Assessing the Goodness-of-Fit of Network Models

Mark S. Handcock

Department of Statistics
University of Washington

Joint work with

David Hunter
Steve Goodreau
Martina Morris

and the

U. Washington Network Modeling Group

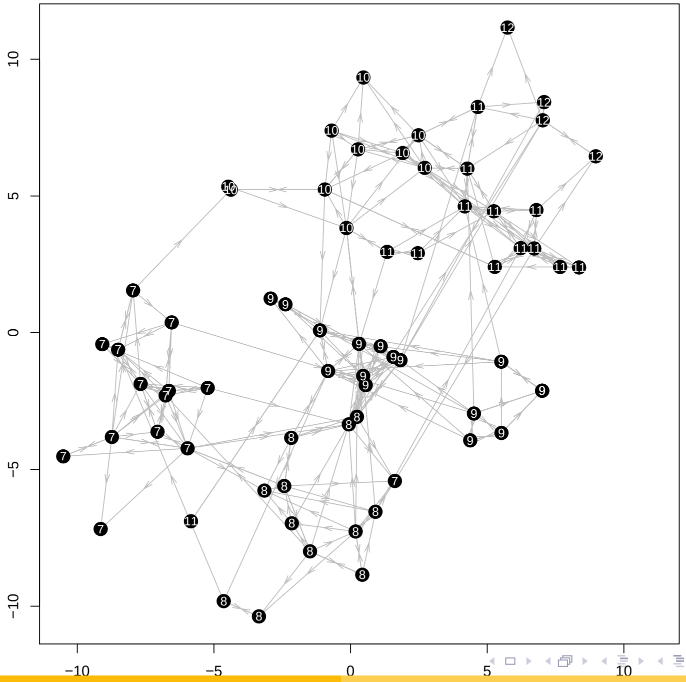
Supported by NIH NIDA Grant DA012831 and NICHD Grant HD041877

NIPS Networks Workshop 2007, December 8 2007

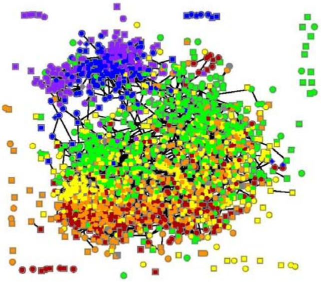
Examples of Friendship Relationships

- The National Longitudinal Study of Adolescent Health
  $\Rightarrow$ `www.cpc.unc.edu/projects/addhealth`

  – “Add Health" is a school-based study of the health-related
  behaviors of adolescents in grades 7 to 12.
- Each nominated up to 5 boys and 5 girls as their friends
- 160 schools: Smallest has 69 adolescents in grades 7–12


−10
−5
0
5
10
−10
−5
0
5
10

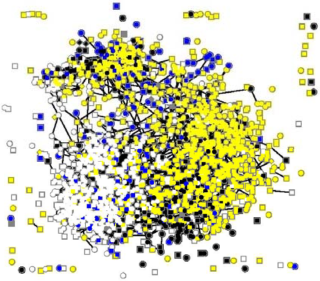
School Community Stratum 44
mutual friendships by Grade



2209 Students

- Grade 7
- Grade 8
- Grade 9
- Grade 10
- Grade 11

School Community Stratum 44
mutual friendships by Race



2209 Students

- **White** (non-Hispanic)
- **Black** (non-Hispanic)
- **Hispanic** (of any race)
- **Asian / Native Am / Other** (non-Hispanic)
- **Race NA**

Notation

A *social network* is defined as a set of n social “actors" and a social relationship between each pair of actors.

Statistical Models for Social Networks

Notation

A *social network* is defined as a set of n social “actors" and a social relationship between each pair of actors.

$$Y_{ij} = \begin{cases} 1 & \text{relationship from actor } i \text{ to actor } j \\ 0 & \text{otherwise} \end{cases}$$

Notation

A *social network* is defined as a set of n social "actors" and a social relationship between each pair of actors.

$$Y_{ij} = \begin{cases} 1 & \text{relationship from actor } i \text{ to actor } j \\ 0 & \text{otherwise} \end{cases}$$

- call $Y \equiv [Y_{ij}]_{n \times n}$ a *sociomatrix*
  - a $N = n(n-1)$ binary array

Notation

A *social network* is defined as a set of n social “actors" and a social relationship between each pair of actors.

$$Y_{ij} = \begin{cases} 1 & \text{relationship from actor } i \text{ to actor } j \\ 0 & \text{otherwise} \end{cases}$$

- call $Y \equiv [Y_{ij}]_{n \times n}$ a *sociomatrix*
  - a $N = n(n-1)$ binary array
- The basic problem of stochastic modeling is to specify a distribution for Y i.e., $P(Y = y)$

A Framework for Network Modeling

Let $\mathcal{Y}$ be the sample space of Y e.g. $\{0,1\}^N$
Any model-class for the multivariate distribution of Y can be *parametrized* in the form:

$$P_\eta(Y=y) = \frac{\exp\{\eta \cdot g(y)\}}{\kappa(\eta,\mathcal{Y})} \qquad y \in \mathcal{Y}$$

Besag (1974), Frank and Strauss (1986)

- $\eta \in \Lambda \subset R^q$ q-vector of parameters
- $g(y)$ q-vector of *network statistics*.
  $\Rightarrow$ $g(Y)$ are jointly sufficient for the model
- For a "saturated" model-class $q = 2^{|\mathcal{Y}|} - 1$
- $\kappa(\eta,\mathcal{Y})$ distribution normalizing constant

$$\kappa(\eta,\mathcal{Y}) = \sum_{y \in \mathcal{Y}} \exp\{\eta \cdot g(y)\}$$

Approximating the loglikelihood

- Suppose $Y_1, Y_2, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} P_{\eta_0}(Y = y)$ for some η_0.
- Using the LOLN, the difference in log-likelihoods is

$$\ell(\eta) - \ell(\eta_0) \quad = \quad \log \frac{\kappa(\eta_0)}{\kappa(\eta)}$$

Approximating the loglikelihood

- Suppose $Y_1, Y_2, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} P_{\eta_0}(Y = y)$ for some η_0.
- Using the LOLN, the difference in log-likelihoods is

$$\begin{aligned}\ell(\eta) - \ell(\eta_0) &= \log \frac{\kappa(\eta_0)}{\kappa(\eta)} \\ &= \log \mathbf{E}_{\eta_0}\left(\exp\{(\eta_0 - \eta)\cdot g(Y)\}\right)\end{aligned}$$

Approximating the loglikelihood

- Suppose $Y_1, Y_2, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} P_{\eta_0}(Y = y)$ for some η_0.
- Using the LOLN, the difference in log-likelihoods is

$$\begin{aligned}
\ell(\eta) - \ell(\eta_0) &= \log \frac{\kappa(\eta_0)}{\kappa(\eta)} \\
&= \log \mathbf{E}_{\eta_0}\left(\exp\{(\eta_0 - \eta)\cdot g(Y)\}\right) \\
&\approx \log \frac{1}{M}\sum_{i=1}^{M} \exp\{(\eta_0 - \eta)\cdot(g(Y_i) - g(y_{\text{obs}}))\} \\
&\equiv \tilde{\ell}(\eta) - \tilde{\ell}(\eta_0).
\end{aligned}$$

Approximating the loglikelihood

- Suppose $Y_1, Y_2, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} P_{\eta_0}(Y = y)$ for some η_0.
- Using the LOLN, the difference in log-likelihoods is

$$
\begin{aligned}
\ell(\eta) - \ell(\eta_0) &= \log \frac{\kappa(\eta_0)}{\kappa(\eta)} \\
&= \log \mathbf{E}_{\eta_0} \left(\exp\{(\eta_0 - \eta)\cdot g(Y)\}\right) \\
&\approx \log \frac{1}{M} \sum_{i=1}^{M} \exp\{(\eta_0 - \eta)\cdot(g(Y_i) - g(y_{\text{obs}}))\} \\
&\equiv \tilde{\ell}(\eta) - \tilde{\ell}(\eta_0).
\end{aligned}
$$

- Simulate $Y_1, Y_2, \ldots, Y_m$ using a MCMC (Metropolis-Hastings) algorithm $\Rightarrow$ Handcock (2002).

Approximating the loglikelihood

- Suppose $Y_1, Y_2, \ldots, Y_m \overset{\text{i.i.d.}}{\sim} P_{\eta_0}(Y = y)$ for some η_0.
- Using the LOLN, the difference in log-likelihoods is

$$\begin{aligned}
\ell(\eta) - \ell(\eta_0) &= \log \frac{\kappa(\eta_0)}{\kappa(\eta)} \\
&= \log \mathbf{E}_{\eta_0}\left(\exp\{(\eta_0 - \eta)\cdot g(Y)\}\right) \\
&\approx \log \frac{1}{M} \sum_{i=1}^{M} \exp\{(\eta_0 - \eta)\cdot(g(Y_i) - g(y_{\text{obs}}))\} \\
&\equiv \tilde{\ell}(\eta) - \tilde{\ell}(\eta_0).
\end{aligned}$$

- Simulate $Y_1, Y_2, \ldots, Y_m$ using a MCMC (Metropolis-Hastings) algorithm $\Rightarrow$ Handcock (2002).
- Approximate the MLE $\hat{\eta} = \text{argmax}_\eta \{\tilde{\ell}(\eta) - \tilde{\ell}(\eta_0)\}$ (MC-MLE) $\Rightarrow$ Geyer and Thompson (1992)

Estimating the change in log-likelihood

- Theoretically, the estimated value of $\ell(\theta) - \ell(\theta_0)$ converges to the true value as the size of the MCMC sample increases, regardless of the value of θ_0.

Estimating the change in log-likelihood

- Theoretically, the estimated value of $\ell(\theta) - \ell(\theta_0)$ converges to the true value as the size of the MCMC sample increases, regardless of the value of θ_0.
- However, in practice this convergence can be agonizingly slow, especially if θ_0 is not chosen close to the maximizer of the likelihood. $\Rightarrow$ Hunter and Handcock (2006)

Measures of the Goodness-of-fit of models

- As this is an Exponential family, natural to measure goodness-of-fit via *deviance*

$$\text{deviance} = 2\left[\ell(\text{saturated model}) - \ell(\hat{\theta})\right]$$

and

$$\text{residual deviance} = 2\left[\ell(\hat{\theta}) - \ell(0)\right]$$

Measures of the Goodness-of-fit of models

- As this is an Exponential family, natural to measure goodness-of-fit via *deviance*

$$\text{deviance} = 2\left[\ell(\text{saturated model}) - \ell(\hat{\theta})\right]$$

and

$$\text{residual deviance} = 2\left[\ell(\hat{\theta}) - \ell(0)\right]$$

- “Standard" asymptotic arguments approximate this by a χ^2 distribution

Measures of the Goodness-of-fit of models

- As this is an Exponential family, natural to measure goodness-of-fit via *deviance*

$$\text{deviance} = 2\left[\ell(\text{saturated model}) - \ell(\hat{\theta})\right]$$

and

$$\text{residual deviance} = 2\left[\ell(\hat{\theta}) - \ell(0)\right]$$

- “Standard" asymptotic arguments approximate this by a χ^2 distribution
- The standard asymptotic approximation can be very bad here...

Measures of the Goodness-of-fit of models

- As this is an Exponential family, natural to measure goodness-of-fit via *deviance*

$$\text{deviance} = 2\left[\ell(\text{saturated model}) - \ell(\hat{\theta})\right]$$

and

$$\text{residual deviance} = 2\left[\ell(\hat{\theta}) - \ell(0)\right]$$

- "Standard" asymptotic arguments approximate this by a χ^2 distribution
- The standard asymptotic approximation can be very bad here... but the deviance may still be a useful measure of fit if properly calibrated. $\Rightarrow$ Hunter and Handcock (2006)

How can we tell if a model class is useful?

Many aspects:

- Is the model-class itself able to represent a range of realistic networks?
  - *model degeneracy:* small range of graphs covered as the parameters vary (Handcock 2003)

How can we tell if a model class is useful?

Many aspects:

- Is the model-class itself able to represent a range of realistic networks?
  - *model degeneracy:* small range of graphs covered as the parameters vary (Handcock 2003)
- What are the properties of different methods of estimation?
  - e.g, MLE, psuedolikelihood, Bayesian framework
  - *computational failure*: estimates do not exist for certain observable graphs

How can we tell if a model class is useful?

Many aspects:

- Is the model-class itself able to represent a range of realistic networks?
  – *model degeneracy:* small range of graphs covered as the parameters vary (Handcock 2003)
- What are the properties of different methods of estimation?
  – e.g, MLE, psuedolikelihood, Bayesian framework
  – *computational failure*: estimates do not exist for certain observable graphs
- Can we assess the goodness-of-fit of models?
  – appropriate measures and tests
  (Besag 2000; Hunter, Goodreau, Handcock 2007)

Model Degeneracy

idea: A random graph model is *near degenerate* if the model places almost all its probability mass on a small number of graph configurations in $\mathcal{Y}$.
e.g. empty graph, full graph, an individual graph, no $2-$stars, mono-degree graphs

Model Degeneracy

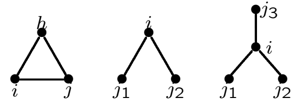
idea: A random graph model is *near degenerate* if the model places almost all its probability mass on a small number of graph configurations in $\mathcal{Y}$.
e.g. empty graph, full graph, an individual graph, no $2-$stars, mono-degree graphs

- Example: The *2-star* model

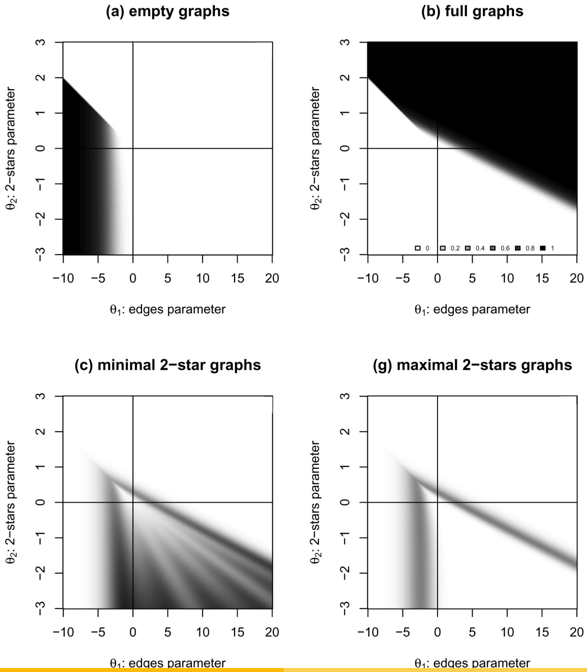
$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

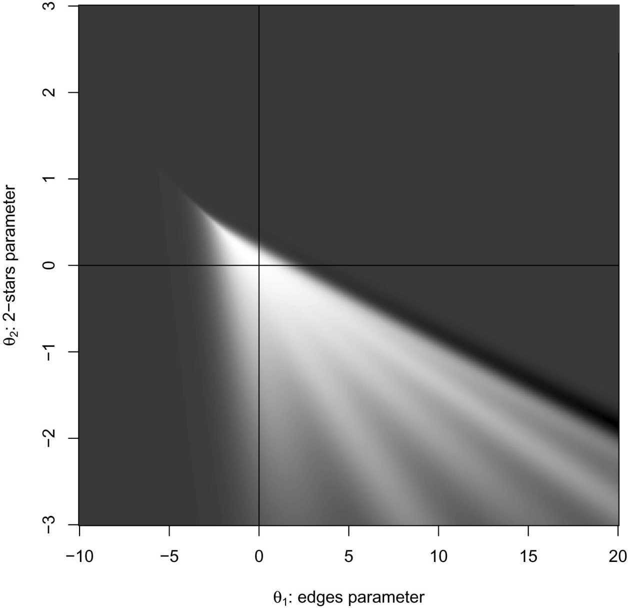
is near-degenerate for most values of $\eta_2 > 0$

$$E(y) = \sum_{i<j} y_{ij} \qquad S(y) = \sum_{i<j<k} y_{ij} y_{ik}$$



triangle = transitive triad   two-star   three-star


(a) empty graphs
(b) full graphs
(c) minimal 2−star graphs
(g) maximal 2−stars graphs
θ_2: 2−stars parameter
θ_1: edges parameter
−10 −5 0 5 10 15 20
−3 −2 −1 0 1 2 3
0 0.2 0.4 0.6 0.8 1


3
2
1
0
−1
−2
−3
θ2: 2−stars parameter
−10
−5
0
5
10
15
20
θ1: edges parameter

Geometry of Exponential Random Graph Models

Consider the alternative parametrization of the models $\mu : \Lambda \rightarrow \mathrm{int}(\mathrm{C})$ defined by

$$\mu(\eta) = \mathbf{E}_{\eta}\left[Z(Y)\right] \equiv \sum_{y \in \mathcal{Y}} Z(y) \frac{\exp\{\eta^T Z(y)\}}{c(\eta)}$$

- The mapping is injective:

$$\mu(\eta_a) = \mu(\eta_b) \rightarrow P_{\eta_a}(Y = y) = P_{\eta_b}(Y = y) \quad \forall y.$$

- The mapping in strictly increasing in the sense that

$$(\eta_a - \eta_b)^T(\mu(\eta_a) - \mu(\eta_b)) \geq 0$$

  with equality only if $P_{\eta_a}(Y = y) = P_{\eta_b}(Y = y)\ \forall y$.
- Represents an alternative *parameterization* of the model

Example of the 2−star model

$$P(Y=y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

where $E(y)$ is the number of edges $(0 – N = \binom{g}{2})$

$S(y)$ is the number of 2−stars $(0 – M = 3\binom{g}{3})$

Example of the 2−star model

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

where $E(y)$ is the number of edges $(0 – N = \binom{g}{2})$

$S(y)$ is the number of 2−stars $(0 – M = 3\binom{g}{3})$

$$\mu_1 = \mathbf{E}_\eta[E(Y)] = \sum_{i<j} \mathbf{E}[Y_{ij}] = N\mathbf{E}[Y_{12}]$$

– μ_1 is the expected number of edges, or
$\frac{1}{N}\mu_1$ is the probability that two actors are linked.

Example of the 2−star model

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

where $E(y)$ is the number of edges $(0 – N = \binom{g}{2})$

$S(y)$ is the number of 2−stars $(0 – M = 3\binom{g}{3})$

$$\mu_1 = \mathbf{E}_\eta[E(Y)] = \sum_{i<j} \mathbf{E}[Y_{ij}] = N\mathbf{E}[Y_{12}]$$

– μ_1 is the expected number of edges, or
$\frac{1}{N}\mu_1$ is the probability that two actors are linked.

$$\mu_2 = \mathbf{E}_\eta[S(Y)] = \sum_{i<j<k} \mathbf{E}[Y_{ij} Y_{ik}] = M\mathbf{E}[Y_{12} Y_{13}]$$

– μ_2 is the expected number of 2−stars, or
$\frac{1}{M}\mu_2$ is the probability that a given actor is tied to
two randomly chosen other actors.

Figure 4: Regions of the parameter space of μ

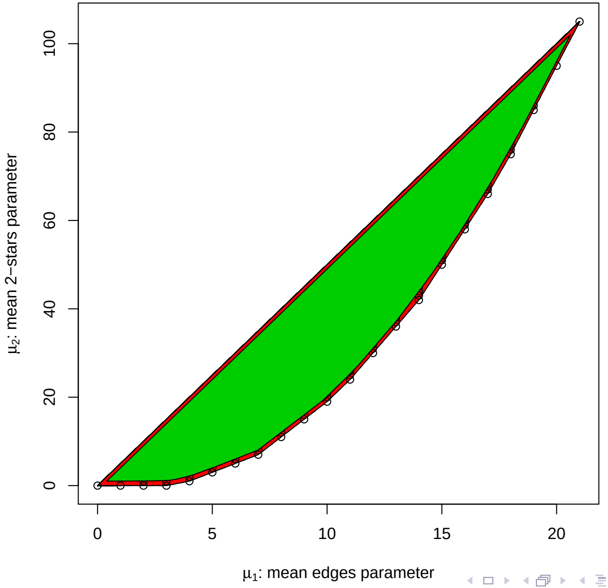
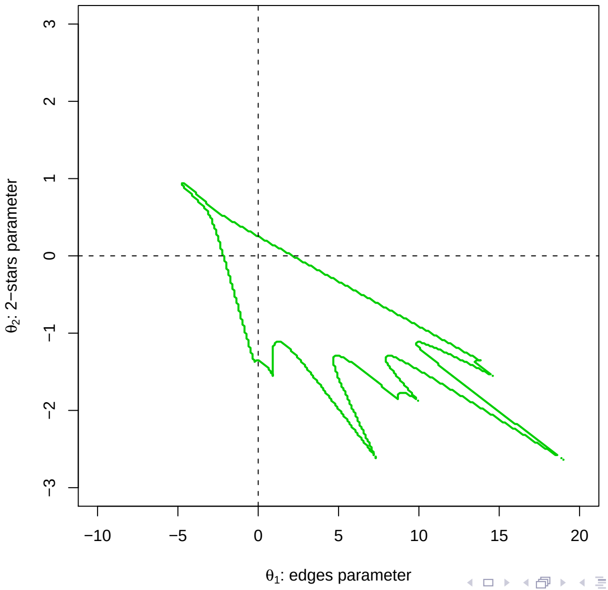


Figure 5: Regions of the parameter space of θ



In some cases mixed parameterizations may be better

Let $(t^{(1)}, t^{(2)})$ be a partition of t such that:

– $t^{(1)}$ is interpretable as a mean value parametrization
– $t^{(2)}$ is interpretable as the “natural" conditional log-odds

Consider similar partitions $(\eta^{(1)}, \eta^{(2)})$ of η and $(\mu^{(1)}(\eta), \mu^{(2)}(\eta))$ of $\mu(\eta)$.
Let $\Lambda^{(2)}$ be the set of values of $\eta^{(2)}$ for η varying in Λ and $C^{(1)}$ be the convex hull of $\{t^{(1)}(y) : y \in \mathcal{Y}\}$.
The mapping $\eta : \Lambda \to \Lambda^{(2)} \times \mathrm{int}(\mathrm{C}^{(1)})$ defined by

$$\eta(\eta) = (\mu^{(1)}(\eta), \eta^{(2)}) \tag{1}$$

is a *mixed* parametization of the model $(\mathcal{Y}, t, \eta)$.
The components $\mu^{(1)}$ and $\eta^{(2)}$ are variationally independent, that is, the range of $\eta(\eta)$ is a product space.

Degeneracy in the mean value parametization

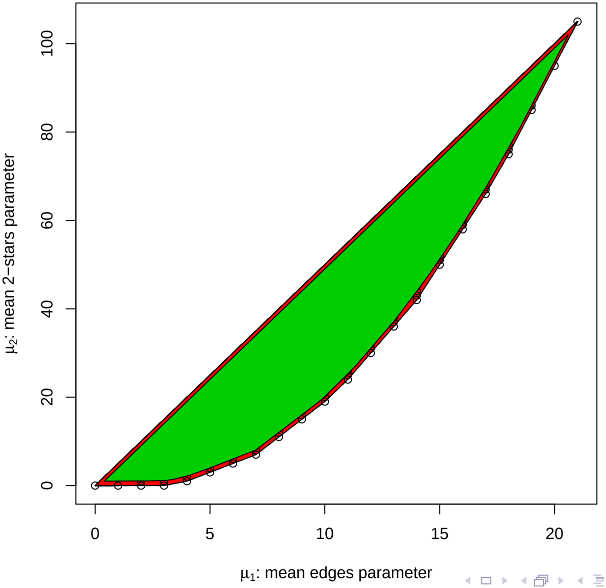
- **Definition:** A model is *near degenerate* if $\mu(\eta)$ is close to the
boundary of C

Degeneracy in the mean value parametization

- **Definition:** A model is *near degenerate* if $\mu(\eta)$ is close to the
boundary of C

Let $\deg \mathcal{Y} = \{y \in \mathcal{Y} : Z(y) \in \mathrm{bdC}\}$ be the set of graph on the boundary of the convex hull.
idea: Based on the geometry of the mean value parametrization the expected sufficient statistics are close to a boundary of the hull and the model will place much probability mass on graphs in $\deg \mathcal{Y}$.

Figure 4: Regions of the parameter space of μ



This statement can be quantified in a number of ways:
Result: Let e be a unit vector in $\mathbf{R}^q$ and $\mathrm{bd}(e) = \sup_{\mu \in \mathrm{intC}}(e^T \mu)$.

1. $\mu(\lambda e) \to \mathrm{bd}(e) e$ as $\lambda \uparrow \infty$.
2. $P_{\lambda e, \mathcal{Y}}(Y \in \deg \mathcal{Y}) \to 1$ as $\lambda \uparrow \infty$.
3. For every $d < \mathrm{bd}(e)$, $P_{\lambda e, \mathcal{Y}}(e^T Z(Y) \leq d) \to 0$ as $\lambda \uparrow \infty$.
4. Let $\eta_0 \in \mathrm{intC}$.
   Then $\mathrm{Kullback-Leibler\ divergence}(\eta_0; \lambda e) \to \infty$ as $\lambda \uparrow \infty$.

Effect of Near-Degeneracy on MCMC Estimation

- Closely related to nice properties of simple MCMC schemes (Geyer 1999).

  – If a random graph model is simulated using a MCMC based on a near-degenerate ψ it will very likely fail.

Effect of Near-Degeneracy on MCMC Estimation

- Closely related to nice properties of simple MCMC schemes (Geyer 1999).

  – If a random graph model is simulated using a MCMC based on a near-degenerate ψ it will very likely fail.
- Full-conditional MCMC with dyad update:

$$M(\psi) = \max_{y \in \mathcal{Y}} |\psi^T \delta(y_{ij}^c)|$$

where $\delta(y_{ij}^c) = Z(y_{ij}^+) - Z(y_{ij}^-)$
– As $\mu(\psi) \to \mathrm{bd(C)}$, $M(\psi) \to \infty$
– There exists $y \in \mathcal{Y}$ with

$$\operatorname{logit}\left[P(Y_{ij} = 1 \mid Y_{ij}^c = y_{ij}^c)\right] = \pm M(\psi)$$

– If ψ is near-degenerate then $M(\psi)$ is large and the MCMC will mix very slowly.

Example of degeneracy of the 2−star model

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

- $M(\eta) = \max\{|\eta_1|, \eta_1 + 2(g-2)\eta_2\}$

Example of degeneracy of the 2−star model

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

- $M(\eta) = \max\{|\eta_1|, \eta_1 + 2(g-2)\eta_2\}$ MCMC will usually mix poorly.

Example of degeneracy of the 2−star model

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

- $M(\eta) = \max\{|\eta_1|, \eta_1 + 2(g-2)\eta_2\}$ MCMC will usually mix poorly.
- If $\mu(\eta)$ close to $(3, 0)$ (e.g., $\eta = (4.5, -18.4)$) then
  $M(\eta) = 4.5$
  So an MCMC will approach $(3, 0)$ and stay there
  (98.9% and 1.1% at $(2, 0) \in \mathrm{bd}(\mathrm{C})$).

Example of degeneracy of the 2−star model

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

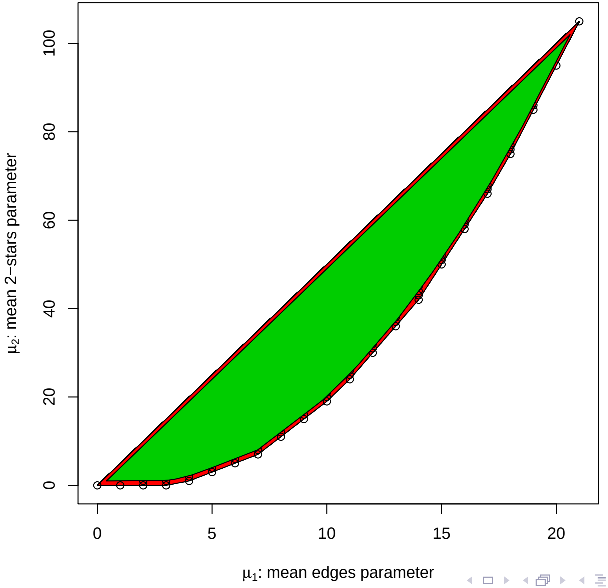
- $M(\eta) = \max\{|\eta_1|, \eta_1 + 2(g-2)\eta_2\}$ MCMC will usually mix poorly.
- If $\mu(\eta)$ close to $(3, 0)$ (e.g., $\eta = (4.5, -18.4)$) then $M(\eta) = 4.5$
  So an MCMC will approach $(3, 0)$ and stay there (98.9% and 1.1% at $(2, 0) \in \mathrm{bd}(\mathrm{C})$).
- If $\mu(\eta)$ close to $(9, 40)$ (e.g., $\eta = (-3.43, 0.683)$) then $M(\eta) = 3.43$. The model places 50% of its mass on graphs with 2 or fewer edges and 36% on graphs with at least 19 edges.

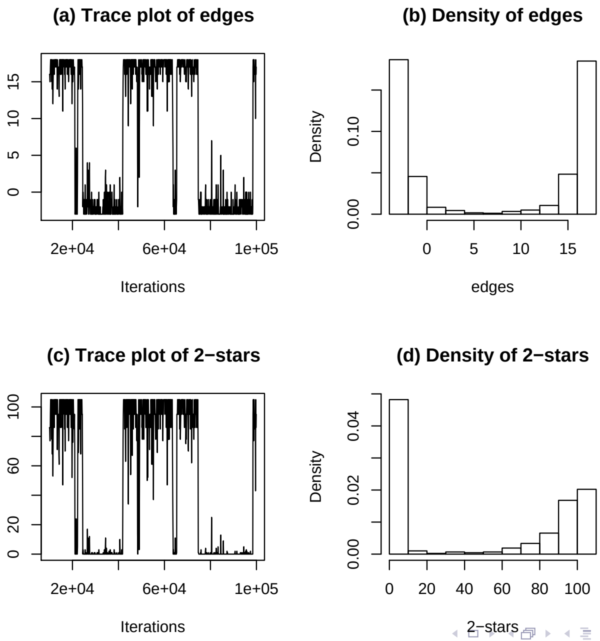
Example of degeneracy of the 2−star model

$$P(Y = y) = \frac{\exp\{\eta_1 E(y) + \eta_2 S(y)\}}{c(\eta_1, \eta_2)} \qquad y \in \mathcal{Y}$$

- $M(\eta) = \max\{|\eta_1|, \eta_1 + 2(g-2)\eta_2\}$ MCMC will usually mix poorly.
- If $\mu(\eta)$ close to $(3, 0)$ (e.g., $\eta = (4.5, -18.4)$) then $M(\eta) = 4.5$
  So an MCMC will approach $(3, 0)$ and stay there (98.9% and 1.1% at $(2, 0) \in \mathrm{bd}(\mathrm{C})$).
- If $\mu(\eta)$ close to $(9, 40)$ (e.g., $\eta = (-3.43, 0.683)$) then $M(\eta) = 3.43$. The model places 50% of its mass on graphs with 2 or fewer edges and 36% on graphs with at least 19 edges.
- The model is also *unstable* e.g., $\eta = (-3.43, 0.67)$) $\mu(\eta) \approx (4.4, 17.1)$ and the model places almost all its mass on empty graphs.

Figure 4: Regions of the parameter space of μ





(a) Trace plot of edges
(b) Density of edges
(c) Trace plot of 2−stars
(d) Density of 2−stars
Density
Iterations
edges
2−stars
2e+04
6e+04
1e+05
0
5
10
15
0.00
0.10
0.02
0.04
20
60
100
40
80

Estimation within the mean value parametization

– If $Z(y_{obs}) \in \mathrm{int}(\mathrm{C})$, the MLE of μ is $Z(y_{obs})$.

Estimation within the mean value parametization

– If $Z(y_{obs}) \in \mathrm{int}(\mathrm{C})$, the MLE of μ is $Z(y_{obs})$.
– If $Z(y_{obs}) \notin \mathrm{int}(\mathrm{C})$ the MLE of μ does not exist.

Estimation within the mean value parametization

– If $Z(y_{obs}) \in \mathrm{int(C)}$, the MLE of μ is $Z(y_{obs})$.
– If $Z(y_{obs}) \notin \mathrm{int(C)}$ the MLE of μ does not exist.
– The MLE $\hat{\mu}$ is unbiased and has minimum variance:

$$\mathbf{E}_\eta(\hat{\mu}) = \mathbf{E}_\eta\left[Z(Y)\right] = \mu(\eta) = \left[\frac{\partial \log c(\eta)}{\partial \eta_i}\right](\eta)$$

$$\mathbf{V}_\eta(\hat{\mu}) = \mathbf{V}_\eta\left[Z(Y)\right] = \left[\frac{\partial^2 \log c(\eta)}{\partial \eta_i \partial \eta_j}\right](\eta)$$

– An estimate of the variance-covariance is available using the same MCMC.

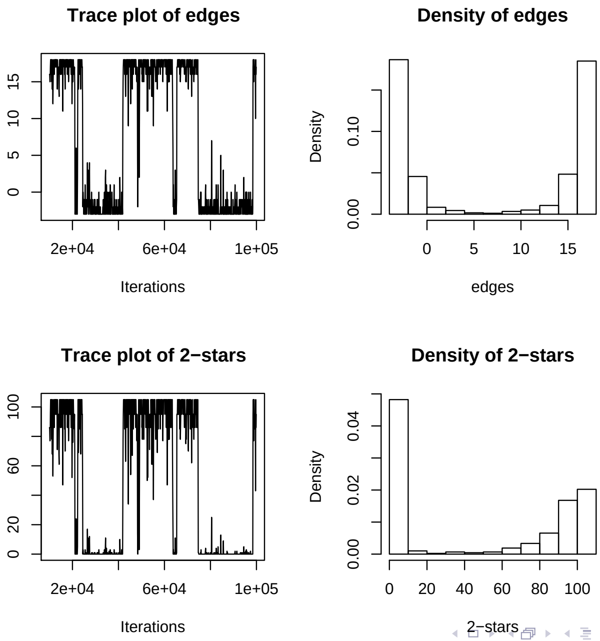
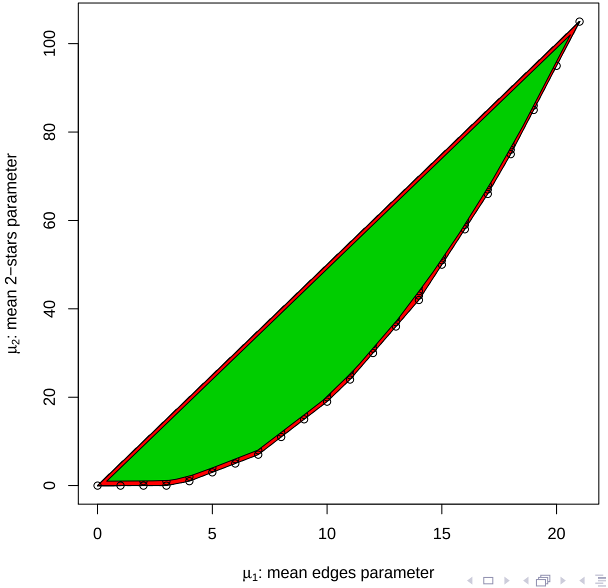

Trace plot of edges
Density of edges
Trace plot of 2−stars
Density of 2−stars
Iterations
edges
2−stars
Density
2e+04
6e+04
1e+05
0
5
10
15
0.00
0.10
20
40
60
80
100
0.02
0.04

Figure 4: Regions of the parameter space of μ



Existence and uniqueness of MLE

Let C be the convex hull of $\{Z(y) : y \in \mathcal{Y}\}$
- the convex hull of the discrete support points.
Let $\mathrm{int(C)}$ be the interior of C.

Existence and uniqueness of MLE

Let C be the convex hull of $\{Z(y) : y \in \mathcal{Y}\}$
- the convex hull of the discrete support points.
Let $\mathrm{int}(\mathrm{C})$ be the interior of C.

Result (Barndorff-Nielsen 1978)
The MLE exists if, and only if, $Z(y_{observed}) \in \mathrm{int}(\mathrm{C})$
If it exists, it is unique and can be found by solving
the likelihood equations or by direct optimization of $\mathcal{L}$.

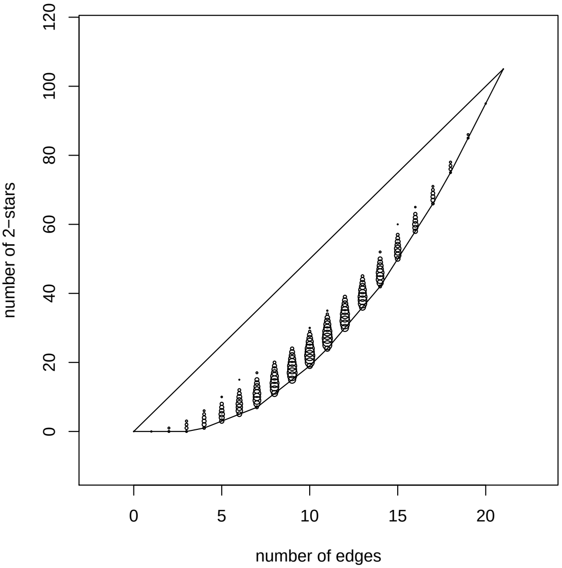


Figure 1: Enumeration of sufficient statistics for graphs with 7 nodes. The circles are centered on

How can we tell if a model class is useful?

Many aspects:

- Is the model-class itself able to represent a range of realistic networks?
  - *model degeneracy:* small range of graphs covered as the parameters vary (Handcock 2003)
- *What are the properties of different methods of estimation?*
  - *e.g, MLE, psuedolikelihood, Bayesian framework*
  - *computational failure: estimates do not exist for certain observable graphs*
- Can we assess the goodness-of-fit of models?
  - appropriate measures and tests
    (Besag 2000; Hunter, Goodreau, Handcock 2007)

Existence and uniqueness of MC-MLE

- Geyer and Thompson (1992) show the MC-MLE converges to the true MLE as the number of simulations increases.

  – also produces estimates of the asymptotic covariance matrix, size of the MCMC induced error, etc.

Let *CO* be the convex hull of *sampled* sufficient statistics.
In practice, three cases:

1. $Z(y) \in \mathrm{int}(\mathrm{CO}) \subset \mathrm{C}$: MC-MLE exists and is unique
2. $Z(y) \notin \mathrm{int}(\mathrm{CO})$ but is in $\mathrm{int}(\mathrm{C})$: MC-MLE does not exist, even though MLE does
3. $Z(y) \notin \mathrm{int}(\mathrm{C})$: MC-MLE and MLE do not exist

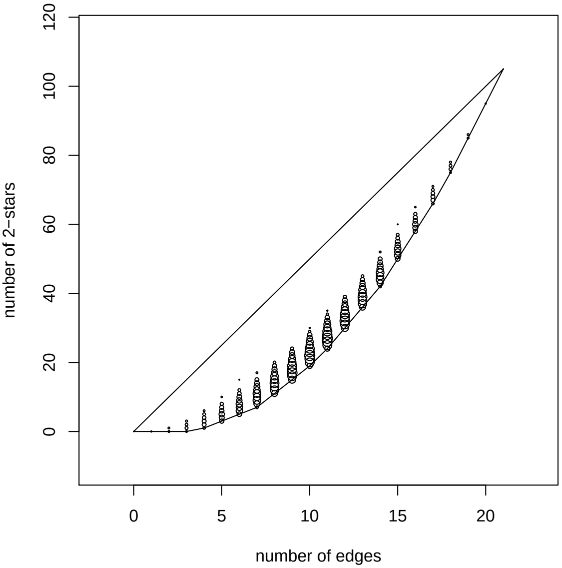


Figure 1: Enumeration of sufficient statistics for graphs with 7 nodes. The circles are centered on

How can we tell if a model class is useful?

Many aspects:

- Is the model-class itself able to represent a range of realistic networks?
  - *model degeneracy:* small range of graphs covered as the parameters vary (Handcock 2003)
- What are the properties of different methods of estimation?
  - e.g, MLE, psuedolikelihood, Bayesian framework
  - *computational failure*: estimates do not exist for certain observable graphs
- *Can we assess the goodness-of-fit of models?*
  - *appropriate measures and tests*

    (Besag 2000; Hunter, Goodreau, Handcock 2007)

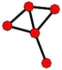
Goodness of fit intuition

ERGM
class
$\exp\{\eta \cdot g(y)\}$

Goodness of fit intuition

ERGM
class
$\exp\{\eta \cdot g(y)\}$ $\longrightarrow$

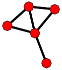
$\uparrow$
y^{obs}



Goodness of fit intuition

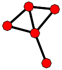
ERGM
class
$\exp\{\eta \cdot g(y)\}$ $\longrightarrow$ (approx) MLE $\widehat{\eta}$

$\uparrow$
y^{obs}



Goodness of fit intuition

ERGM class		(approx) MLE		Fitted ERGM
$\exp\{\eta \cdot g(y)\}$	$\longrightarrow$	$\widehat{\eta}$	$\longrightarrow$	$\exp\{\widehat{\eta} \cdot g(y)\}$
	$\uparrow$			
	y^{obs}			

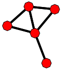


Goodness of fit intuition

ERGM class $\exp\{\eta \cdot g(y)\}$ $\longrightarrow$ (approx) MLE $\widehat{\eta}$ $\longrightarrow$ Fitted ERGM $\exp\{\widehat{\eta} \cdot g(y)\}$

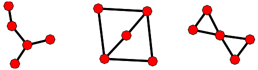
$\uparrow$

y^{obs}



$\downarrow$

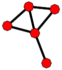
Randomly generated networks $\tilde{Y}_1, \tilde{Y}_2, \ldots$



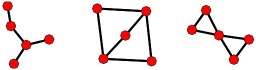
Goodness of fit intuition

ERGM class $\exp\{\eta \cdot g(y)\}$ $\longrightarrow$ (approx) MLE $\widehat{\eta}$ $\longrightarrow$ Fitted ERGM $\exp\{\widehat{\eta} \cdot g(y)\}$

$\uparrow$ y^{obs}



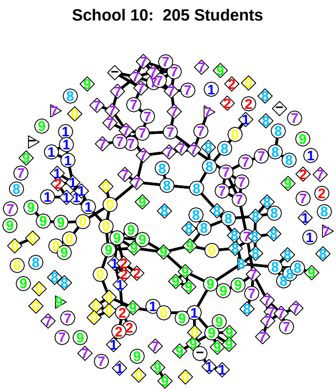
$\downarrow$ Randomly generated networks $\tilde{Y}_1, \tilde{Y}_2, \ldots$



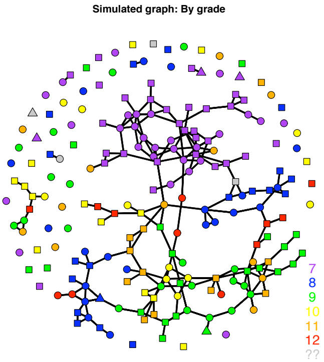
- Question: How does y^{obs} “look” as a representative of the sample $\tilde{Y}_1, \tilde{Y}_2, \ldots$?

The eyeball test

The data:

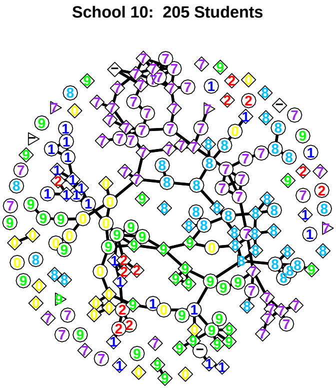


Simulated network, model A:

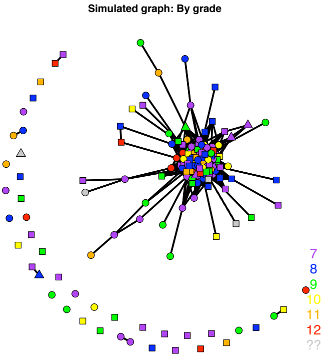


The eyeball test (cont'd)

The data:



Simulated network, model B:



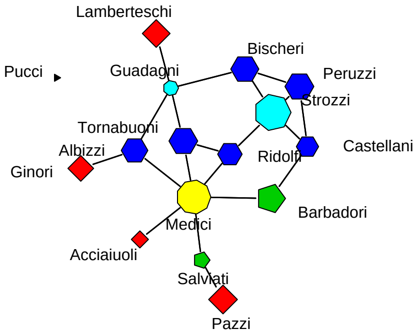
(Yikes!)

The models

- Model A: $g(y)$ contains terms for
  - # of edges
  - Homophily effects of grade, sex, and race factors
  - Main effects of grade, sex, and race factors
  - $\sum_i (.632)^i \mathrm{EP}_i$, where $\mathrm{EP}_i =$# edges with i shared partners
- Model B: $g(y)$ contains terms for
  - # of edges
  - # of neighbors of the same sex (homophily effect)
  - # of 2-stars
  - # of triangles

(Note: It was necessary to use MPLE to fit Model B)

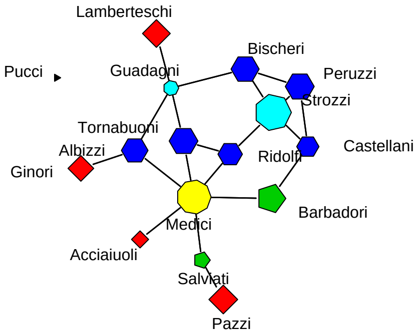
A well-known example:



Florentine marriage data

- Edge indicates marriage tie between families
- Sides=degree + 3
- Color=degree
- Size=log(wealth)

A well-known example:



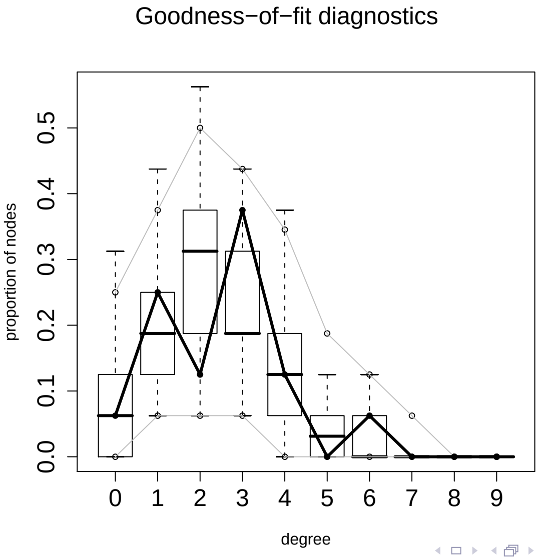
Florentine marriage data

- Edge indicates marriage tie between families
- Sides=degree + 3
- Color=degree
- Size=log(wealth)

```
model1 <- ergm(flomarriage ~ edges + kstar(2))
```

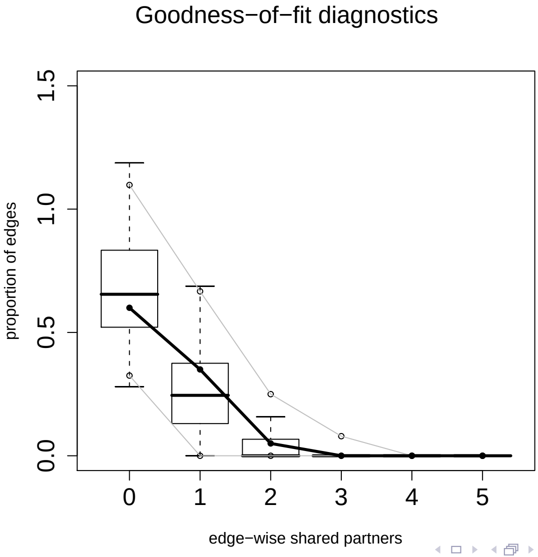
Graphical GOF check: degree distribution

```
model1 <- ergm(flomarriage ~ edges + kstar(2))
```



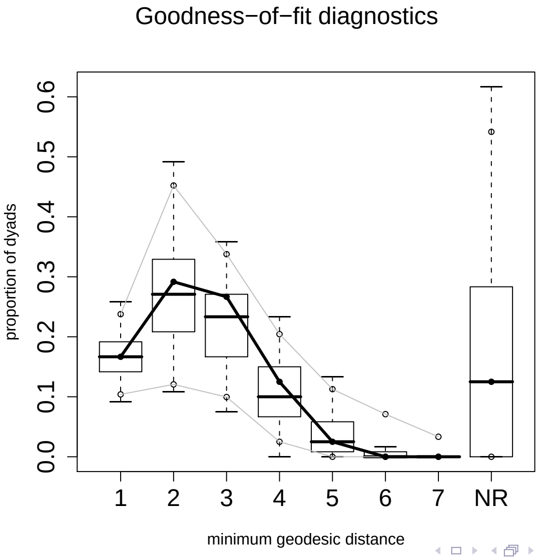
Graphical GOF: edgewise shared partner distribution

```
model1 <- ergm(flomarriage ~ edges + kstar(2))
```



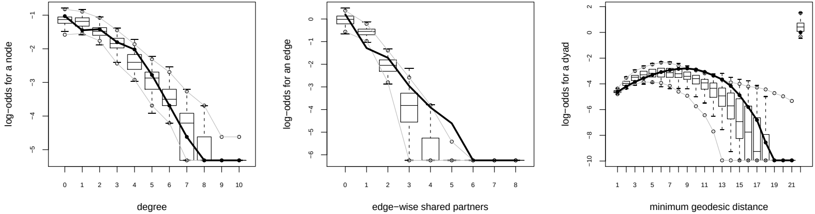
Graphical GOF check: geodesic distance distribution

```
model1 <- ergm(flomarriage ~ edges + kstar(2))
```

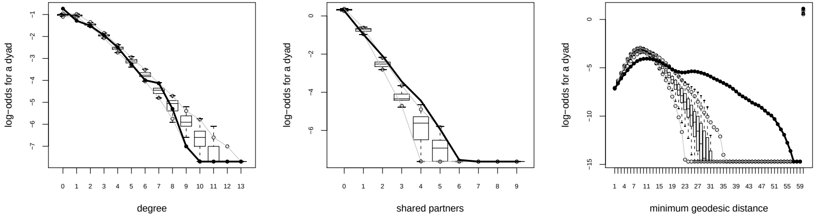


GOF check: Examples from Add Health networks

n=205



n=2209



MCMC Testing

- Significance tests based on comparing the observed value of a statistics to a null probability distribution.
- MCMC $p-$values $\Rightarrow$ Besag and Clifford (1991), Besag (2000)

Illustration: Empirical evidence of competition among Darwin's Finches

Finch	Island																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Large ground finch	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Medium ground finch	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0
Small ground finch	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0
Sharp-beaked ground finch	0	0	1	1	1	0	0	1	0	1	0	1	1	0	1	1	1
Cactus ground finch	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	0
Large cactus ground finch	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
Large tree finch	0	0	1	1	1	1	1	1	1	0	0	1	0	1	1	0	0
Medium tree finch	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Small tree finch	0	0	1	1	1	1	1	1	1	1	0	1	0	0	1	0	0
Vegetarian finch	0	0	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0
Woodpecker finch	0	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0
Mangrove finch	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Warbler finch	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table: Darwin's finch data

MCMC Testing and $p-$values

- Does the observed grouping of finch species on islands happened by random chance or if it was the result of a struggle in which only species which depended on different food sources could coexist on an island.
- To test this hypothesis, consider the test statistic

$$\bar{S}^2 = \frac{1}{m(m-1)} \sum_{i \neq j} s_{ij}^2,$$

where m is the number of finch species, $S = (s_{ij}) = AA^T$, and $A = (a_{ij})$ is the bipartite graph in the table.

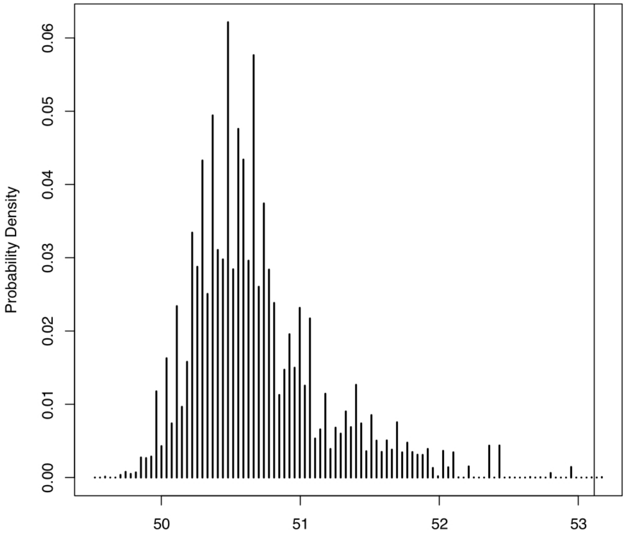


Figure: Null distribution of the test statistic $\bar{S}^2$

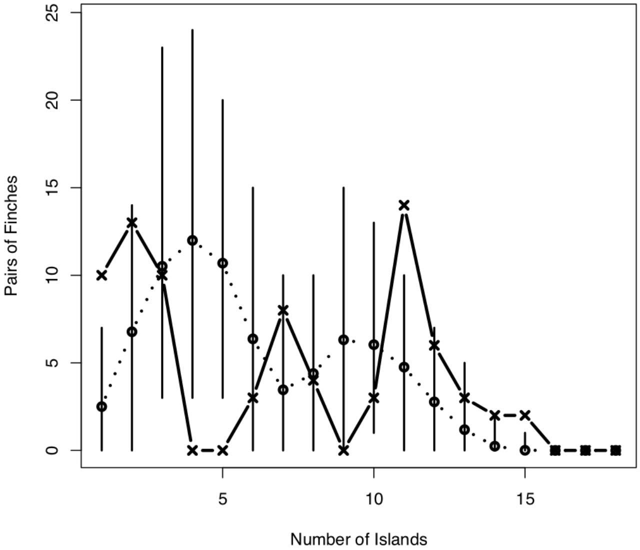


Figure: Number of pairs of finches sharing x islands, $x = 0, 1, ..., 17$

Summary

- Network representations intersect with most sciences
- Sparse models are being used to capture structural properties
- The models must depend on the scientific objective.
- Some seemingly simple models are not so.
- The inclusion of attributes is very important
  - actor attributes
  - dyad attributes e.g. homophily, race, location
  - structural terms e.g. transitive homophily

- We need better and more local models for social networks: e.g. “nearest neighbor" ideas for local dependence
  - $\Rightarrow$ Baddeley and Moller (1989)
  - $\Rightarrow$ Snijders, Robins, Pattison, Handcock (2006)
- Taking into account class membership is very important

  – known classes “block models"
  - $\Rightarrow$ Wang and Wong (1987)

- latent class and trait models are important
  - an underlying latent “social space" of actors
    - $\Rightarrow$ Hoff, Raftery and Handcock (2002)
    - $\Rightarrow$ Hoff (2003, 2004 ,...)
  - latent class models are very promising
    - $\Rightarrow$ Nowicki and Snijders (2001)
  - latent class and trait models
    - $\Rightarrow$ Handcock, Raftery, Tantrum (2007); Krivitsky et. al (2007)
    - $\Rightarrow$ Hoff (2005, 2007)
  - grade of membership models
    - $\Rightarrow$ Airoldi, Blei, Feinberg (2007)