Computational Issues with ERGM:
Pseudo-likelihood for constrained degree models

Mark S. Handcock

University of California - Los Angeles

MURI-UCI June 3, 2011

For details, see:

- van Duijn, Marijtje A. J., Gile, Krista J. and Handcock, Mark 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]
- Gile, Krista J. and Handcock, Mark S. (2011). Network Model-Assisted Inference from Respondent-Driven Sampling Data, UCLA working paper.

[1]Research supported by 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 $\mathscr{Y}$ 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:

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

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
- $\kappa(\eta, \mathscr{Y})$ distribution normalizing constant

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

Statistical Inference for η

Base inference on the loglikelihood function,

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

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

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

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

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

Maximum Pseudolikelihood

Consider the conditional formulation of the ERGM:

$$\text{logit}[P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, \eta)] = \eta \cdot \delta(y_{ij}^c) \qquad y \in \mathscr{Y} \tag{1}$$

where $\delta(y_{ij}^c) = g(y_{ij}^+, \mathbf{z}) - g(y_{ij}^-, \mathbf{z})$, the change in $g(y, \mathbf{z})$ when y_{ij} changes from 0 to 1 while the remainder of the network remains y_{ij}^c The log-pseudolikelihood function is then

$$\ell_P(\eta; y) = \sum \log[P(Y_{ij} = y_{ij} | Y_{ij}^c = y_{ij}^c)]$$

The pseudo-likelihood for the model is:

$$\ell_P(\eta; y) \equiv \eta \cdot \sum_{ij} \delta(y_{ij}^c, \mathbf{z}) y_{ij} - \sum_{ij} \log\left[1 + \exp(\eta \cdot \delta(y_{ij}^c, \mathbf{z}))\right]. \tag{2}$$

This is the standard form of pseudo-likelihood, which we refer to as the *dyadic pseudo-likelihood*.

Result: The *maximum pseudolikelihood estimate* is then the value that maximizes $\ell_P(\eta; y)$ as a function of η..

Models Conditional on Degree and Covariate Sequences

Let the n-vector $\mathbf{z}$, represent a vector of covariates and $\mathbf{d}_i = \sum_j y_{ij}$ the nodal *degree*

Here focus on $\mathscr{Y} \equiv \mathscr{Y}(\mathbf{z}, \mathbf{d})$ consisting of all binary networks consistent with $\mathbf{d}$ and $\mathbf{z}$.

This standard form of pseudo-likelihood is inappropriate for the ERGM as it does not take into account the network space $\mathscr{Y}(\mathbf{z}, \mathbf{d})$.

This is because $P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c, \eta)]$ is either 1 or 0 depending on if the value $y_{ij} = 1$ produces a joint degree and covariate sequence consistent with $\mathbf{d}$ and $\mathbf{z}$. Hence the dyadic MPLE will usually produce non-sensical results.

Instead of a dyadic pseudo-likelihood we develop a *tetradic pseudo-likelihood*.

Consider the set of all tetrads (four-node subnetworks) of the network. For a given tetrad, consider the (counter-factual) equivalence set of tetrads with the same node set for which the degree and covariate sequences of the corresponding full network are the same as the actual one.

Let y_{ijkl} be the four ties in the tetrad among nodes $i, j, k,$ and l, for which the equivalence set has at least two elements in it. Assume w.l.o.g. that $i, j, k,$ and l, are in decreasing order.

We focus on tetrads where one of the pair has i–j, k–l, but not j–k and the other has i–k, j–k, but not i–j or k–l.

That is a pair with the y_{ij} is toggled from 1 to 0 while y_{jk} is toggled from 0 to 1 in such a way as to retain the the degree and covariate sequences of the corresponding full network. Let y^c_{ijkl} denote the remainder of the full network not determined by the triadic pair.

For this pair:

$$\text{logit}[P(Y_{ijkl} = 1 | Y^c_{ijkl} = y^c_{ijkl}, \eta)] = \eta \cdot \delta(y^c_{ijkl}) \qquad y \in \mathscr{Y}(\mathbf{z}, \mathbf{d}) \tag{3}$$

where $\delta(y^c_{ijkl}) = g(y^+_{ijkl}, \mathbf{z}) - g(y^-_{ijkl}, \mathbf{z})$, the change in $g(y, \mathbf{z})$ when y_{ijkl} changes from 0 to 1 while y_{jk} is toggled from 0 to 1 in such a way as to retain the the degree and covariate sequences of the corresponding full network with y^c_{ijkl} unchanged. The tetradic pseudo-likelihood for the ERGM is:

$$\ell_{PT}(\eta; y) \equiv \eta \cdot \sum_{ijkl} \delta(y^c_{ijkl}, \mathbf{z}) y_{ijkl} - \sum_{ijkl} \log \left[1 + \exp(\eta \cdot \delta(y^c_{ijkl}, \mathbf{z})) \right]. \tag{4}$$

As the number of tetrad pairs is large, we take a large random sample of them ($N = 100000$) and use the sample mean of them instead. This procedure is implemented in the `ergm` `R` package

Performance

While the MPLE is know to be inferior to the MLE for dyadic dependence models (van Duijn, Gile and Handcock 2009) it is equivalent to the MLE for some dyadic independence models.

For the model the network statistic is close to independent on the set of networks with the given degree and covariate sequences.

Hence the maximum tetradic pseudo-likelihood (MTPLE) might be expected to perform well for this model.

In simulations (not shown here) as it appears to be indistinguishable from the MCMC-MLE

The advantages of the tetradic MPLE are that it is computationally stable and fast while being numerically indistinguishable from the MCMC-MLE.

Improvements

This estimator could be improved by adding hexadic configurations to the pseudo-likelihood. These are necessary for sampling algorithms to cover the full network space (Rao and Rao 1996)

However they also lead to more complex algorithms and will be considered in other work.

A Bias-corrected Pseudo-likelihood Estimator

The penalized pseudo-likelihood

$$\ell_{BP}(\eta; y) \equiv \ell_P(\eta; y) + \frac{1}{2} \log |I(\eta)| \tag{5}$$

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).

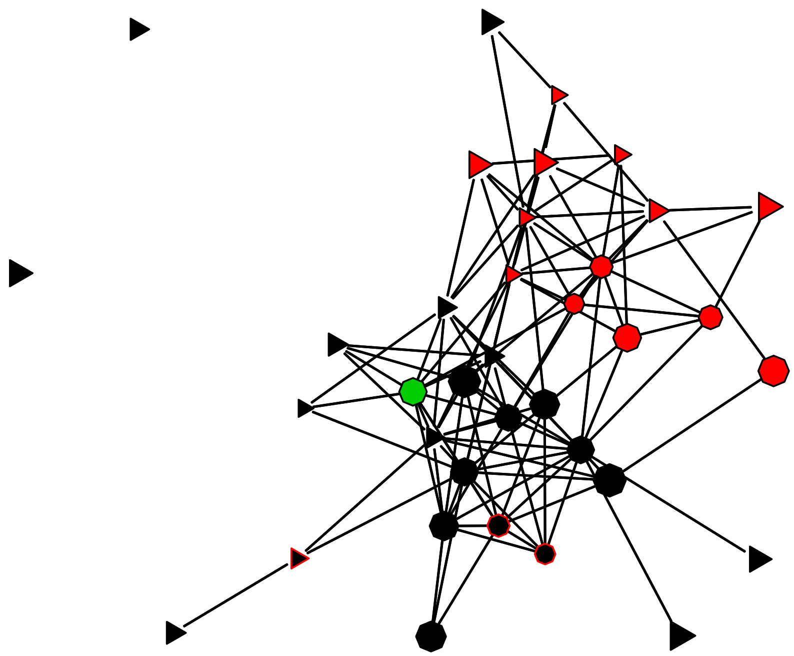
Simulation study of MLE, MPLE and MBLE

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:



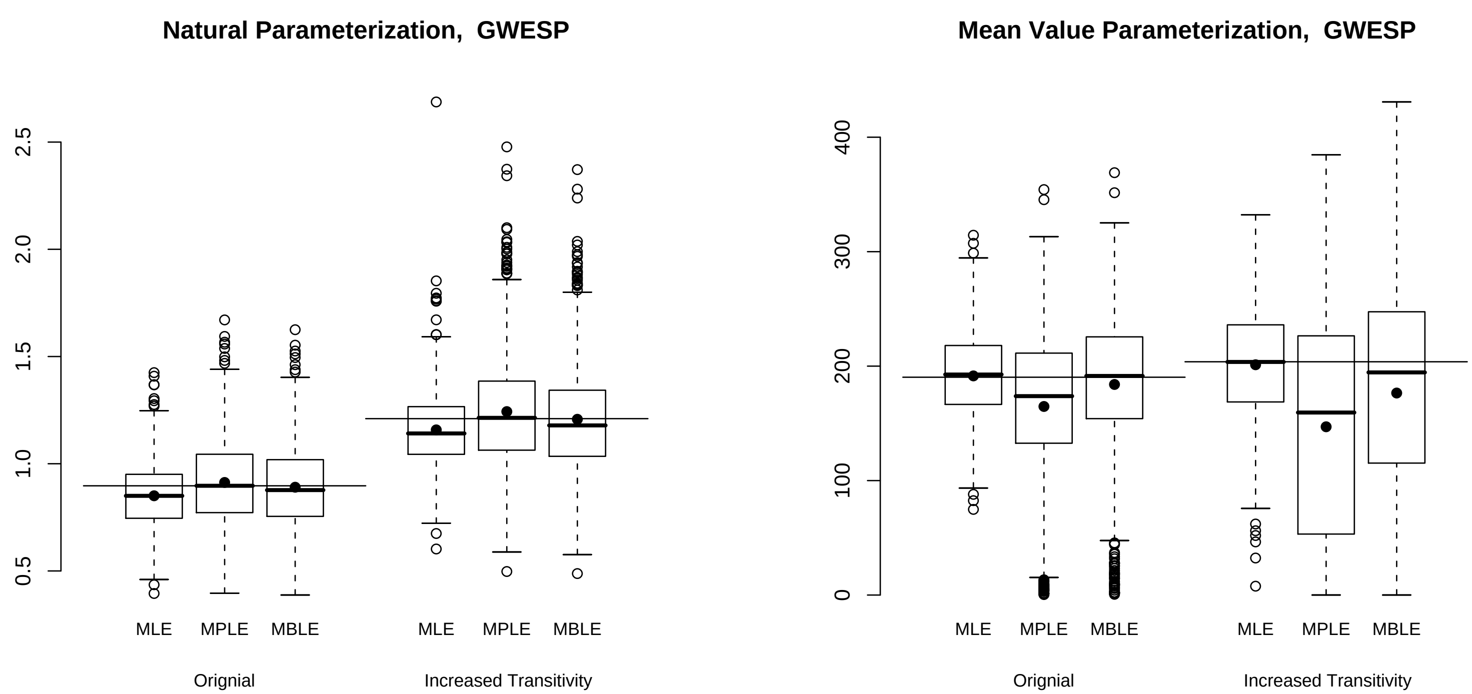
- Each partner asked to identify the others with whom (s)he collaborated.
- Seniority, Sex, Practice (corporate or litigation) and Office (3 locations) available for all 36 partners.

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

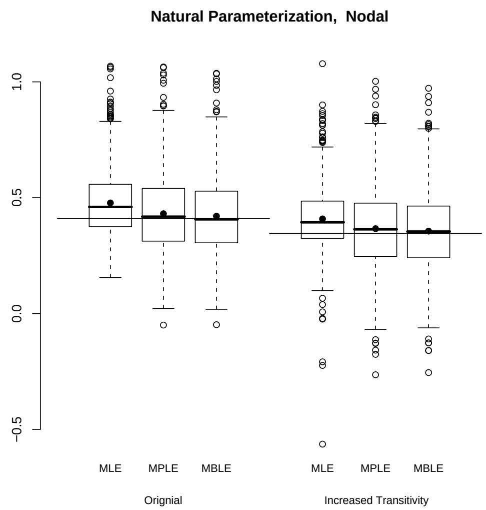
Parameter	Natural Parameterization		Mean Value Parameterization	
	Original	Increased Transitivity	Original	Increased Transitivity
Structural				
edges	−6.506	−6.962	115.00	115.00
GWESP	0.897	1.210	190.31	203.79
Nodal				
seniority	0.853	0.779	130.19	130.19
practice	0.410	0.346	129.00	129.00
Homophily				
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

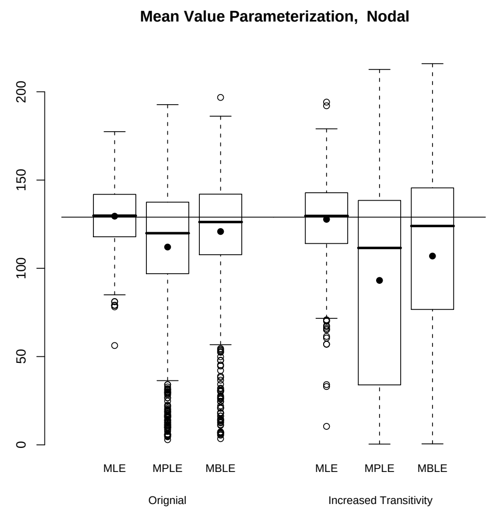
(a) (b)



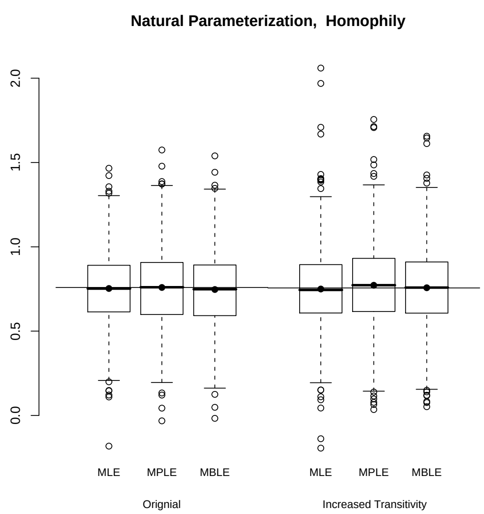
(c)


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)

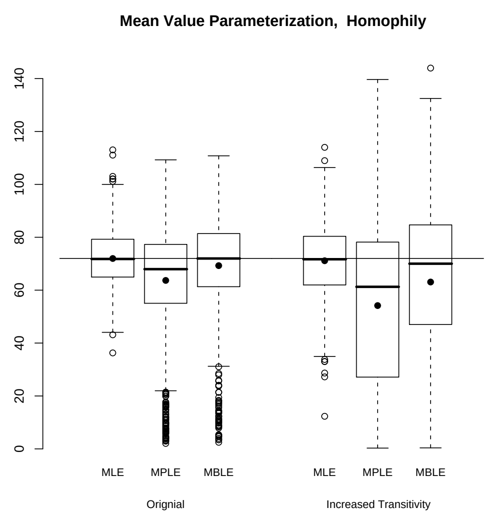

Mean Value Parameterization, Homophily
140
120
100
80
60
40
20
0
MLE
MPLE
MBLE
MLE
MPLE
MBLE
Orignial
Increased Transitivity

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)

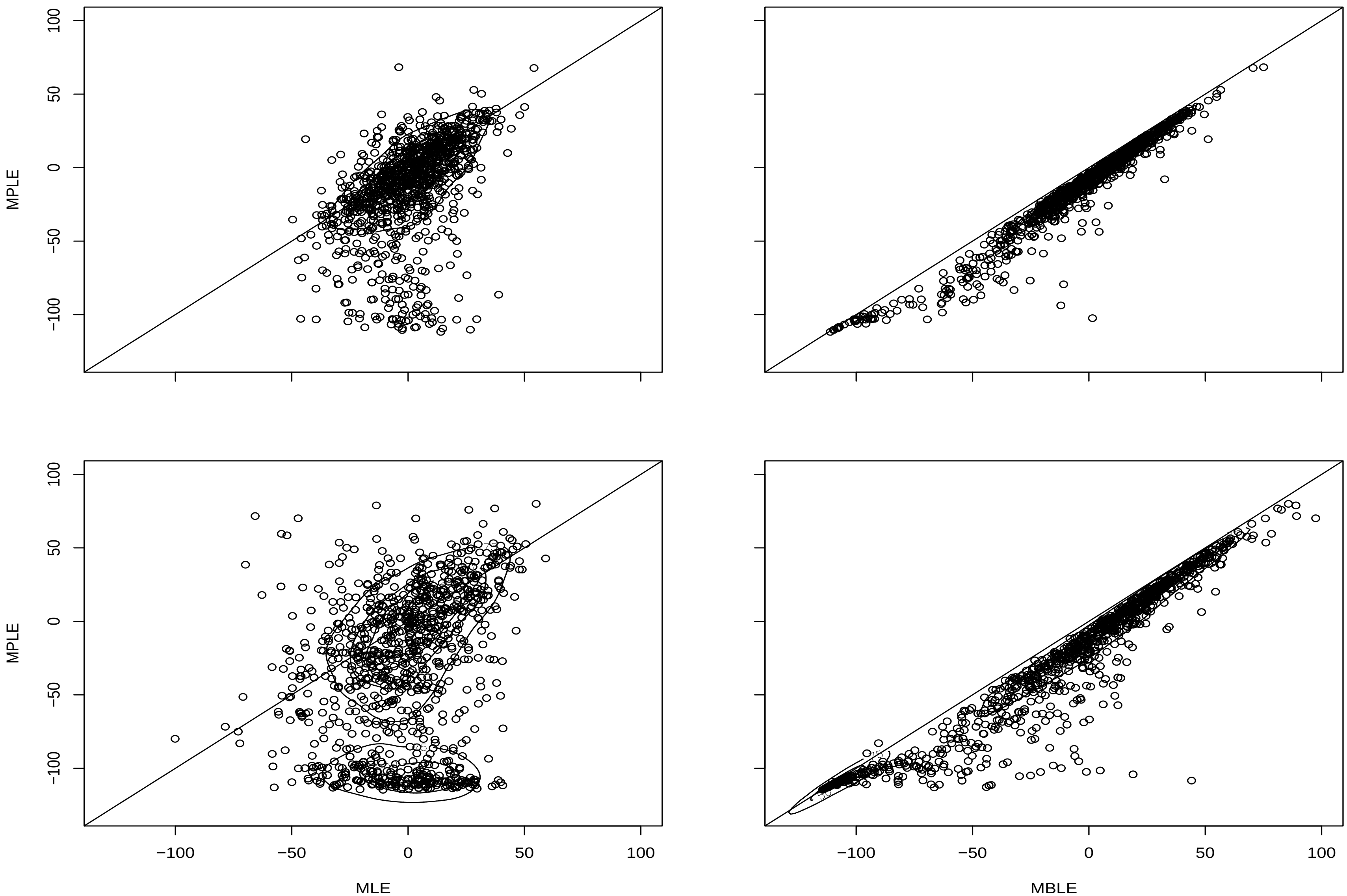


Figure 2: Comparison of error in mean value parameter estimates for edges in original (top) and increased transitivity (bottom) models.