

Inferential Network Analysis with Exponential Random Graph Models

Bruce A. Desmarais
University of Massachusetts Amherst

Data Matters Institute Workshop

June 25, 2015

Is Hagan a Likely Supporter of Burr?



?



What does it mean to model the network?

Construct a probability distribution that
accurately approximates the network

Why build models?

- ▶ Test hypotheses

Example: Does the organization's network exhibit clustering?

- ▶ Simulation for theoretical exploration

Example: How should seats be assigned in a classroom to encourage cross-racial friendships?

- ▶ Tie prediction

Example: Will cross-departmental communication ties persist?

Advantage of ERGM

Can model how ties depend upon each other

Modeling Interdependence

Two Classes of Questions: Covariate and Interdependence

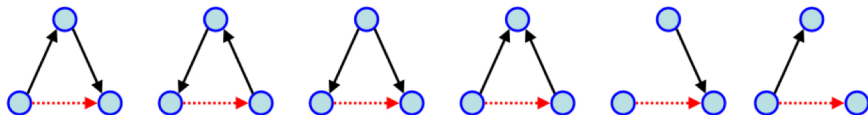
1. Covariate

- ▶ Do legislators in the same political party collaborate more frequently than those in opposite parties?
- ▶ Do states with democratic governments have more alliances than those with autocratic regimes?

2. Interdependence

- ▶ Are two states at war with the same third state less likely to be at war with each other?
- ▶ Are there popularity effects in the choice of co-authors?

ERGM: integrate effects for any forms of (1) and (2) into a unified model of a network.



Estimation

Finding parameter values that result in a distribution that would have produced something similar to what we observe.

Parametric Probabilistic Modeling and the Likelihood Framework of Inference

We observe x , a draw of a random variable X .

$$X \sim f(X, \boldsymbol{\theta})$$

f is a family of probability distributions and $\boldsymbol{\theta}$ is unknown.
 X could be

- ▶ The dependent variable in a regression model
- ▶ An adjacency matrix
- ▶ The text in a document

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg \max_{\boldsymbol{\theta}} [f(x, \boldsymbol{\theta})]$$

1. In many cases, $\hat{\boldsymbol{\theta}}_{MLE}$ is asymptotically normally distributed
2. If f is exponential family, $\ln [f(x, \boldsymbol{\theta})]$ globally concave in $\boldsymbol{\theta}$

The Exponential Random Graph Model (ERGM)

The probability (likelihood function) of observing network N is:

$$\mathcal{P}(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}'\mathbf{h}(N^*)\}}$$

Decomposition:

$$\underbrace{\mathbf{h}(N)}_{\text{Net Stats}} \quad \underbrace{\boldsymbol{\theta}}_{\text{Effects}} \quad \underbrace{\exp\{\boldsymbol{\theta}'\mathbf{h}(N)\}}_{+ \text{ Weight}} \quad \underbrace{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}'\mathbf{h}(N^*)\}}_{\text{Normalizer}}$$

Flexible: \mathbf{h} can capture virtually any form of interdependence among the edges + covariates

Normalizing constant can make estimation difficult

ERGM Task List

What the Modeler Does:

- ▶ Conceptually defines dependencies that should/might exist in the network.
- ▶ Defines (selects) empirical measures of those dependencies. (i.e., $\mathbf{h}(N)$)

What ERGM Software Does:

- ▶ Finds most likely set of θ .
- ▶ Simulates networks so you can check model fit.

Defining h

How would we measure **reciprocity**?

A statistic we would expect to be high if ties were reciprocated a lot and low if they were not reciprocated.

Unpacking h

- Dyadic Covariate

$$h_D(N, X) = \sum_{ij} N_{ij} X_{ij}$$

- Sender Covariate

$$h_{VS}(N, VS) = \sum_i VS_i \sum_{j \neq i} N_{ij}$$

- Receiver Covariate

$$h_{VR}(N, VR) = \sum_i VR_i \sum_{j \neq i} N_{ji}$$

- Reciprocity

$$h_R(N) = \sum_{i < j} N_{ij} N_{ji}$$



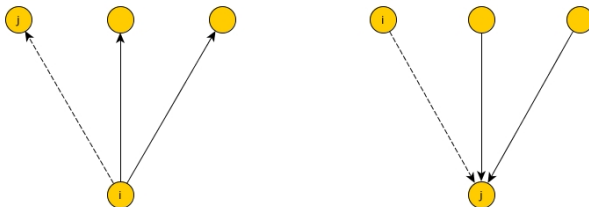
Unpacking h

► Popularity

$$h_P(N) = \sum_{i,j,k} N_{ji}N_{ki} + N_{kj}N_{ij} + N_{ik}N_{jk}$$

► Sociality

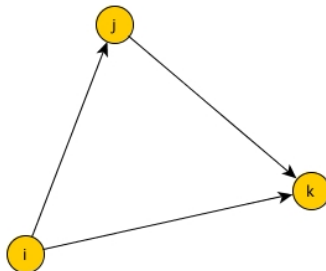
$$h_S(N) = \sum_{i,j,k} N_{ij}N_{ik} + N_{jk}N_{ji} + N_{ki}N_{kj}$$



Unpacking h

- Transitivity

$$h_T(N) = \sum_i \sum_{i \neq j, k} N_{ij} N_{ik} N_{jk}$$



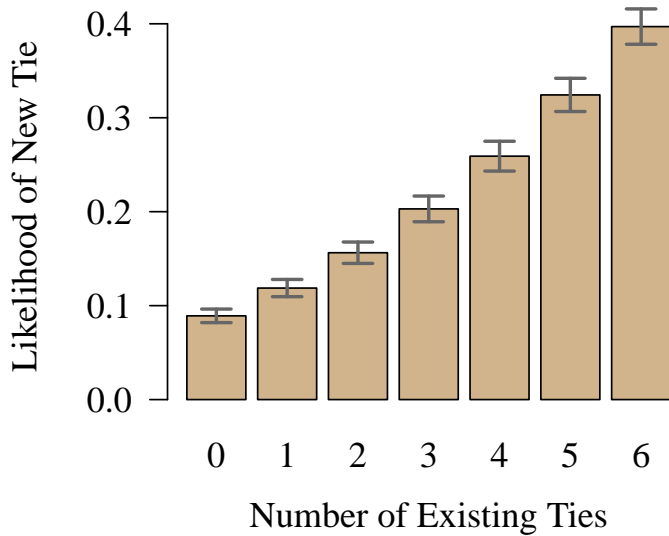
- For detailed discussions on the selection of network statistics, see Snijders et al. (2006) and Goodreau (2007)

Interpretation of ERGM

ERGM offers an incredibly flexible model – it can be used to investigate individual, dyad, node and network level effects.

Two levels of interpretation

1. **(Network)** The relative likelihood of observing N^{j+} to observing N^j is $\exp(\theta_j)$, where
 - ▶ θ_j is the estimate of the parameter that corresponds to statistic j .
 - ▶ N^{j+} is one unit greater than N^j on statistic j (e.g., one more closed triangle, one more edge), *ceteris paribus*.
2. **(Edge)** $P(N_{ij} = 1 | N_{-ij}, \theta) = \text{logit}^{-1} \left(\sum_{r=1}^k \theta_r \delta_r^{(ij)}(N) \right)$
 - ▶ N_{-ij} indicates the network excluding N_{ij}
 - ▶ $\delta_r^{(ij)}(N)$ is equal to the change in h_r when N_{ij} is changed from zero to one
 - ▶ $\text{logit}^{-1}(x) = 1/(1 + \exp(-x))$ (i.e., inverse logit function)



ERGM (ML) Estimation

ML Estimation is very attractive because ERGM is a canonical exponential family form

- ▶ Globally concave loglikelihood
- ▶ \mathbf{h} is minimally sufficient for $\boldsymbol{\theta}$
- ▶ $\hat{\boldsymbol{\theta}}$ is asymptotically normal

Estimation requires we confront... $\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}' \mathbf{h}(N^*)\}$
Sum over m networks...

Nodes	m (i.e., unique undirected networks)
5	1,024
10	35,184,370,000,000
15	40,564,820,000,000,000,000,000,000,000,000

Simulation-based MLE

- ▶ Normalizer is a sum over a population of m networks.
- ▶ **Solution:** approximate with a random sample of networks.

Markov Chain Monte Carlo Maximum Likelihood

m is the MCMC sample size

v is the number of vertices in the network

$\boldsymbol{\theta}^{[0]} = \mathbf{0}$ or $\boldsymbol{\theta}^{[\text{MPLE}]}$

$i = 0$

repeat

$i = i + 1$

draw $\tilde{N} \sim \mathcal{P}(N, \boldsymbol{\theta}^{[i-1]})$ by MCMC, m networks with v

vertices

$$\widehat{C(\boldsymbol{\theta})} = \ln \left(\sum_{j=1}^m \exp \left[(\boldsymbol{\theta} - \boldsymbol{\theta}^{[i-1]})' \mathbf{h}(\tilde{N}_j) \right] \right)$$

$$\boldsymbol{\theta}^{[i]} = \arg \max_{\boldsymbol{\theta}} \left[\sum_{t=1}^n \boldsymbol{\theta}' \mathbf{h}(N_t) - \widehat{C(\boldsymbol{\theta})} \right]$$

until convergence

Degeneracy

Complexity and Flexibility: Combinatorial properties of high order dependence functions induce unintended consequences.

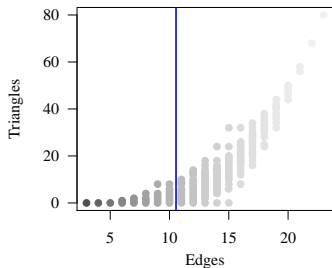
- ▶ Each edge can be involved in $n - 2$ triangles

Degeneracy: Most probability mass concentrated on a few networks, most commonly, the **completely full** or **completely empty** network.

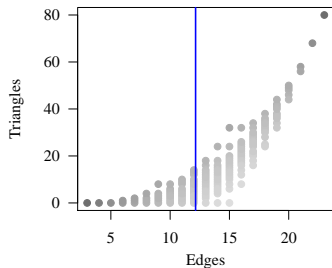
Avoid Degenerate Models!! They constitute completely unrealistic characterizations of the data generating process.

5 node directed net with the number of **Edges** and **Triangles**

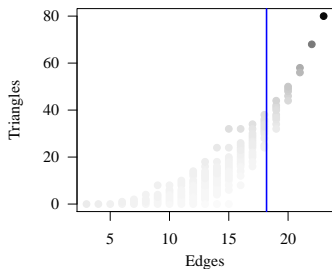
$$\theta_E = 0.50, \theta_T = 0$$



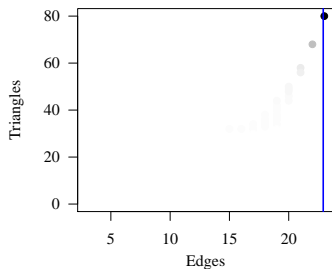
$$\theta_E = 0.50, \theta_T = 0.125$$



$$\theta_E = 0.50, \theta_T = 0.25$$



$$\theta_E = 0.50, \theta_T = 0.50$$



Dealing with Degeneracy

Common Solution: use statistics that down-weight repeated structures that involve the same edge.

Example: Transitivity

- ▶ Classic measure: Number of triangles in the network

$$\sum_{ijk} N_{ij} N_{ik} N_{jk}$$

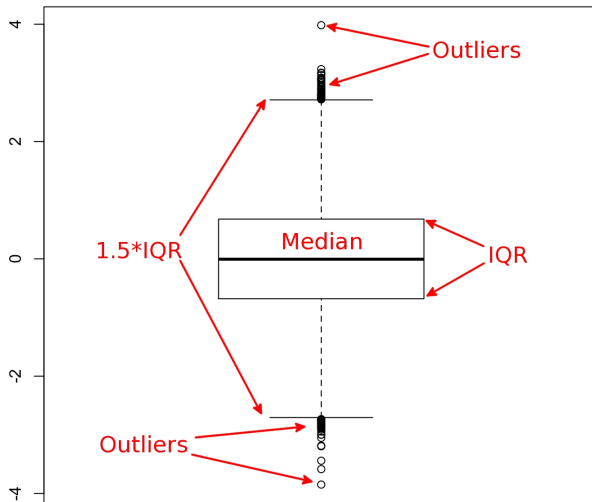
- ▶ Prone to degeneracy. There are probably decreasing marginal returns to indirect connections
 - ▶ I am probably not twice as likely to befriend the friend of two of my friends as I am the friend of one of my friends.
- ▶ Geometrically Weighted Edgewise Shared Partners

$$\sum_{i=1}^{n-2} \left[1 - \left(1 - e^{-\phi} \right)^i \right] EP_i(N)$$

Pre-Interpretation Checklist

- ▶ **Assessing Convergence of Markov Chains:** Assures that the MC process is accurate.
- ▶ **Assessing Degeneracy:** Assures you don't report a degenerate model.
- ▶ **Posterior predictive checking (i.e., gof):** Assures model fits the network well (i.e., you're not missing anything big).

Posterior Predictive Checks: Boxplots



Three Thoughts on Specification

1. Specify baseline model from theory
2. Add so that data is not an outlier WRT model
3. Don't curve-fit!

Running ERGMs

```
> spec4 <- ergm(adviceNet~edges+mutual+ostar(2:3)+gwesp(0,fixed=T)
+edgecov("reportsto")+nodeicov("Tenure")+nodeocov("Tenure")+absdiff("Tenure")
+nodeicov("Age")+nodeocov("Age")+absdiff("Age"),
control=control.ergm(MCMC.samplesize=50000,MCMC.burnin=100000,MCMLL.maxit=10))
```

Checking Convergence: `mcmc.diagnostics()`

Sample statistics burn-in diagnostic (Geweke):

Chain 1

Fraction in 1st window = 0.1

Fraction in 2nd window = 0.5

edges	mutual	ostar2	ostar3
-0.04568	-0.55592	-0.08828	-0.12607
gwesp	edgecov.reportsto	nodeicov.Tenure	nodeocov.Tenure
-0.09754	0.45148	0.20936	0.12223
absdiff.Tenure	nodeicov.Age	nodeocov.Age	absdiff.Age
0.72878	-0.02782	-0.16250	-0.56763

Individual P-values (lower = worse):

edges	mutual	ostar2	ostar3
0.9635690	0.5782674	0.9296563	0.8996785
gwesp	edgecov.reportsto	nodeicov.Tenure	nodeocov.Tenure
0.9222977	0.6516459	0.8341710	0.9027132
absdiff.Tenure	nodeicov.Age	nodeocov.Age	absdiff.Age
0.4661376	0.9778072	0.8709090	0.5702871

Joint P-value (lower = worse): 0.02416657 .

Checking Degeneracy:

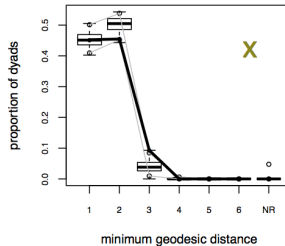
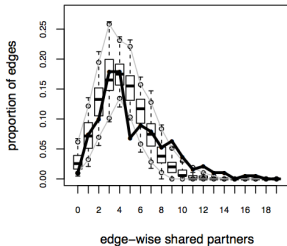
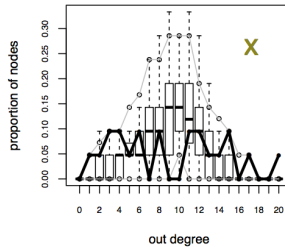
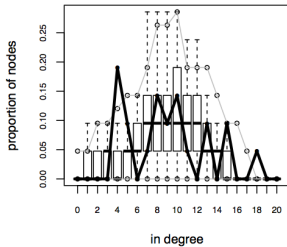
```
summary(gof(spec4, GOF= model))
```

Goodness-of-fit for model statistics

	obs	min	mean	max	MC	p-value
edges	190	133	188.09	269		0.92
mutual	45	24	43.62	81		0.80
ostar2	1062	542	1044.81	1922		0.86
ostar3	4322	1758	4250.62	9438		0.84
edgecov.reportsto	19	15	19.04	20		1.00
nodeicov.Tenure	2588	1971	2563.76	3488		0.96
nodeocov.Tenure	1871	1337	1849.11	2466		0.82
absdiff.Tenure	1401	988	1386.93	2019		0.80
nodeicov.Age	7514	5394	7440.83	10649		0.94
nodeocov.Age	7384	5018	7330.23	10469		0.86
absdiff.Age	1732	1156	1718.62	2606		0.86

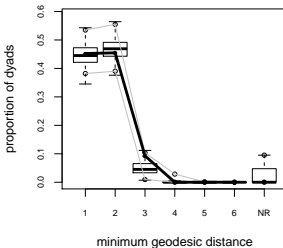
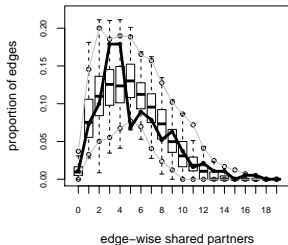
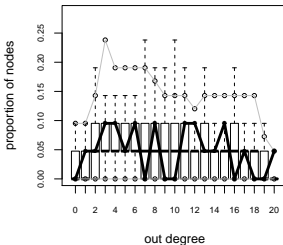
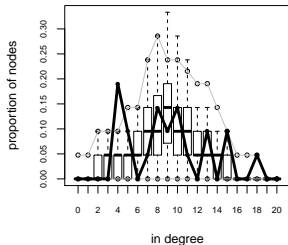
Checking Fit: `plot(gof())`

Goodness-of-fit diagnostics



Checking Fit: `plot(gof())`

Goodness-of-fit diagnostics



Examining Results: summary()

```
> summary(spec4)
```

```
=====
```

```
Summary of model fit
```

```
=====
```

```
Formula: adviceNet ~ edges + mutual + ostar(2:3) + gwesp(0, fixed = T) +  
edgecov("reportsto") + nodeicov("Tenure") + nodeocov("Tenure") +  
absdiff("Tenure") + nodeicov("Age") + nodeocov("Age") + absdiff("Age")
```

```
Iterations: 10
```

```
Monte Carlo MLE Results:
```

	Estimate	Std. Error	MCMC %	p-value
edges	-3.656105	1.196786	0	0.002399 **
mutual	0.481182	0.345224	0	0.164129
ostar2	0.226369	0.099784	0	0.023814 *
ostar3	0.001898	0.010772	0	0.860255
gwesp.fixed.0	1.372966	0.753753	0	0.069262 .
edgecov.reportsto	3.774979	1.109190	0	0.000731 ***
nodeicov.Tenure	0.138554	0.023185	0	< 1e-04 ***
nodeocov.Tenure	-0.022392	0.013914	0	0.108332
absdiff.Tenure	-0.049885	0.019012	0	0.009021 **
nodeicov.Age	-0.047557	0.019370	0	0.014498 *
nodeocov.Age	0.015321	0.009177	0	0.095776 .
absdiff.Age	-0.017134	0.015228	0	0.261174

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Null Deviance: 582.2 on 420 degrees of freedom  
Residual Deviance: 415.2 on 408 degrees of freedom
```

```
AIC: 439.2 BIC: 487.6 (Smaller is better.)
```

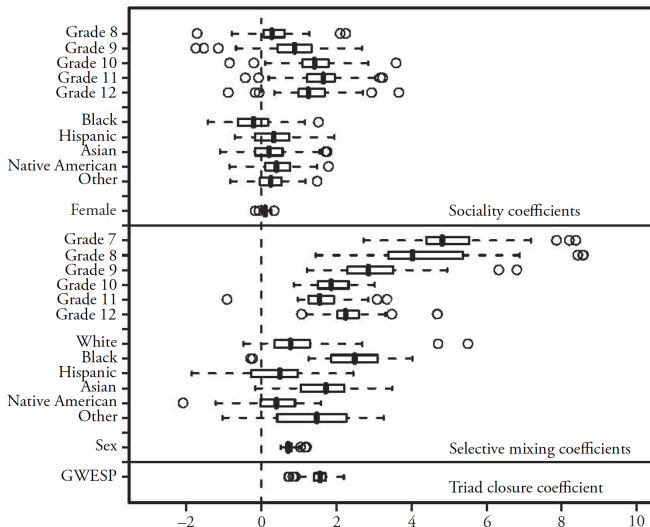
BIRDS OF A FEATHER, OR FRIEND OF A FRIEND? USING EXPONENTIAL RANDOM GRAPH MODELS TO INVESTIGATE ADOLESCENT SOCIAL NETWORKS*

STEVEN M. GOODREAU, JAMES A. KITTS, AND MARTINA MORRIS

In this article, we use newly developed statistical methods to examine the generative processes that give rise to widespread patterns in friendship networks. The methods incorporate both traditional demographic measures on individuals (age, sex, and race) and network measures for structural processes operating on individual, dyadic, and triadic levels. We apply the methods to adolescent friendship networks in 59 U.S. schools from the National Longitudinal Survey of Adolescent Health (Add Health). We model friendship formation as a selection process constrained by individuals' sociality (propensity to make friends), selective mixing in dyads (friendships within race, grade, or sex categories are differentially likely relative to cross-category friendships), and closure in triads (a friend's friends are more likely to become friends), given local population composition. Blacks are generally the most cohesive racial category, although when whites are in the minority, they display stronger selective mixing than do blacks when blacks are in the minority. Hispanics exhibit disassortative selective mixing under certain circumstances; in other cases, they exhibit assortative mixing but lack the higher-order cohesion common in other groups. Grade levels are always highly cohesive, while females form triangles more than males. We conclude with a discussion of how network analysis may contribute to our understanding of sociodemographic structure and the processes that create it.

Goodreau, Kitts and Morris, *Demography* 2009

Figure 3. Coefficients From the Full Model, Plotted Across All 59 Schools



Goodreau, Kitts and Morris, *Demography* 2009

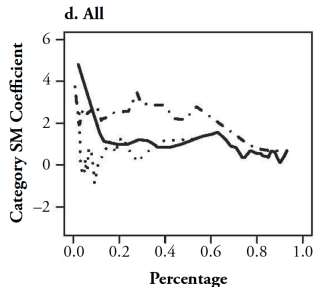
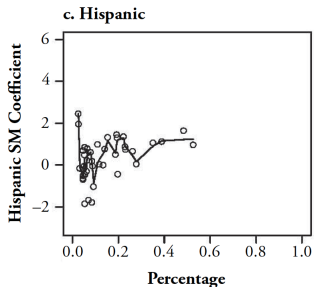
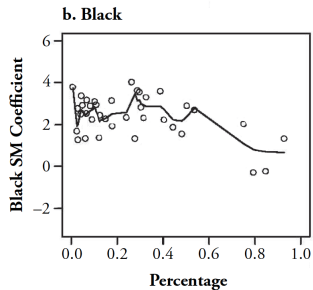
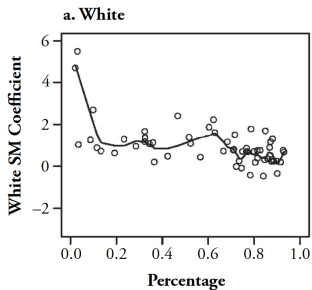
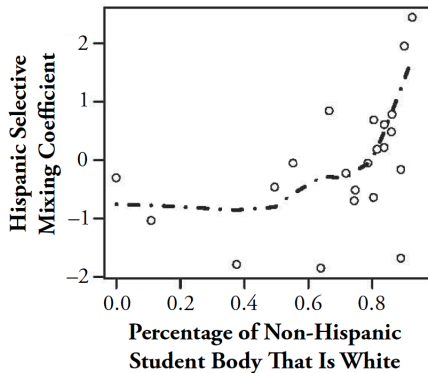
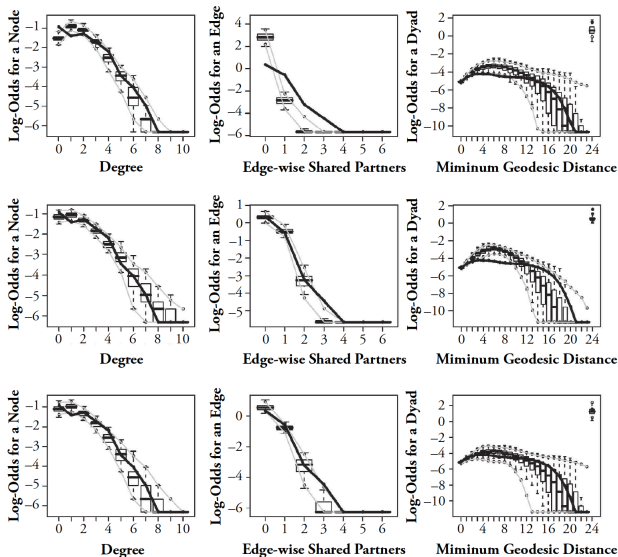


Figure 7. Hispanic Selective Mixing, by Proportion White: Full Model



Goodreau, Kitts and Morris, *Demography* 2009

Figure 8. Goodness-of-Fit Plots: "School 18"



Wrap-up

- ▶ ERGM
 - ▶ Evaluate relationship between network and node attributes
 - ▶ Test for signature network structures (e.g., clusters)
 - ▶ Nothing like it in the literature
- ▶ Extensions to be aware of
 - ▶ Weighted/valude Ties
 - ▶ Network time series
 - ▶ Multipartite and “constrained” networks