A comparison of various estimation methods for ERGMs

Mark S. Handcock
University of Washington

joint work with Krista J. Gile
Nuffield College, Oxford

Marijtje van Duijn
Groningen, Netherlands

MURI-UCI August 25, 2009

For details, see:

- van Duijn, M. A. J., Gile, K. and Handcock, M.S. (2008). A Framework for the Comparison of Maximum Pseudo Likelihood and Maximum Likelihood Estimation of Exponential Family Random Graph Models. *Social Networks*, doi:10.1016/j.socnet.2008.10.003[1]

[1]Research supported by NICHD grant 7R29HD034957 and NIDA 7R01DA012831, and ONR award N00014-08-1-1015.

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.

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

- 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 $q = |\mathcal{Y}| - 1$ e.g. $2^N - 1$
- $\kappa(\eta, \mathcal{Y})$ distribution normalizing constant

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

Statistical Inference for η

Base inference on the loglikelihood function,

$$\ell(\eta) = \eta \cdot g(y_{\text{obs}}) - \log \kappa(\eta)$$

$$\kappa(\eta) = \sum_{\substack{\text{all possible} \\ \text{graphs } z}} \exp\{\eta \cdot g(z)\}$$

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

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

- Simulate $Y_1, Y_2, \ldots, Y_m$ using a MCMC (Metropolis-Hastings) algorithm
  $\Rightarrow$ Snijders (2002); Handcock (2002).
- Approximate the MLE $\hat{\eta} = \text{argmax}_\eta \{\tilde{\ell}(\eta) - \tilde{\ell}(\eta_0)\}$ (MC-MLE)
  $\Rightarrow$ Geyer and Thompson (1992)
- Given a random sample of networks from P_{η_0}, we can thus approximate (and subsequently maximize) the loglikelihood shifted by a constant.

Conditional log-odds of an edge

Notation: For a network y and a pair (i, j) of nodes,

- $y_{ij} = 0$ or 1, depending on whether there is an edge
- y_{ij}^c denotes the status of all pairs in y other than (i, j)
- y_{ij}^+ denotes the same network as y but with $y_{ij} = 1$
- y_{ij}^- denotes the same network as y but with $y_{ij} = 0$

Conditional log-odds of an edge

Notation: For a network y and a pair (i, j) of nodes,

- $y_{ij} = 0$ or 1, depending on whether there is an edge
- y_{ij}^c denotes the status of all pairs in y other than (i, j)
- y_{ij}^+ denotes the same network as y but with $y_{ij} = 1$
- y_{ij}^- denotes the same network as y but with $y_{ij} = 0$

Conditional on $Y_{ij}^c = y_{ij}^c$, Y has only two possible states, depending on whether $Y_{ij} = 0$ or $Y_{ij} = 1$.

Conditional log-odds of an edge

Notation: For a network y and a pair (i, j) of nodes,

- $y_{ij} = 0$ or 1, depending on whether there is an edge
- y_{ij}^c denotes the status of all pairs in y other than (i, j)
- y_{ij}^+ denotes the same network as y but with $y_{ij} = 1$
- y_{ij}^- denotes the same network as y but with $y_{ij} = 0$

Conditional on $Y_{ij}^c = y_{ij}^c$, Y has only two possible states, depending on whether $Y_{ij} = 0$ or $Y_{ij} = 1$.
Let's calculate the ratio of the two respective probabilities.

[We'll use $P_\theta(Y = y) = \exp\{\theta^t g(y)\}/\kappa(\theta)$.]

Conditional log-odds of an edge

Notation: For a network y and a pair (i, j) of nodes,

- $y_{ij} = 0$ or 1, depending on whether there is an edge
- y_{ij}^c denotes the status of all pairs in y other than (i, j)
- y_{ij}^+ denotes the same network as y but with $y_{ij} = 1$
- y_{ij}^- denotes the same network as y but with $y_{ij} = 0$

$$\frac{P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)}{P(Y_{ij} = 0 | Y_{ij}^c = y_{ij}^c)} = \frac{\exp\{\theta^t g(y_{ij}^+)\}}{\exp\{\theta^t g(y_{ij}^-)\}}$$

A lot of cancellation happened on the right hand side!

Conditional log-odds of an edge

Notation: For a network y and a pair (i, j) of nodes,

- $y_{ij} = 0$ or 1, depending on whether there is an edge
- y_{ij}^c denotes the status of all pairs in y other than (i, j)
- y_{ij}^+ denotes the same network as y but with $y_{ij} = 1$
- y_{ij}^- denotes the same network as y but with $y_{ij} = 0$

$$\frac{P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)}{P(Y_{ij} = 0 | Y_{ij}^c = y_{ij}^c)} = \exp\{\theta^t [g(y_{ij}^+) - g(y_{ij}^-)]\}$$

A lot of cancellation happened on the right hand side!

Conditional log-odds of an edge

Notation: For a network y and a pair (i, j) of nodes,

- $y_{ij} = 0$ or 1, depending on whether there is an edge
- y_{ij}^c denotes the status of all pairs in y other than (i, j)
- y_{ij}^+ denotes the same network as y but with $y_{ij} = 1$
- y_{ij}^- denotes the same network as y but with $y_{ij} = 0$

$$\log \frac{P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)}{P(Y_{ij} = 0 | Y_{ij}^c = y_{ij}^c)} = \theta^t [g(y_{ij}^+) - g(y_{ij}^-)]$$

Conditional log-odds of an edge

Notation: For a network y and a pair (i, j) of nodes,

- $\Delta(g(y))_{ij}$ denotes the vector of change statistics,

$$\Delta(g(y))_{ij} = g(y_{ij}^{+}) - g(y_{ij}^{-}).$$

So $\Delta(g(y))_{ij}$ is the conditional log-odds of edge (i, j).

$$\log \frac{P(Y_{ij} = 1 | Y_{ij}^{c} = y_{ij}^{c})}{P(Y_{ij} = 0 | Y_{ij}^{c} = y_{ij}^{c})} = \theta^{t} \Delta(g(y))_{ij}$$

Simulating random networks

Goal:

Simulate random network(s) Y from an ERGM.

Note: There is no model, only a model class, unless we have a specific parameter vector θ; we'll need to fix an θ somehow.

Simulating random networks

Goal:

Simulate random network(s) Y from an ERGM.

Note: There is no model, only a model class, unless we have a specific parameter vector θ; we'll need to fix an θ somehow.

We'll discuss one way to achieve the goal, called a Metropolis algorithm.

The Metropolis algorithm is one of a broad class of algorithms called Markov Chain Monte Carlo (MCMC) algorithms.

Obtaining samples via Markov Chain Monte Carlo

Whence the name MCMC?

- Markov Chain: A sequence $Y_1, Y_2, \ldots$ where the $(i+1)$th network is randomly generated based on the ith network.
- Monte Carlo: The computational implementation of the "randomly generated" part.

MCMC Idea for simulating networks:

Simulate a carefully designed Markov chain on the sample space of networks for a while. When we stop it, we'll have our random network.

Metropolis algorithm

- First, select a pair of nodes at random, say (i, j).

Metropolis algorithm

- First, select a pair of nodes at random, say (i, j).
- Calculate the ratio

$$\begin{aligned} \pi &= \frac{P(Y_{ij} \text{ changes} | Y_{ij}^c = y_{ij}^c)}{P(Y_{ij} \text{ does not change} | Y_{ij}^c = y_{ij}^c)} \\ &= \exp\{\pm\theta_0^t \Delta(g(y))_{ij}\} \end{aligned}$$

Metropolis algorithm

- First, select a pair of nodes at random, say (i, j).
- Calculate the ratio

$$\begin{aligned} \pi &= \frac{P(Y_{ij} \text{ changes} | Y_{ij}^c = y_{ij}^c)}{P(Y_{ij} \text{ does not change} | Y_{ij}^c = y_{ij}^c)} \\ &= \exp\{\pm\theta_0^t \Delta(g(y))_{ij}\} \end{aligned}$$

- Accept the change of Y_{ij} with probability $\min\{1, \pi\}$ (i.e., always make the change if $\pi \geq 1$; otherwise, only make the change sometimes).

Metropolis algorithm

- First, select a pair of nodes at random, say (i, j).
- Calculate the ratio

$$\begin{aligned} \pi &= \frac{P(Y_{ij} \text{ changes} | Y_{ij}^c = y_{ij}^c)}{P(Y_{ij} \text{ does not change} | Y_{ij}^c = y_{ij}^c)} \\ &= \exp\{\pm\theta_0^t \Delta(g(y))_{ij}\} \end{aligned}$$

- Accept the change of Y_{ij} with probability $\min\{1, \pi\}$ (i.e., always make the change if $\pi \geq 1$; otherwise, only make the change sometimes).

Note: The values of $g(y_{ij}^+)$ and $g(y_{ij}^-)$ are never needed; only the difference $\Delta(g(y))_{ij}$ *matters.*

How should θ_0 be chosen?

- 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.

How should θ_0 be chosen?

- 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.

How should θ_0 be chosen?

- 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.
- A choice that sometimes works is the MPLE (maximum pseudolikelihood estimate)

Outline

Maximum Pseudolikelihood: Intuition

- What if we assume that there is no dependence (or very weak dependence) among the Y_{ij}?

Maximum Pseudolikelihood: Intuition

- What if we assume that there is no dependence (or very weak dependence) among the Y_{ij}?
- In other words, what if we approximate the marginal $P(Y_{ij} = 1)$ by the conditional $P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)$?

Maximum Pseudolikelihood: Intuition

- What if we assume that there is no dependence (or very weak dependence) among the Y_{ij}?
- In other words, what if we approximate the marginal $P(Y_{ij} = 1)$ by the conditional $P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)$?
- Then the Y_{ij} are independent with

$$\log \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} = \theta^t \Delta(g(y^{\text{obs}}))_{ij},$$

so we obtain an estimate of θ using straightforward logistic regression.

Maximum Pseudolikelihood: Intuition

- What if we assume that there is no dependence (or very weak dependence) among the Y_{ij}?
- In other words, what if we approximate the marginal $P(Y_{ij} = 1)$ by the conditional $P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)$?
- Then the Y_{ij} are independent with

$$\log \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} = \theta^t \Delta(g(y^{\text{obs}}))_{ij},$$

  so we obtain an estimate of θ using straightforward logistic regression.
- Result: The maximum pseudolikelihood estimate.

Maximum Pseudolikelihood: Intuition

- What if we assume that there is no dependence (or very weak dependence) among the Y_{ij}?
- In other words, what if we approximate the marginal $P(Y_{ij} = 1)$ by the conditional $P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)$?
- Then the Y_{ij} are independent with

$$\log \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} = \theta^t \Delta(g(y^{\text{obs}}))_{ij},$$

  so we obtain an estimate of θ using straightforward logistic regression.
- Result: The maximum pseudolikelihood estimate.
- For independence models, MPLE = MLE!

Warnings about MPLE

Unfortunately, little is known about the quality of MPL estimates in general, but we do know some ways in which they can be misleading.

- If the model is bad, you'll get nice-looking MPLE results that do not reveal the problem.
- If the model is good, in many cases the MPLE looks "close" to the MLE; however, "close" can be deceiving, since small changes in θ can sometimes lead to large differences in the behavior of randomly generated networks.

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:

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

- 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 $(0 - M = 3\binom{g}{3})$

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

– 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

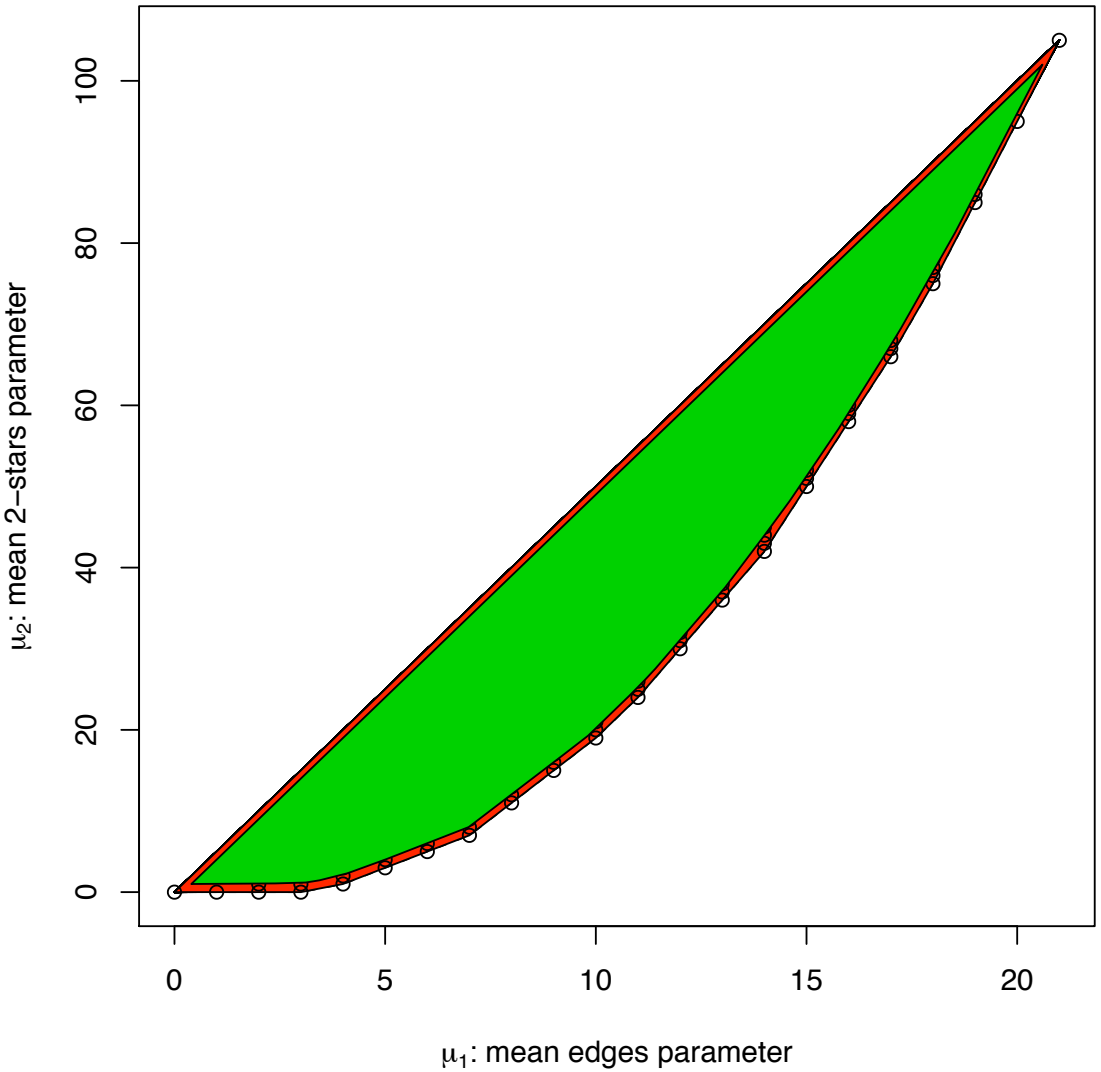
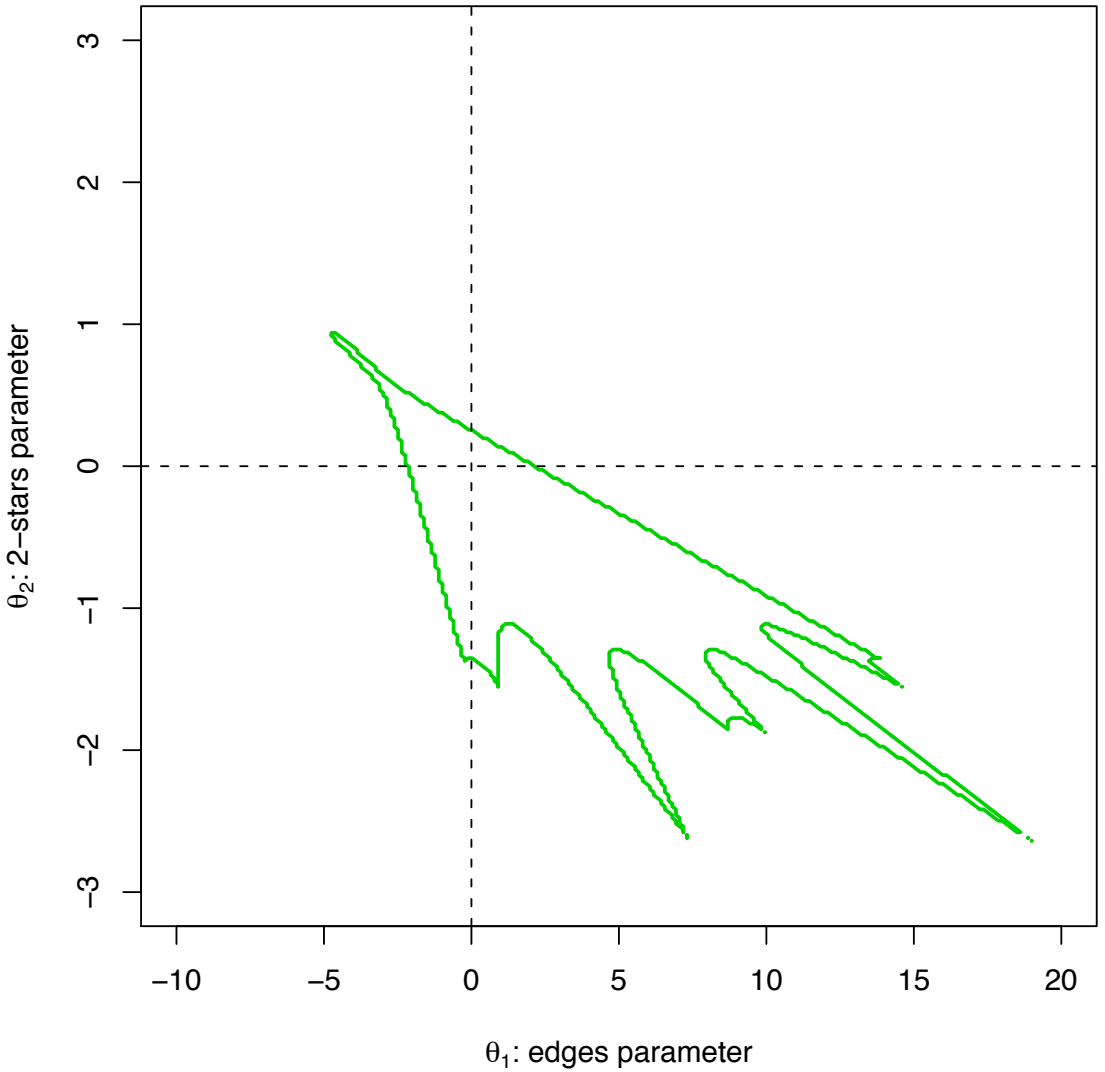


Figure 5: Regions of the parameter space of θ



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 $\eta : \Lambda \longrightarrow \Lambda^{(2)} \times \mathrm{int}(\mathrm{C}^{(1)})$ defined by

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

is a *mixed* parametrization 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.

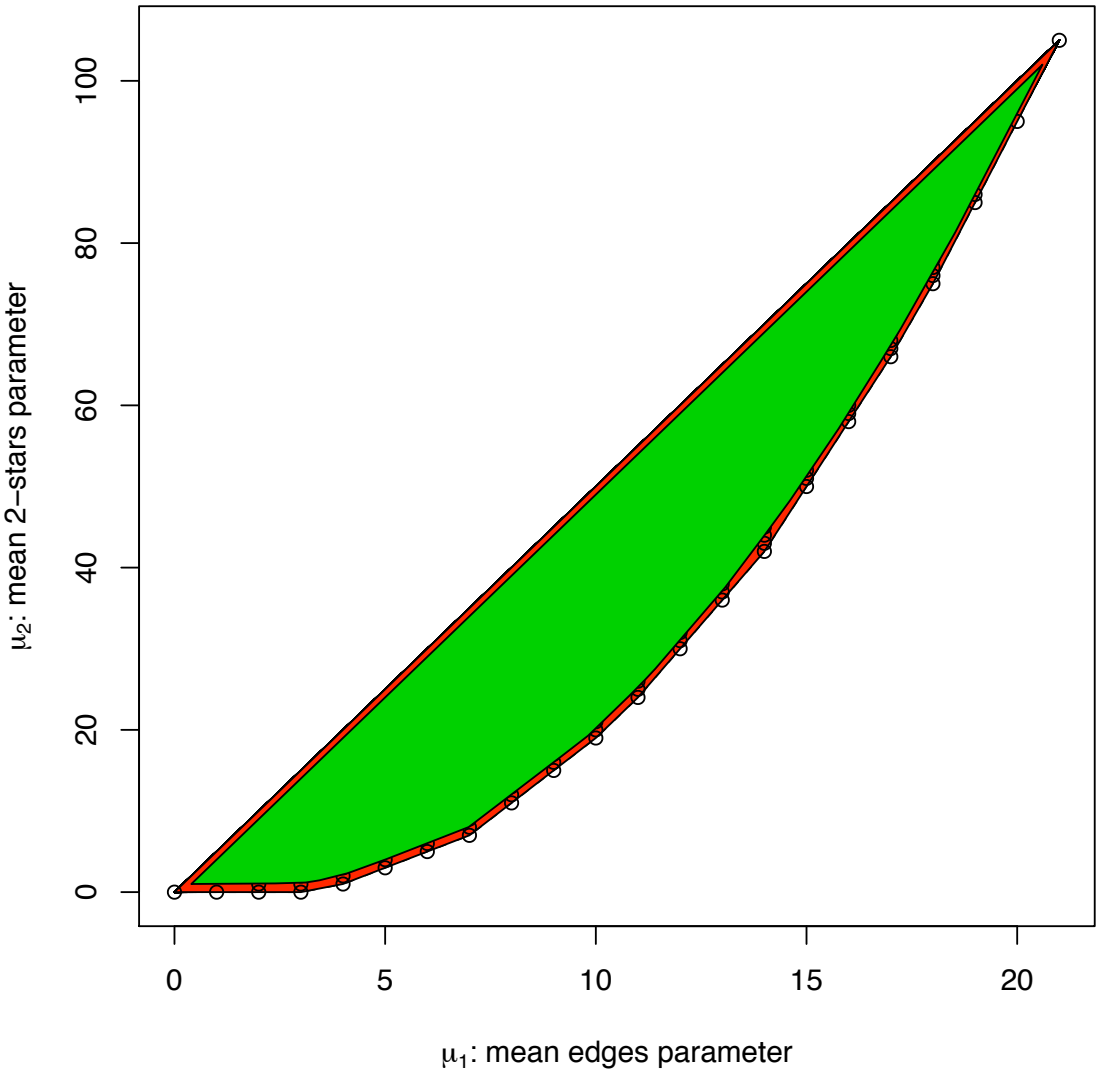
Degeneracy in the mean value parametrization

- **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 $\mathrm{Kullback} - \mathrm{Leibler\ divergence}(\eta_0; \lambda e) \to \infty$ 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.

- 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 $\mu(\psi) \to \mathrm{bd}(\mathrm{C})$, $M(\psi) \to \infty$
  – There exists $y \in \mathcal{Y}$ with

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

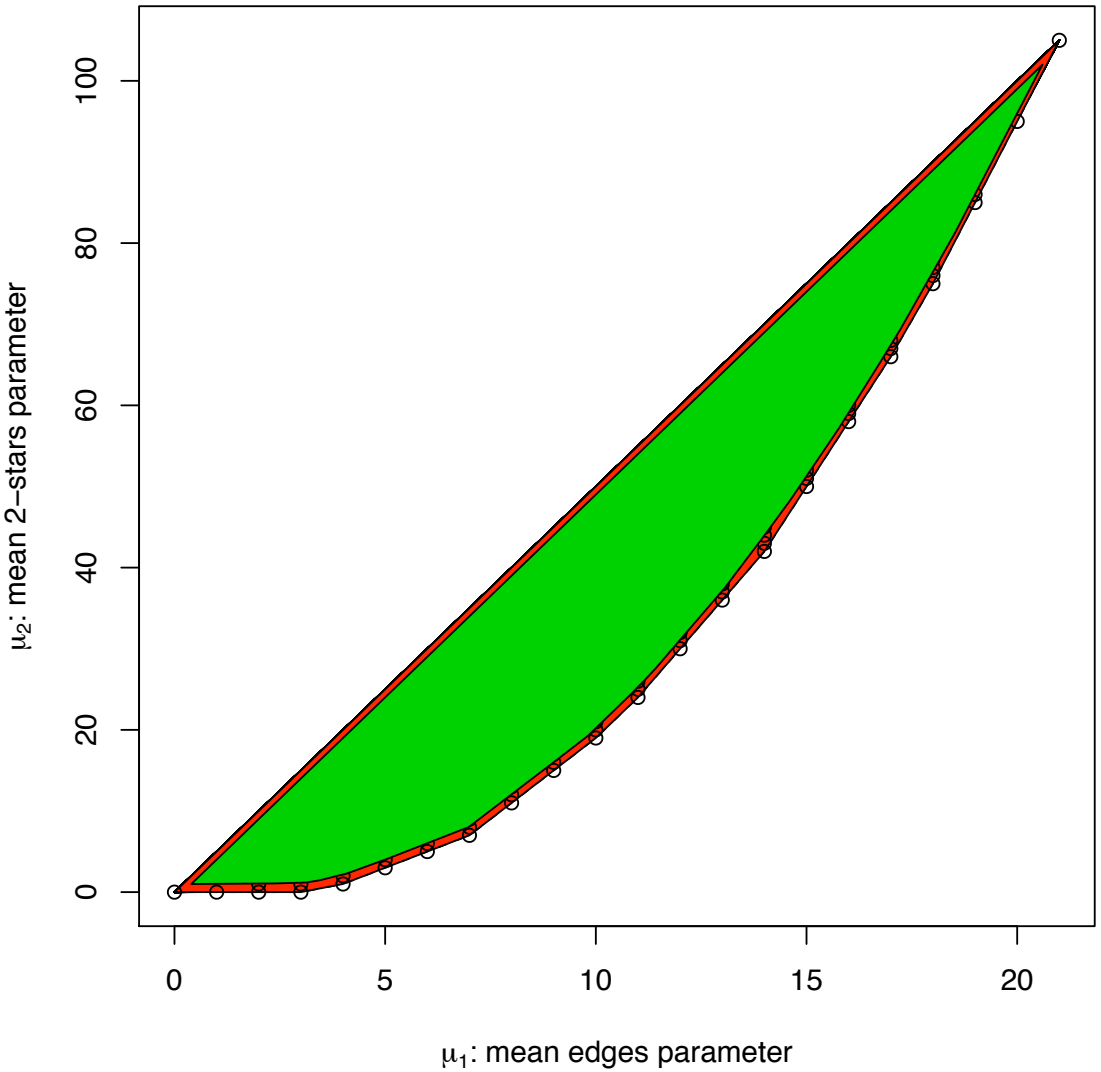
  – If ψ is near-degenerate then $M(\psi)$ is large and the MCMC will mix very slowly.

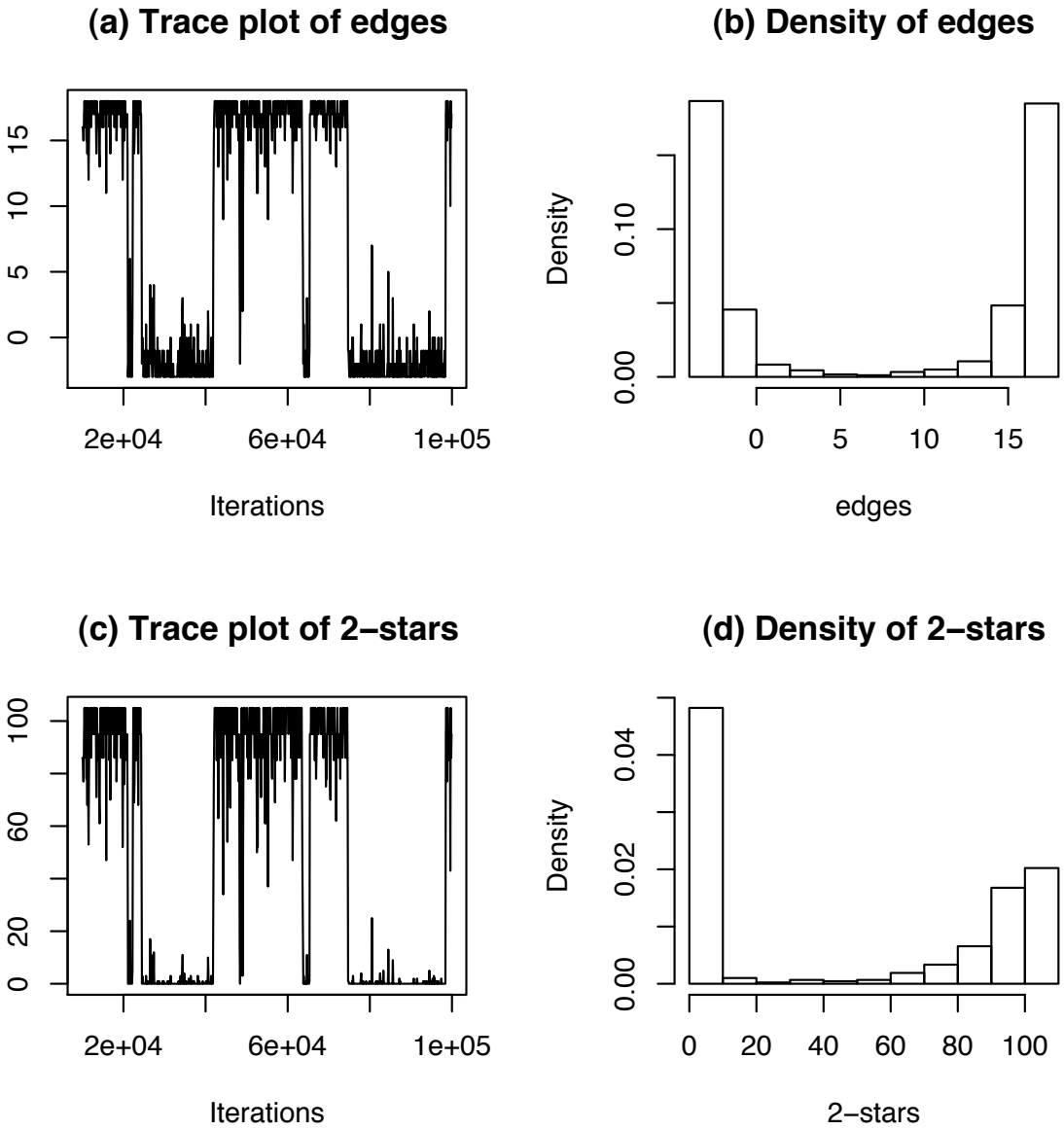
Example of degeneracy of the $2-$star model

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

- $M(\eta) = \max\{|\eta_1|, \eta_1 + 2(g-2)\eta_2\}$ 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
0
5
10
15
2e+04
6e+04
1e+05
Iterations
(b) Density of edges
Density
0.00
0.10
0
5
10
15
edges
(c) Trace plot of 2−stars
0
20
60
100
2e+04
6e+04
1e+05
Iterations
(d) Density of 2−stars
Density
0.00
0.02
0.04
0
20
40
60
80
100
2−stars

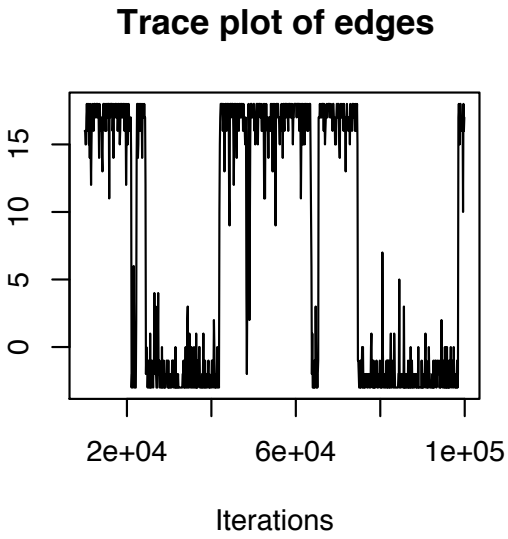
Estimation within the mean value parametrization

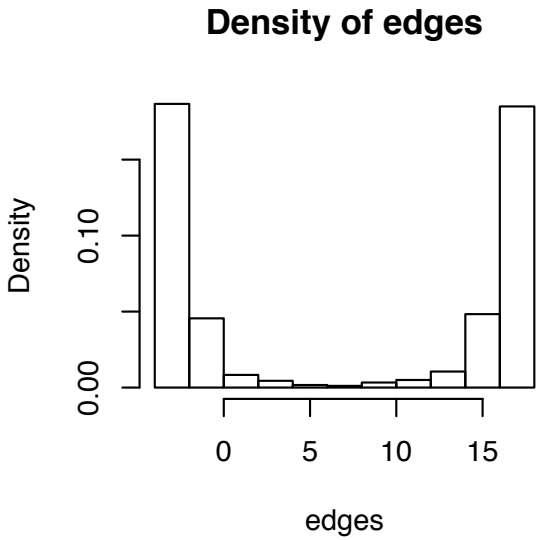
– If $Z(y_{obs}) \in \mathrm{int}(\mathrm{C})$, the MLE of μ is $Z(y_{obs})$.
– If $Z(y_{obs}) \notin \mathrm{int}(\mathrm{C})$ the MLE of μ does not exist.
– The MLE $\hat{\mu}$ is unbiased and has minimum variance:

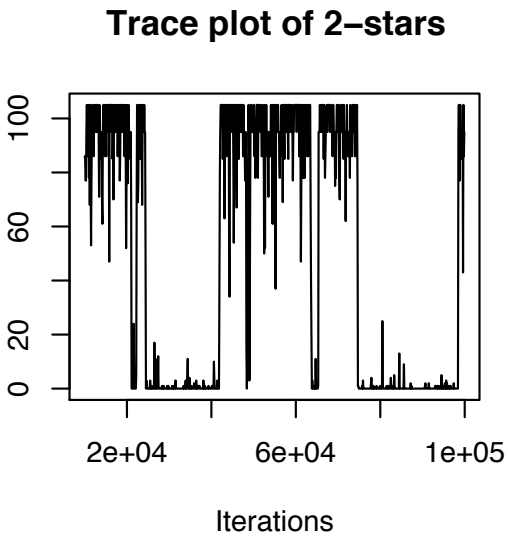
$$\mathbf{E}_\eta(\hat{\mu}) = \mathbf{E}_\eta\left[Z(Y)\right] = \mu(\eta) = \left[\frac{\partial \log c(\eta)}{\partial \eta_i}\right](\eta)$$

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

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


Trace plot of edges
0
5
10
15
2e+04
6e+04
1e+05
Iterations


Density of edges
Density
0.00
0.10
0
5
10
15
edges


Trace plot of 2−stars
0
20
60
100
2e+04
6e+04
1e+05
Iterations

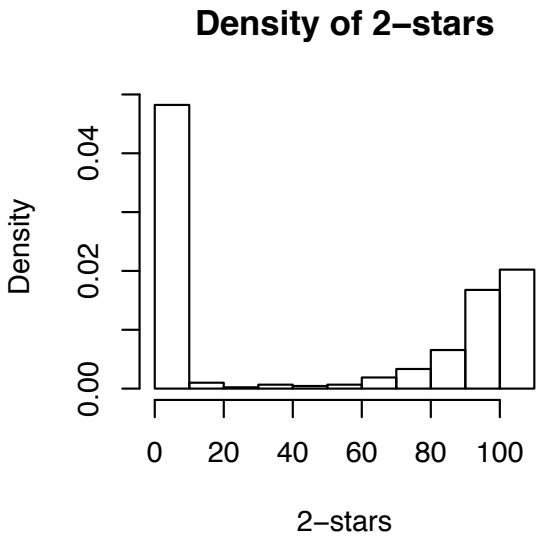
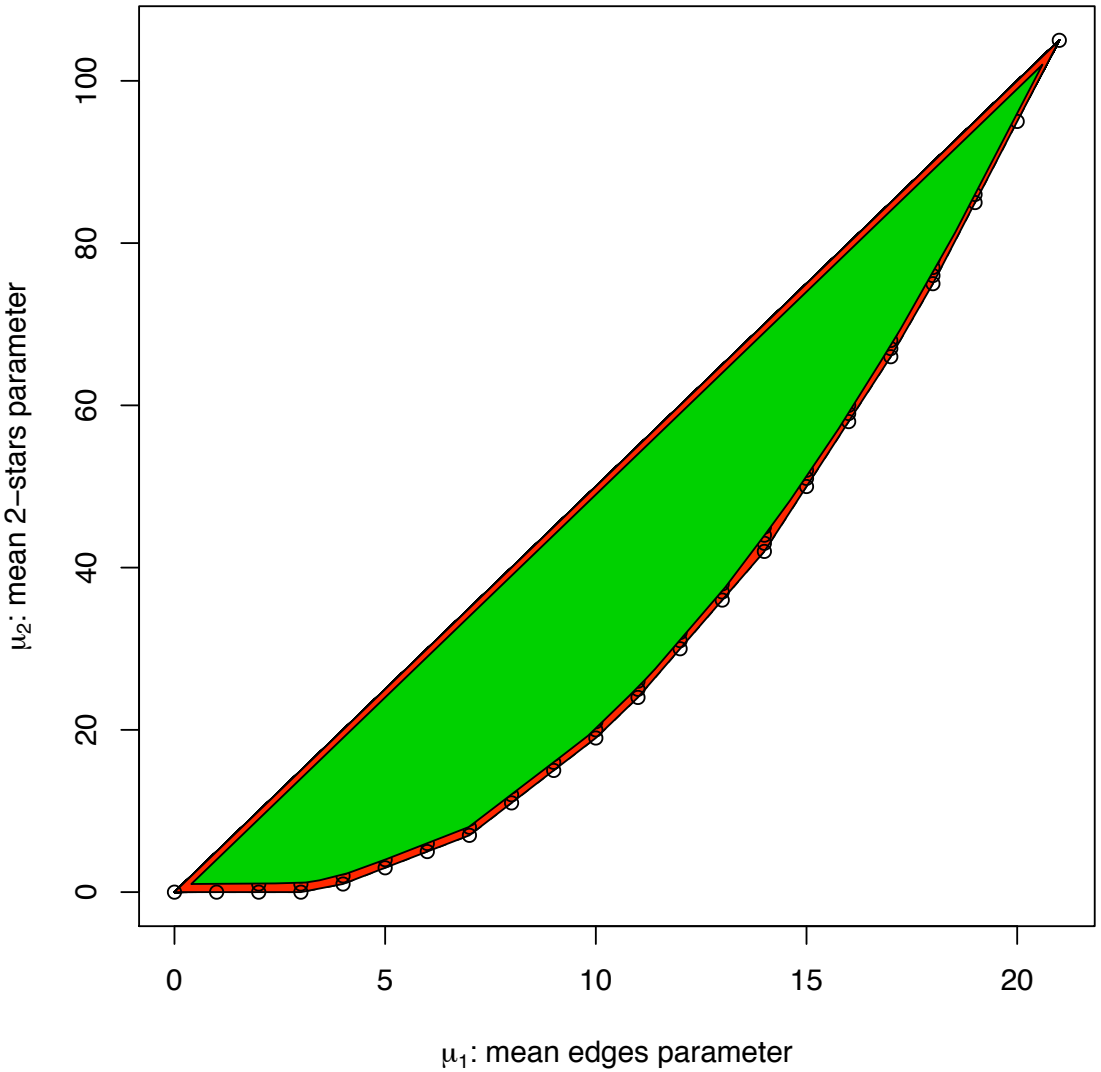

Density of 2−stars
Density
0.00
0.02
0.04
0
20
40
60
80
100
2−stars

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.

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

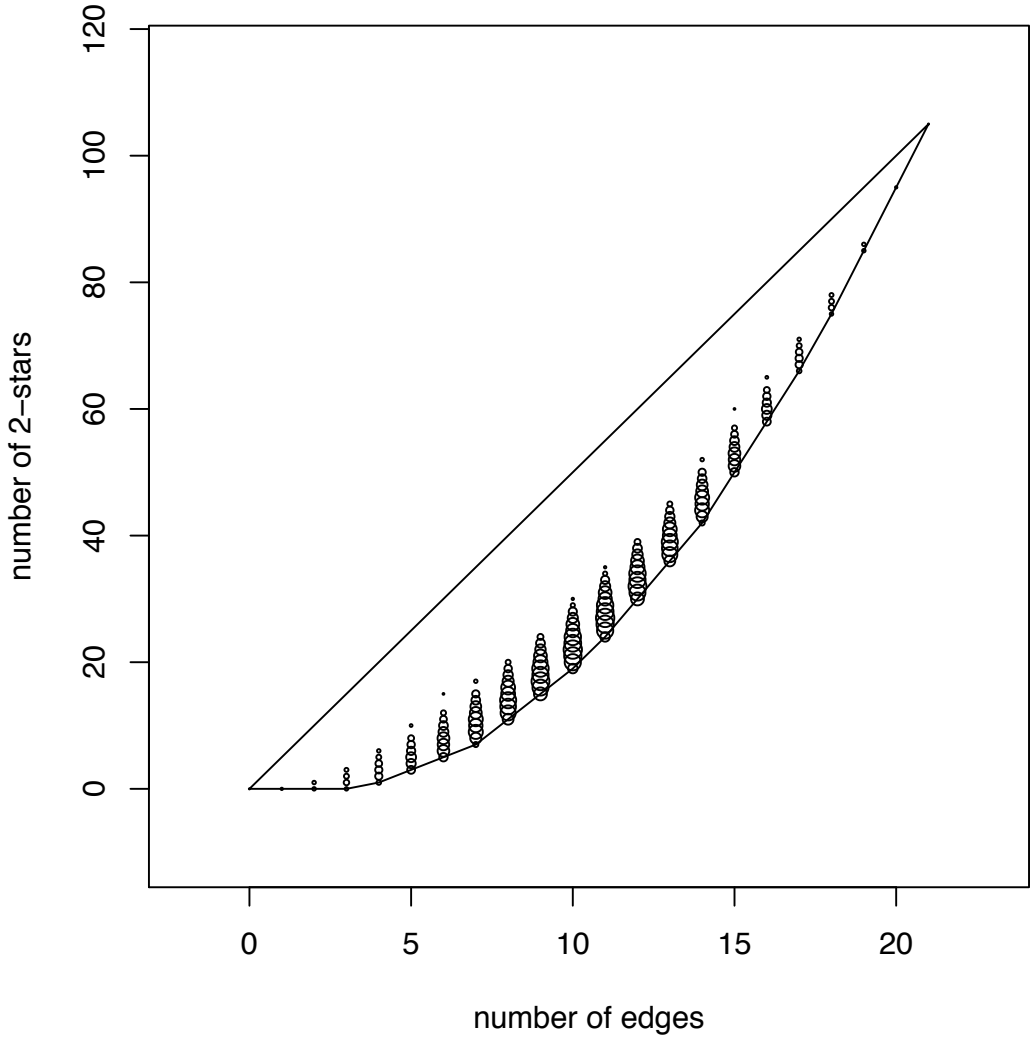


Figure 1: Enumeration of sufficient statistics for graphs with 7 nodes. The circles are centered on the possible values and the area of the circle is proportional to the number of graphs with that value of the sufficient statistic. There are a total of 2,097,152 graphs.

A Bias-corrected Pseudo-likelihood Estimator

The penalized pseudo-likelihood

$$\ell_{BP}(\eta; y) \equiv \ell_P(\eta; y) + \frac{1}{2} \log |I(\eta)| \tag{2}$$

where $I(\eta)$ denotes the expected Fisher information matrix for the formal logistic model underlying the pseudo-likelihood evaluated at η.

Motivated by Firth (1993) as a general approach to reducing the asymptotic bias of MLEs

We refer to the estimator that maximizes $\ell_{BP}(\eta; y_{obs})$ as the *maximum bias-corrected pseudo-likelihood estimator* (MBLE).

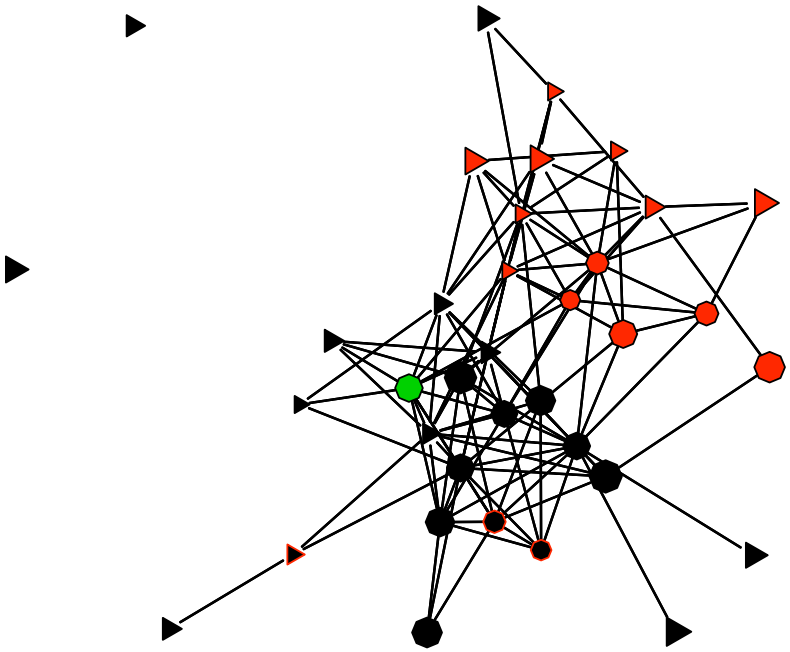
Simulation study of MLE, MPLE and MBLE

The general structure of the simulation study is as follows:

- Begin with the MLE model fit of interest for a given network.
- Simulate networks from this model fit.
- Fit the model to each sampled network using each method under comparison.
- Evaluate the performance of each estimation procedure in recovering the known true parameter values, along with appropriate measures of uncertainty.

Introduction to Law Firm Collaboration Example

From the Emmanuel Lazega's study of a Corporate Law Firm:



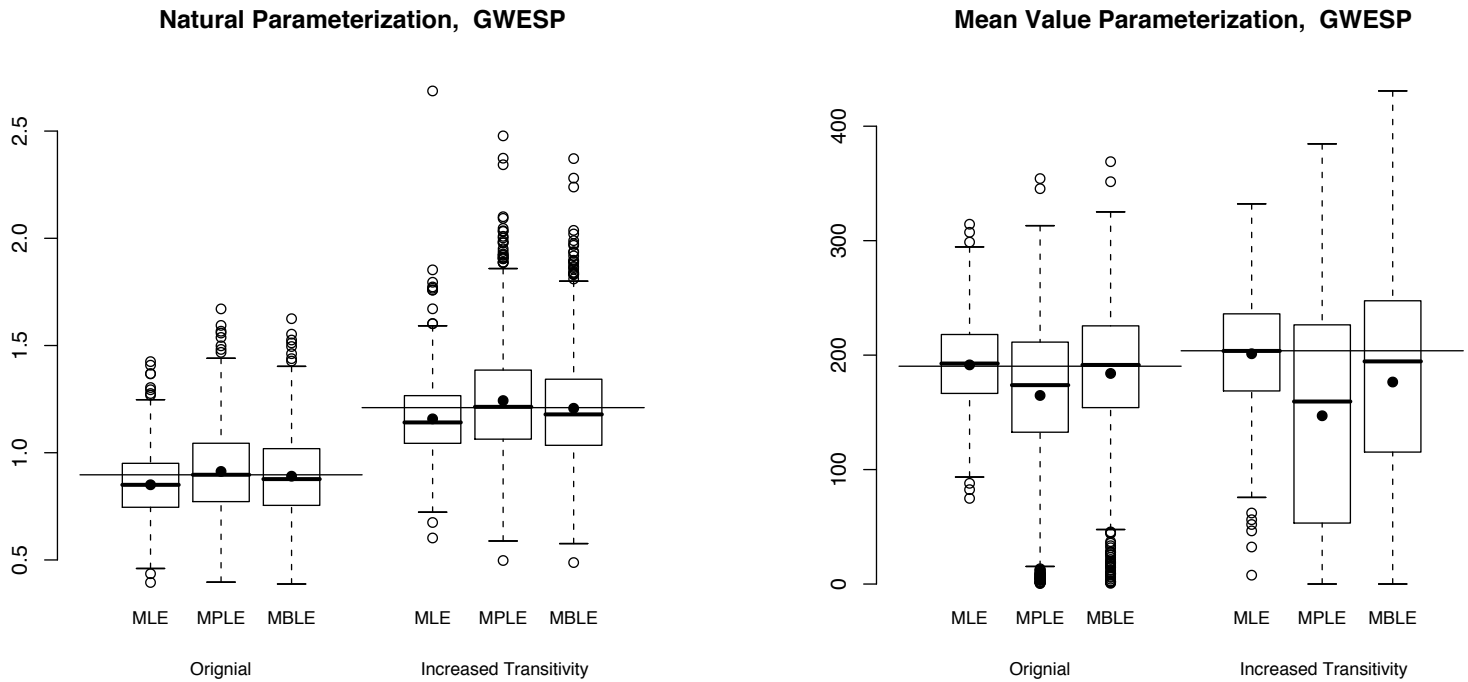
- Each partner asked to identify the others with whom (s)he collaborated.
- Seniority, Sex, Practice (corporate or litigation) and Office (3 locations) available for all 36 partners.

Table 1: Natural and mean value model parameters for Original model for Lazega data, and for model with increased transitivity.

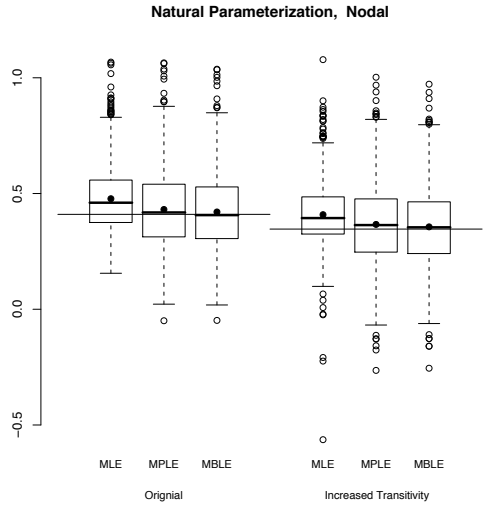
Parameter	Natural Parameterization		Mean Value Parameterization	
	Original	Increased Transitivity	Original	Increased Transitivity
Structural				
edges	−6.506	−6.962	115.00	115.00
GWESP	0.897	1.210	190.31	203.79
Nodal				
seniority	0.853	0.779	130.19	130.19
practice	0.410	0.346	129.00	129.00
Homophily				
practice	0.759	0.756	72.00	72.00
gender	0.702	0.662	99.00	99.00
office	1.145	1.081	85.00	85.00

Figure 1: Boxplots of the distribution of the MLE, the MPLE and the MBLE of the geometrically weighted edgewise shared partner statistic (GWESP), differential activity by practice statistic (Nodal), and homophily on practice statistic (Homophily) under the natural and mean value parameterization for 1000 samples of the original Lazega network and 1000 samples of the Lazega network with increased transitivity

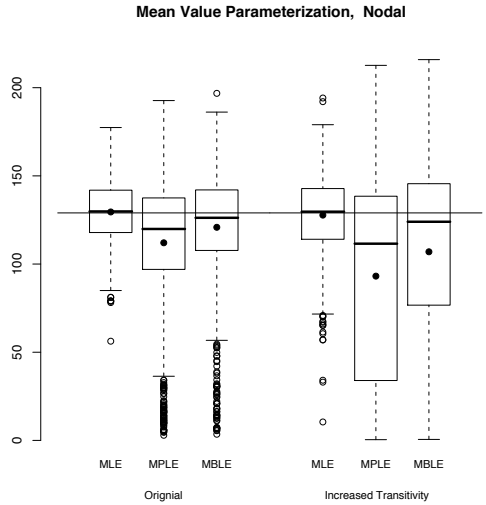
(a) (b)



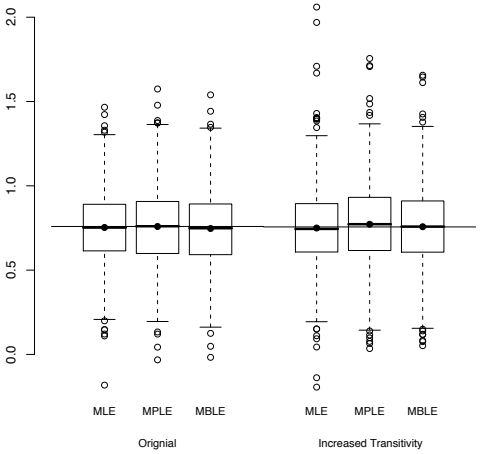
(c)


Natural Parameterization, Nodal
1.0
0.5
0.0
−0.5
MLE
MPLE
MBLE
MLE
MPLE
MBLE
Orignial
Increased Transitivity

(d)


Mean Value Parameterization, Nodal
200
150
100
50
0
MLE
MPLE
MBLE
MLE
MPLE
MBLE
Orignial
Increased Transitivity

(e)


Natural Parameterization, Homophily
2.0
1.5
1.0
0.5
0.0
MLE
MPLE
MBLE
MLE
MPLE
MBLE
Orignial
Increased Transitivity

(f)

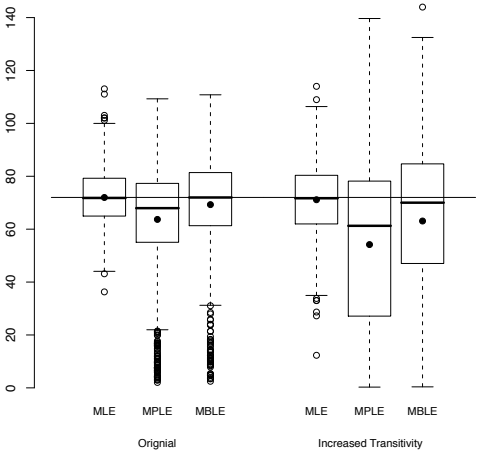

Mean Value Parameterization, Homophily
140
120
100
80
60
40
20
0
MLE
MPLE
MBLE
MLE
MPLE
MBLE
Orignial
Increased Transitivity

Table 2: Relative efficiency of the MPLE, and the MBLE with respect to the MLE

	Natural Parameterization						Mean Value Parameterization					
	Original			Increased Transitivity			Original			Increased Transitivity		
Parameter	MLE	MPLE	MBLE	MLE	MPLE	MBLE	MLE	MPLE	MBLE	MLE	MPLE	MBLE
Structural												
edges	1	0.80	0.94	1	0.66	0.80	1	0.21	0.29	1	0.15	0.20
GWESP	1	0.64	0.68	1	0.50	0.55	1	0.28	0.37	1	0.19	0.24
Nodal												
seniority	1	0.87	0.92	1	0.78	0.83	1	0.22	0.30	1	0.17	0.22
practice	1	0.91	0.96	1	0.72	0.77	1	0.19	0.27	1	0.12	0.16
Homophily												
practice	1	0.91	0.96	1	0.94	1.01	1	0.23	0.32	1	0.15	0.19
gender	1	0.81	0.91	1	0.78	0.86	1	0.23	0.31	1	0.17	0.22
office	1	0.92	1.00	1	0.79	0.87	1	0.23	0.32	1	0.15	0.20

Table 3: Coverage rates of nominal 95% confidence intervals for the MLE, the MPLE, and the MBLE of model parameters for original and increased transitivity models. Nominal confidence intervals are based on the estimated curvature of the model and the t distribution approximation.

	Natural Parameterization						Mean Value Parameterization					
	Original			Increased Transitivity			Original			Increased Transitivity		
Parameter	MLE	MPLE	MBLE	MLE	MPLE	MBLE	MLE	MPLE	MBLE	MLE	MPLE	MBLE
Structural												
edges	94.9	97.5	98.0	96.4	98.2	98.2	93.1	44.9	49.4	85.5	23.8	28.5
GWESP	92.7	74.6	74.1	94.2	78.8	77.6	91.4	56.7	62.7	85.9	31.3	36.6
Nodal												
seniority	94.4	97.8	98.0	95.4	98.4	98.7	91.6	45.5	49.0	84.4	22.8	27.6
practice	94.0	98.1	98.6	95.5	98.4	98.8	93.2	51.0	57.9	89.9	35.9	39.3
Homophily												
practice	94.8	98.1	98.1	94.6	97.9	98.0	92.6	52.0	57.1	89.7	31.1	37.3
gender	95.8	98.7	98.8	95.3	98.1	98.8	92.0	46.5	51.6	84.8	22.7	28.5
office	94.2	98.1	98.4	95.1	98.2	98.4	92.5	50.2	54.4	87.8	27.0	32.3

Summary

This is a framework to assess estimators for (ERG) models.

Key features:

- The use of the mean-value parametrization space as an alternate metric space to assess model fit.
- The adaptation of a simulation study to the specific circumstances of interest to the researcher: e.g. network size, composition, dependency structure.
- It assesses the efficiency of point estimation via mean-squared error in the different parameter spaces.
- It assesses the performance of measures of uncertainty and hypothesis testing via actual and nominal interval coverage rates.
- It provides methodology to modify the dependence structure of a model in a known way, for example, changing one aspect while holding the other aspects fixed.
- It enables the assessment of performance of estimators to be to alternative specifications of the underlying model.

Case study:

- MLE superior to MPLE and MBLE for structural and covariate effects.
  - due to the dependence between the GWESP estimates and others
  - Greater variability in the GWESP results translates to broad CI
  - GWESP standard errors are underestimated resulting in too narrow CI
- Inference based on the MPLE is suspect
  - Tests for structural parameters tend to be liberal
  - Tests for nodal and dyadic attributes conservative
- MLE drastically superior on the mean value scale (30% of MSE of MP(B)LE)
  - MPLE nominal 95% CI coverage is 50%.
  - Gets worse as dependence increases.
- MBLE
  - Smallest bias for the natural parameter estimates.
  - MBLE consistently out-performs the MPLE
    (for both natural and mean-value parameters)

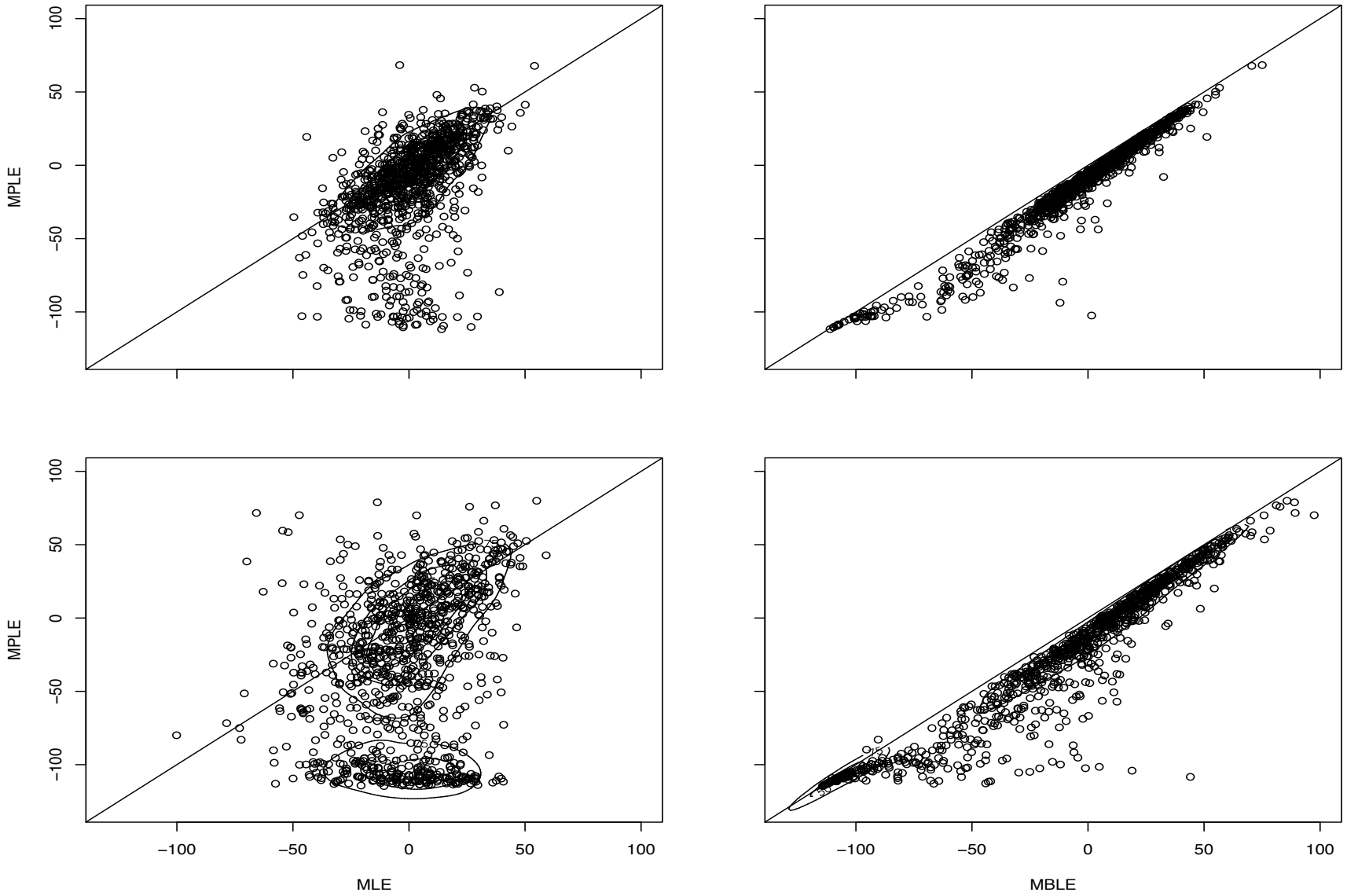


Figure 2: Comparison of error in mean value parameter estimates for edges in original (top) and increased transitivity (bottom) models.