,p,i}\beta_{r,p}) + \sum_{p=1}^{P_c}(x_{c,p,j}\beta_{c,p}) + \sum_{p=1}^{P_d}(x_{d,p,i,j}\beta_{d,p}) + c_i + d_j + \epsilon_{i,j}$$

- X_r: Observed covariate information for row covariates
- X_c: Observed covariate information for column covariates
- X_d: Observed covariate information for dyadic covariates
- β: Coefficients of interest estimating covariate effects on connections

Putting this into matrix form

$$\begin{bmatrix} z_{1,1} & z_{1,2} & z_{1,3} \\ z_{2,1} & z_{2,2} & z_{2,3} \\ z_{3,1} & z_{3,2} & z_{3,3} \end{bmatrix} = \begin{bmatrix} x_{r,1} & x_{r,1} & x_{r,1} \\ x_{r,2} & x_{r,2} & x_{r,2} \\ x_{r,3} & x_{r,3} & x_{r,3} \end{bmatrix} \beta_r + \begin{bmatrix} x_{c,1} & x_{c,2} & x_{c,3} \\ x_{c,1} & x_{c,2} & x_{c,3} \\ x_{c,1} & x_{c,2} & x_{c,3} \end{bmatrix} \beta_c + \ldots$$

Better Models...

Better Models...

Adding in multiplicative effects (AMEN [Hoff, 2018])

- Sometimes there are higher order latent dependencies between nodes
  - **Example:** Nodes may be *homophilous*. Individuals who are similar to one another are more likely to be friends (clustering)

Adding in multiplicative effects (AMEN [Hoff, 2018])

- Sometimes there are higher order latent dependencies between nodes
  - **Example:** Nodes may be *homophilous*. Individuals who are similar to one another are more likely to be friends (clustering)

$$a_{i,j} = \mathbb{1}_{z_{i,j}>0}$$
$$z_{i,j} = \beta_0 + \sum_{p=1}^{P_r}(x_{r,p,i}\beta_{r,p}) + \sum_{p=1}^{P_c}(x_{c,p,j}\beta_{c,p}) + \sum_{p=1}^{P_d}(x_{d,p,i,j}\beta_{d,p}) + c_i + d_j + u_i^T v_j + \epsilon_{i,j}$$

- U, V: Latent factor matrices of rank R ($n \times R$ matrices)

Adding in multiplicative effects (AMEN [Hoff, 2018])

- Sometimes there are higher order latent dependencies between nodes
  - **Example:** Nodes may be *homophilous*. Individuals who are similar to one another are more likely to be friends (clustering)

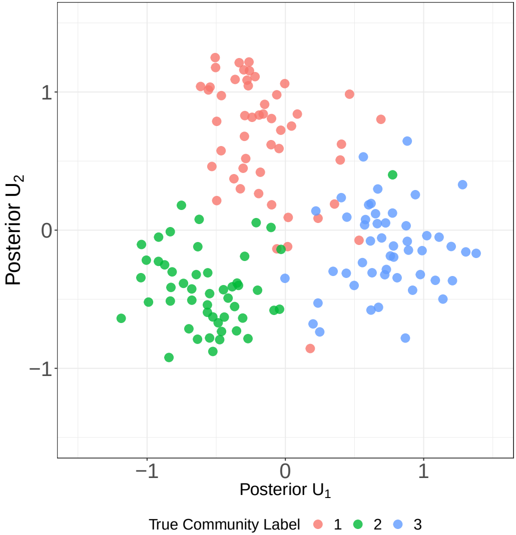
$$a_{i,j} = \mathbb{1}_{z_{i,j}>0}$$

$$z_{i,j} = \beta_0 + \sum_{p=1}^{P_r}(x_{r,p,i}\beta_{r,p}) + \sum_{p=1}^{P_c}(x_{c,p,j}\beta_{c,p}) + \sum_{p=1}^{P_d}(x_{d,p,i,j}\beta_{d,p}) + c_i + d_j + u_i^T v_j + \epsilon_{i,j}$$

- U, V: Latent factor matrices of rank R ($n \times R$ matrices)
- u_i gives us information about a node as a sender
- v_j gives us information about a node as a reciever
- They can describe notions of stochastic equivalence (if u_i is similar to u_j, then they may share similar behaviors)

But what if we have latent community structure?

- Latent multiplicative effects can capture latent community structure



How do we use this model?

- The standard AMEN model is implemented using a Markov Chain Monte Carlo (MCMC) algorithm
- For the standard model, this is provided in the ‘amen’ R package


BRACE YOURSELVES
BAYESIAN ANALYSIS IS COMING

Brief overview of Bayesian Methods

- In frequentist methods, we have a parameter that we want to estimate, θ, that is considered to be fixed, but unknown

Brief overview of Bayesian Methods

- In frequentist methods, we have a parameter that we want to estimate, θ, that is considered to be fixed, but unknown
- In Bayesian methods, we are still interested in estimating θ, however we believe it is an unknown, **random** quantity

Brief overview of Bayesian Methods

- In frequentist methods, we have a parameter that we want to estimate, θ, that is considered to be fixed, but unknown
- In Bayesian methods, we are still interested in estimating θ, however we believe it is an unknown, **random** quantity
- Rather than come up with a point estimate, we want to derive a posterior distribution for θ. What does that mean exactly?

Bayesian Methods

Bayesian Methods

- We observe a dataset, y, which comes from a sample space, $\mathcal{Y}$ where $\mathcal{Y}$ represents all possible datasets that y could come from

Bayesian Methods

- We observe a dataset, y, which comes from a sample space, $\mathcal{Y}$ where $\mathcal{Y}$ represents all possible datasets that y could come from
- We are interested in estimating a population parameter, θ, that comes from Θ (parameter space of θ)

Bayesian Methods

- We observe a dataset, y, which comes from a sample space, $\mathcal{Y}$ where $\mathcal{Y}$ represents all possible datasets that y could come from
- We are interested in estimating a population parameter, θ, that comes from Θ (parameter space of θ)
- In a Bayesian setting, we put a *prior* distribution on our parameter, $\theta \in \Theta$. This describes our beliefs about the true population parameter ($p(\theta)$)

Bayesian Methods

- We observe a dataset, y, which comes from a sample space, $\mathcal{Y}$ where $\mathcal{Y}$ represents all possible datasets that y could come from
- We are interested in estimating a population parameter, θ, that comes from Θ (parameter space of θ)
- In a Bayesian setting, we put a *prior* distribution on our parameter, $\theta \in \Theta$. This describes our beliefs about the true population parameter ($p(\theta)$)
- We then have a *sampling model*, $p(y|\theta)$ that describes our beliefs about y had we known the true θ

Bayesian Methods

- We observe a dataset, y, which comes from a sample space, $\mathcal{Y}$ where $\mathcal{Y}$ represents all possible datasets that y could come from
- We are interested in estimating a population parameter, θ, that comes from Θ (parameter space of θ)
- In a Bayesian setting, we put a *prior* distribution on our parameter, $\theta \in \Theta$. This describes our beliefs about the true population parameter ($p(\theta)$)
- We then have a *sampling model*, $p(y|\theta)$ that describes our beliefs about y had we known the true θ
- Our goal is then to get $p(\theta|y)$ which describes our beliefs for the true value of θ after observing our data y

Bayes Rule

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_{\tilde{\theta}\in\Theta} p(y|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}}$$

Back to AMEN: How do we implement this model?

Back to AMEN: How do we implement this model?

- We have a model: $Z \sim P(Z|\theta)$, $\theta \in \Theta$

Back to AMEN: How do we implement this model?

- We have a model: $Z \sim P(Z|\theta)$, $\theta \in \Theta$
- When Y is binary,

$$S(A) = \{Z : a_{i,j} > 0 \Rightarrow z_{i,j} > 0, a_{i,j} = 0 \Rightarrow z_{i,j} \leq 0\}$$

Back to AMEN: How do we implement this model?

- We have a model: $Z \sim P(Z|\theta)$, $\theta \in \Theta$
- When Y is binary,

$$S(A) = \{Z : a_{i,j} > 0 \Rightarrow z_{i,j} > 0, a_{i,j} = 0 \Rightarrow z_{i,j} \leq 0\}$$

- Likelihood is:

$$L_B(\theta, Z) = Pr(Z \in S(A)|\theta) = \int_{S(A)} P(Z|\theta) d\mu(Z)$$

Back to AMEN: How do we implement this model?

- We have a model: $Z \sim P(Z|\theta)$, $\theta \in \Theta$
- When Y is binary,

$$S(A) = \{Z : a_{i,j} > 0 \Rightarrow z_{i,j} > 0, a_{i,j} = 0 \Rightarrow z_{i,j} \leq 0\}$$

- Likelihood is:

$$L_B(\theta, Z) = Pr(Z \in S(A)|\theta) = \int_{S(A)} P(Z|\theta) d\mu(Z)$$

- We can then use a Gibbs Sampler with MH to approximate $P(\theta|Z \in S(A))$

Why AMEN? Different likelihoods!

Why AMEN? Different likelihoods!

- Up to this point, we have focused on adjacency matrices/ binary networks

Why AMEN? Different likelihoods!

- Up to this point, we have focused on adjacency matrices/ binary networks
- However, networks can have many different forms

Why AMEN? Different likelihoods!

- Up to this point, we have focused on adjacency matrices/ binary networks
- However, networks can have many different forms
- AMEN allows us to easily consider some of these
  - **Censored Binary**: Still 1 or 0, but not all connections are observed
  - **Ranked**: Connections are ranked by some level of importance
  - **Fixed Rank Nomination**: A node can only have a FIXED number of connections and they are ranked

Fixed Rank Nomination

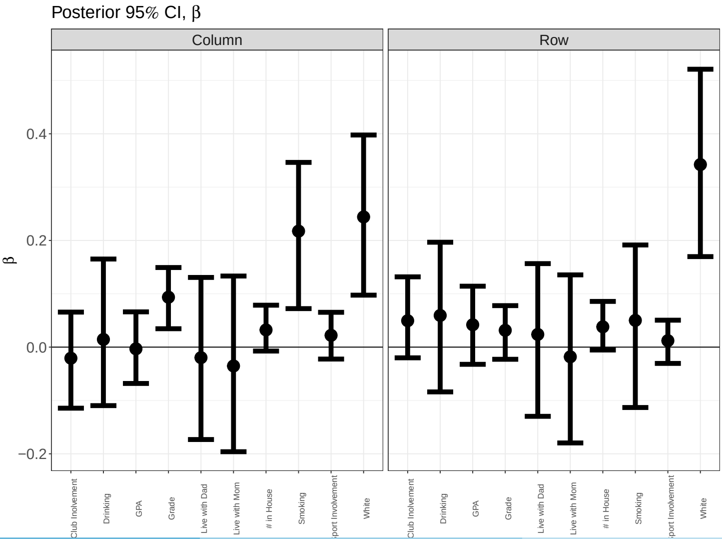
AddHealth

- Each person could rank up to 5 male friends and 5 female friends
- This introduces a *censorship* issue.

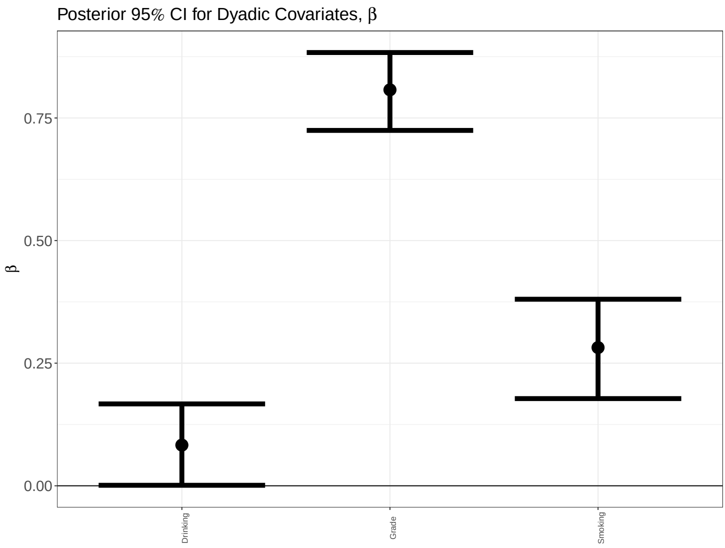
  If person i ranks the 5 people, are they friends with the 6^{th} and just didn't have room to rank them? or do they not like that person?
- If someone ranks less than 5 people, then we assume that they are not friends with person 6
- Another issue... perhaps person i just has never met person 6 but they would be great friends if they had

Back to AddHealth: Regression Results

- 95% CI for β estimates when fitting AMEN model with $R = 2$ on AddHealth network



Standard AMEN R=2



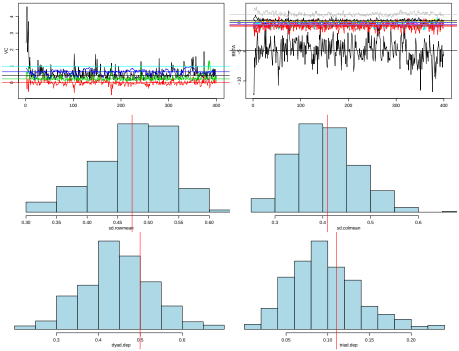
What appears to be important?

- Column: Grade, Smoking, White
- Row: White
- Dyadic: Sharing same drinking behavior, grade, smoking behavior

Diagnostics: Goodness of Fit Statistics

AMEN provides us with some posterior predictive goodness of fit statistics:

1. Empirical standard deviation for row means
2. Empirical standard deviation for column means
3. Empirical within-dyad correlation
4. Normalized measure of triadic dependence



And back to the lab...

And back to the lab...

https://igraph.org/r/doc/igraph.pdf (igraph)
https://cran.r-project.org/web/packages/ergm/ergm.pdf (ergm)
https://arxiv.org/pdf/1506.08237.pdf (amen)

Abbe, E. and Sandon, C. (2015).
Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery.
In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE.

Binkiewicz, N., Vogelstein, J. T., and Rohe, K. (2017).
Covariate-assisted spectral clustering.
Biometrika, 104(2):361–377.

Deitrick, W. and Hu, W. (2013).
Mutually enhancing community detection and sentiment analysis on twitter networks.
Journal of Data Analysis and Information Processing, 1(03):19.

Durante, D., Dunson, D. B., et al. (2018).
Bayesian inference and testing of group differences in brain networks.
Bayesian Analysis, 13(1):29–58.

Feld, S. L. (1991).
Why your friends have more friends than you do.
American Journal of Sociology, 96(6):1464–1477.

Harris, K., Halpern, E., Whitsel, E., Hussey, J., and Udry, J. (2009).
The national longitudinal study of adolescent health: Research design.

Hoff, P. (2018).
Additive and multiplicative effects network models.
arXiv preprint arXiv:1807.08038.

Hoff, P., Fosdick, B., Volfovsky, A., and He, Y. (2017).
amen: Additive and Multiplicative Effects Models for Networks and Relational Data.
R package version 1.3.

Hoff, P., Fosdick, B., Volfovsky, A., and Stovel, K. (2013).
Likelihoods for fixed rank nomination networks.
Network Science, 1(03):253–277.

Hoff, P., Raftery, A., and Handcock, M. (2002).
Latent space approaches to social network analysis.
Journal of the american Statistical association, 97(460):1090–1098.

Holland, P., Laskey, K., and Leinhardt, S. (1983a).
Stochastic blockmodels : First steps.
Social Networks, 5:109–137.

Holland, P., Laskey, K., and Leinhardt, S. (1983b).
Stochastic blockmodels: First steps.
Social Networks, 5:109–137.

Holland, P. and Leinhardt, S. (1981).
An exponetial family of probability distributions for directed graphs.
Journal of the American Statistical Association, 76:33–50.

Lewis, K., Gonzalez, M., and Kaufman, J. (2012).
Social selection and peer influence in an online social network.
Proceedings of the National Academy of Sciences, 109(1):68–72.

Manley, E. (2014).
Identifying functional urban regions within traffic flow.
Regional Studies, Regional Science, 1(1):40–42.

Mossel, E., Neeman, J., and Sly, A. (2014).
Belief propagation, robust reconstruction and optimal recovery of block models.
In *Conference on Learning Theory*, pages 356–370.

Rohe, K., Chatterjee, S., and Yu, B. (2011).
Spectral clustering and the high -dimensional stochastic blockmodel.
The Annals of Statistics, 39(4):1878–1915.

Rohe, K., Qin, T., and Yu, B. (2012).
Co-clustering for directed graphs: the stochastic co-blockmodel and spectral algorithm di-sim.
arXiv preprint arXiv:1204.2296.

Shiokawa, H., Fujiwara, Y., and Onizuka, M. (2013).
Fast algorithm for modularity-based graph clustering.
In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Tambunan, T. (2005).
Promoting small and medium enterprises with a clustering approach: A policy experience from indonesia.
Journal of Small Business Management, 43(2):138–154.

Warner, R. M., Kenny, D. A., and Stoto, M. (1979).
A new round robin analysis of variance for social interaction data.
Journal of Personality and Social Psychology, 37(10):1742.

Wu, Z.-H., Lin, Y.-F., Gregory, S., Wan, H.-Y., and Tian, S.-F. (2012).
Balanced multi-label propagation for overlapping community detection in social networks.
Journal of Computer Science and Technology, 27(3):468–479.