Inference versus estimation in exponential random graph models

Alex Stivala[1] Maksym Byshkin[2] Garry Robins[1]

[1]Melbourne School of Psychological Sciences, The University of Melbourne, Australia
[2]Università della Svizzera italiana, Lugano, Switzerland

INSNA Sunbelt XXXVII, May 30 – June 4, 2017
Beijing, China

Outline

1. ERGM basics.
2. There appears to be a problem with inference on some parameters.
3. What the problem is not.
4. New algorithms let us see the problem more clearly with larger networks.
5. What the problem is, with help from a classic ERGM paper.
6. What might we do about it?

Exponential random graph models (ERGMs)

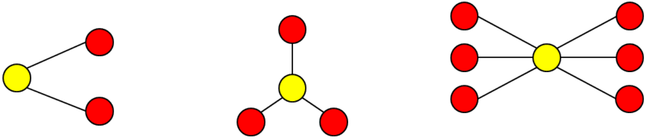
$$\Pr(X = x) = \frac{1}{\kappa} \exp\left(\sum_{A} \theta_A z_A(x)\right)$$

where

- $X = [X_{ij}]$ is a 0-1 matrix of random tie variables,
- x is a realization of X,
- A is a subgraph configuration,
- $z_A(x)$ is the network statistic for configuration A,
- θ_A is a model parameter corresponding to configuration A,
- κ is a normalizing constant to ensure a proper distribution.

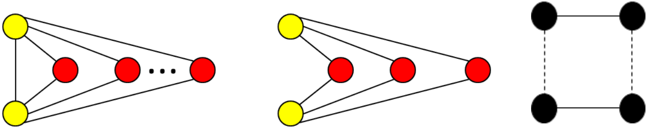
Model configurations — structural

k-stars: useful for capturing degree distribution



Two-star Three-star *k*-star

k-triangles (AKT), *k*-2-paths (A2P): useful for modelling social circuit dependence

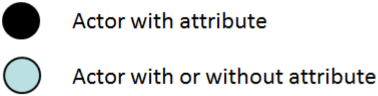


k-triangles *k*-2-paths

Model configurations — binary actor attributes

Activity ρ

Interaction (homophily) ρ_B



Inference vs estimation

- *Estimation* is to estimate parameter values,
- with an associated estimate of the standard error.
- *Inference* is making an inference as to whether or not there is a statistically significant effect (positive or negative).
- Even if the point estimates are not very accurate, if the standard error estimates are reliable, inference will still be sound,
- (at our chosen significance level, conventionally 5%).

The problem appears using snowball sampling on large networks

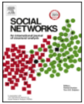
Social Networks 47 (2016) 167–188

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

ELSEVIER



Snowball sampling for estimating exponential random graph models for large networks

CrossMark

Alex D. Stivala[a,*], Johan H. Koskinen[b], David A. Rolls[a], Peng Wang[a], Garry L. Robins[a]

[a] Melbourne School of Psychological Sciences, The University of Melbourne, Australia
[b] The Mitchell Centre for SNA, and Social Statistics Discipline Area, University of Manchester, United Kingdom

A R T I C L E I N F O

Article history:
Available online 18 March 2016

Keywords:
Exponential random graph model (ERGM)
Snowball sampling
Parallel computing

A B S T R A C T

The exponential random graph model (ERGM) is a well-established statistical approach to modelling social network data. However, Monte Carlo estimation of ERGM parameters is a computationally intensive procedure that imposes severe limits on the size of full networks that can be fitted. We demonstrate the use of snowball sampling and conditional estimation to estimate ERGM parameters for large networks, with the specific goal of studying the validity of inference about the presence of such effects as network closure and attribute homophily. We estimate parameters for snowball samples from the network in parallel, and combine the estimates with a meta-analysis procedure. We assess the accuracy of this method by applying it to simulated networks with known parameters, and also demonstrate its application to networks that are too large (over 40 000 nodes) to estimate social circuit and other more advanced ERGM specifications directly. We conclude that this approach offers reliable inference for closure and homophily.

The problem: very low power on some parameters

A.D. Stivala et al. / Social Networks 47 (2016) 167

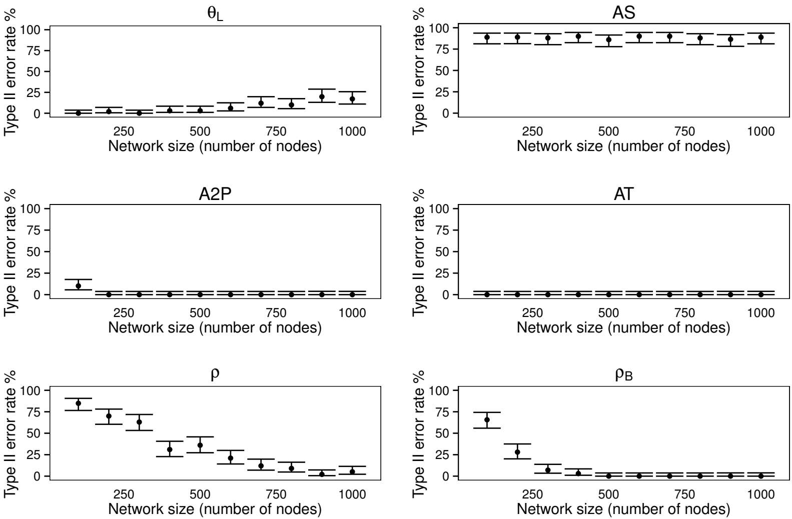
Table 5
Error statistics and number of converged snowball samples (out of 20) for the simulated networks, incl
networks, using the median point estimator, when density is not fixed. True parameter values for each mo

N	Attributes	Effect	Bias	RMSE	Type II error rate (%)		
					Estim.	95% C.I.	
						Lower	Uppe
					...		
5000	50/50	A2P	0.0199	0.0311	1	0	5
5000	50/50	AT	0.0061	0.0423	0	0	4
5000	50/50	Edge	8.4380	14.2800	100	96	100
5000	50/50	AS	−2.2470	3.6810	100	96	100
5000	50/50	ρ	−0.0121	0.0979	80	71	87
5000	50/50	ρ_B	0.0132	0.1042	8	4	15
10000	None	A2P	0.0101	0.0347	0	0	4
10000	None	AT	−0.0060	0.0510	0	0	4
10000	None	Edge	−0.3986	16.1600	95	89	98
10000	None	AS	−0.0418	4.1230	97	92	99

What the problem is not

- There appears to be no problem with the Type I error rate: inference is still "safe", just with very low power.
- It is not specific to snowball sampling. Also happens with "conventional" full network algorithms i.e. PNet, statnet (Robbins-Monro, MCMLE (Geyer-Thompson), "Stepping").
- It is not (just) because of the small AS parameter magnitude. Also happens for larger values.
- It is not (just) because of a positive AS parameter. Also happens for negative values (even with larger magnitude), although it is not as bad.
- It is not specific to alternating k-star (AS). It also happens with 2-star, 3-star, and 4-star in place of AS.
- Estimation with fixed density does not solve the problem either.

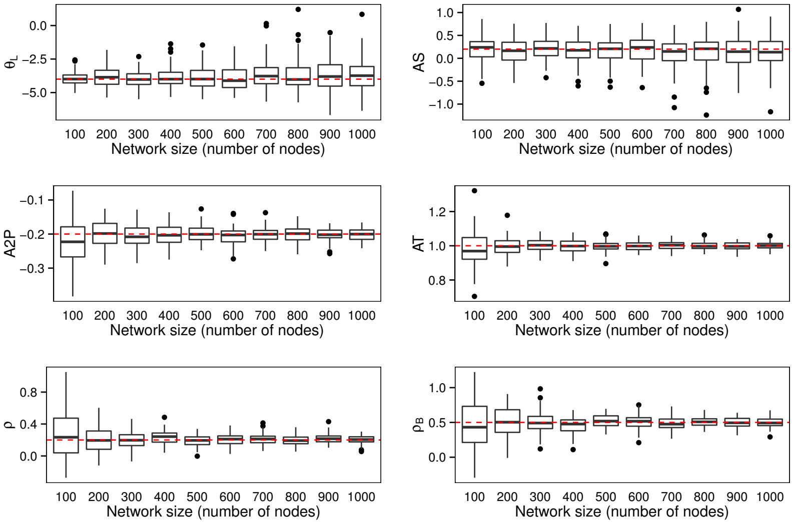
PNet estimates Type II error rate for different network sizes



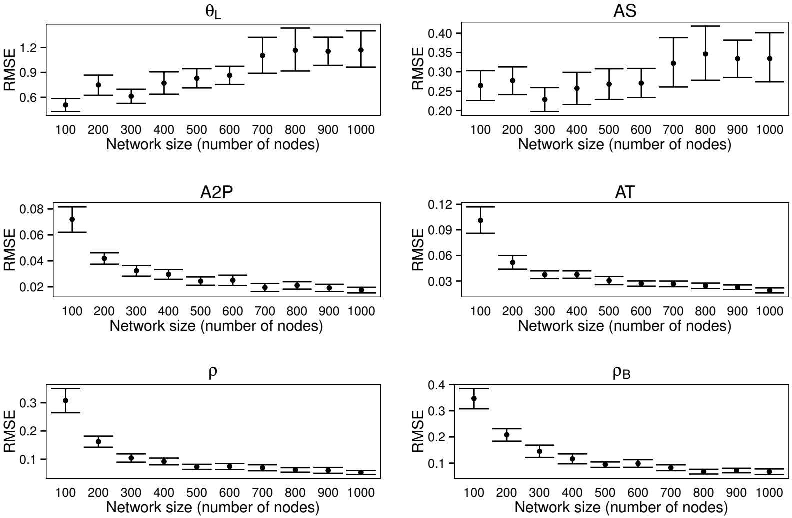
A closer look at the problem

By looking at the estimate bias and RMSE, and estimated standard error for different network sizes, we will see that the bias and RMSE in estimates for Edge and AS, unlike other parameters, does not get smaller with larger networks.

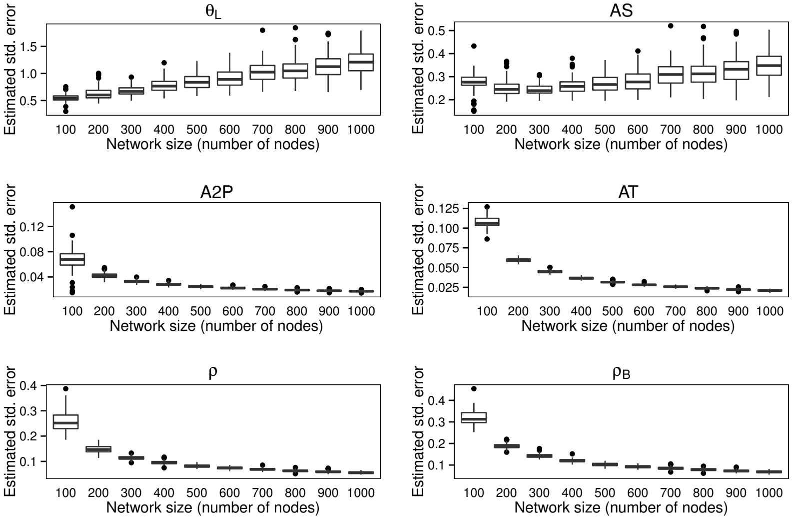
PNet estimates of canonical parameters for different sizes



RMSE on PNet estimates



Estimated standard error of PNet estimates



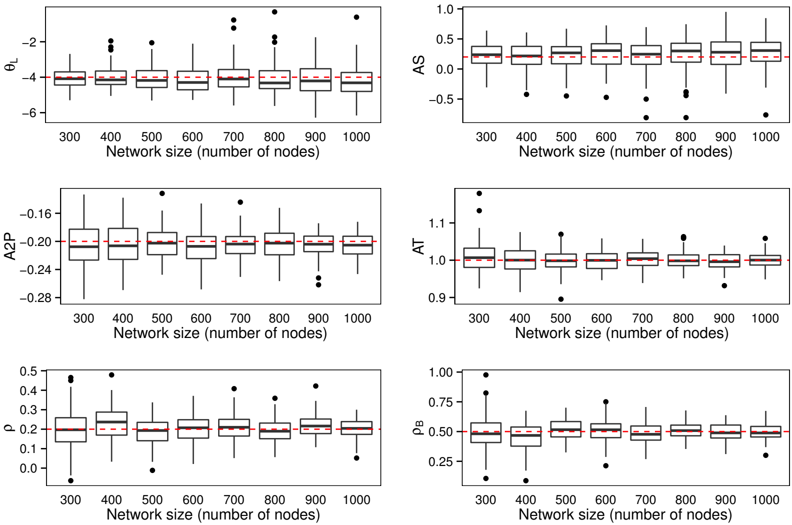
The cause of the problem?

- It looks like there is error in the point estimates of Edge and AS.
- It gets worse as network size increases.
- The estimated standard error then, correctly, is also larger.
- And hence we get a very high Type II error rate on those parameters.

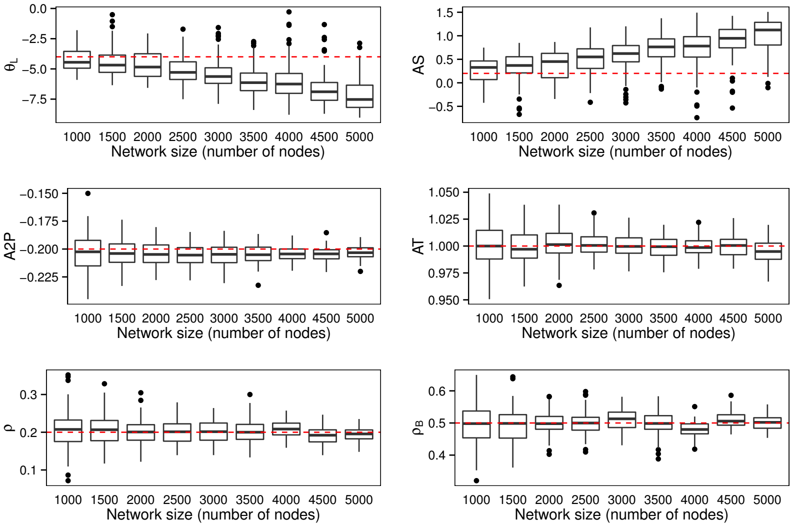
Even larger networks

- With PNet, statnet, etc. estimating networks larger than around 1000 nodes can be impractically slow.
- A new method called "Equilibrium Expectation" (Byshkin et al.) is much faster and can easily handle networks of tens or hundreds of thousands of nodes (or more).
- ("Efficient MCMC Estimation for Exponential Random Graph Models", Maksym Byshkin, Fri. June 2 09:20, BICC 310)
- The problem also occurs with this method, and the following slides show the error in the Edge and AS parameters gets even larger as the network size grows...

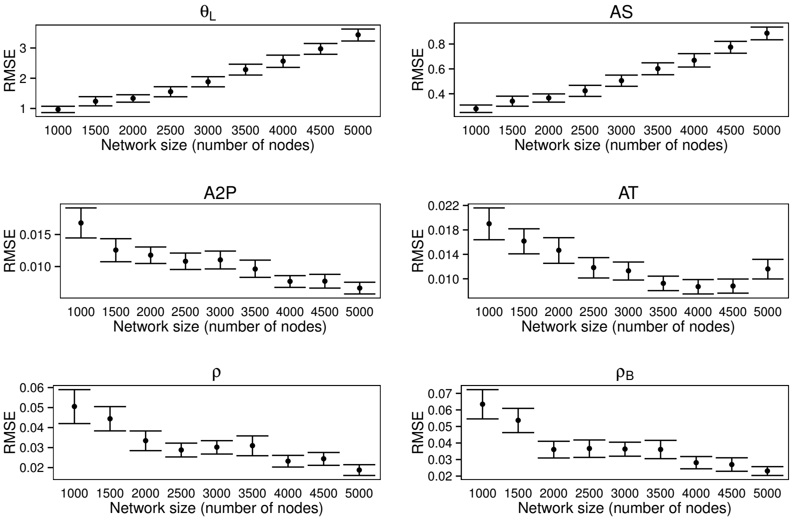
EE estimates of canonical parameters for different sizes



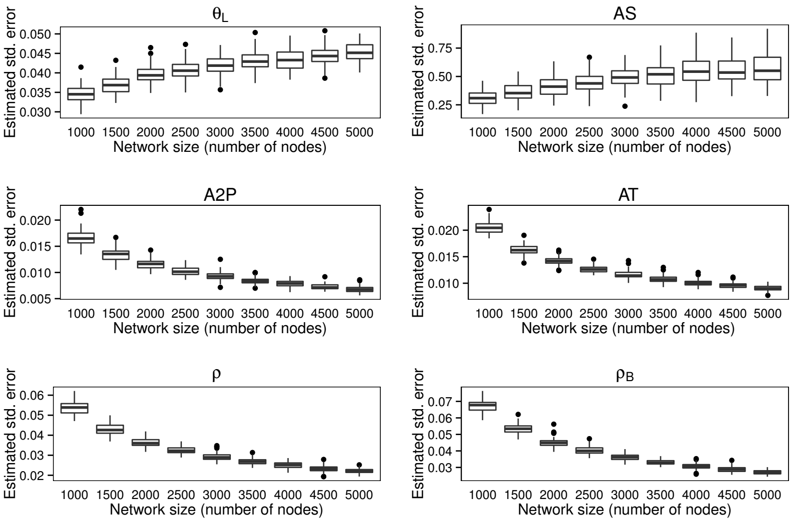
EE estimates of canonical parameters for larger networks



RMSE on EE estimates of canonical parameters for larger networks



Estimated Std. Error on EE estimates for larger networks



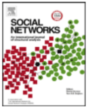
An important paper helps us find the root cause

Social Networks 31 (2009) 52–62

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet



A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models

Marijtje A.J. van Duijn[a,1], Krista J. Gile[b,2,3], Mark S. Handcock[b,*,3]

[a] Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands
[b] Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4332, United States

ARTICLE INFO

Keywords:
Networks
statnet
Dyad dependence
Mean-value parameterization
Markov Chain Monte Carlo

ABSTRACT

The statistical modeling of social network data is difficult due to the complex dependence structure of the tie variables. Statistical exponential families of distributions provide a flexible way to model such dependence. They enable the statistical characteristics of the network to be encapsulated within an exponential family random graph (ERG) model. For a long time, however, likelihood-based estimation was only feasible for ERG models assuming dyad independence. For more realistic and complex models inference has been based on the pseudo-likelihood. Recent advances in computational methods have made likelihood-based inference practical, and comparison of the different estimators possible.

In this paper, we present methodology to enable estimators of ERG model parameters to be compared. We use this methodology to compare the bias, standard errors, coverage rates and efficiency of maximum likelihood and maximum pseudo-likelihood estimators. We also propose an improved pseudo-likelihood estimation method aimed at reducing bias. The comparison is performed using simulated social network data based on two versions of an empirically realistic network model, the first representing Lazega's law firm data and the second a modified version with increased transitivity. The framework considers estimation of both the natural and the mean-value parameters.

The results clearly show the superiority of the likelihood-based estimators over those based on pseudo-likelihood, with the bias-reduced pseudo-likelihood out-performing the general pseudo-likelihood. The use of the mean-value parameterization provides insight into the differences between the estimators and when these differences will matter in practice.

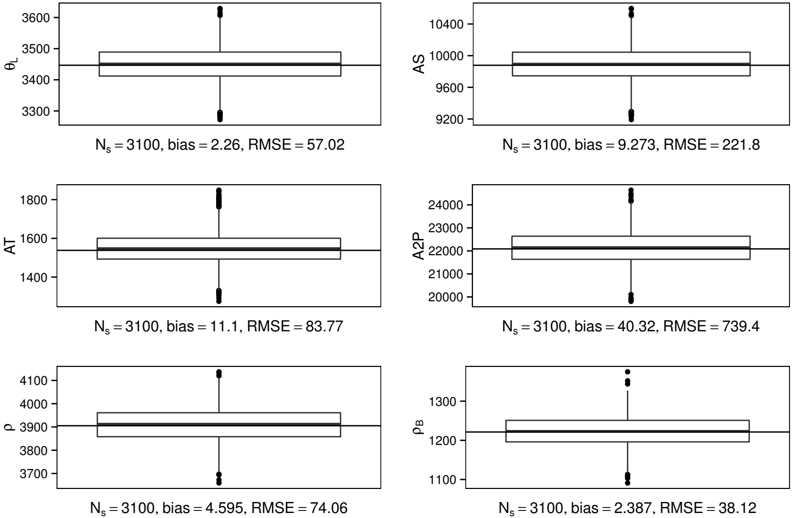
Canonical and mean-value parameterization

- The *canonical* (or “natural”) parameters are the ones we use for inference.
- There is an alternative parameterization, the *mean-value* parameterization, in which the parameters are just the network sufficient statistics.
- There is a one-to-one strictly increasing mapping from the canonical to mean-value parameterization (a general property of exponential families).
- An MLE is unbiased in the mean-value parameter space by construction.
- But *because* of this fact, and the strictly increasing mapping, it is biased in the natural parameter space.

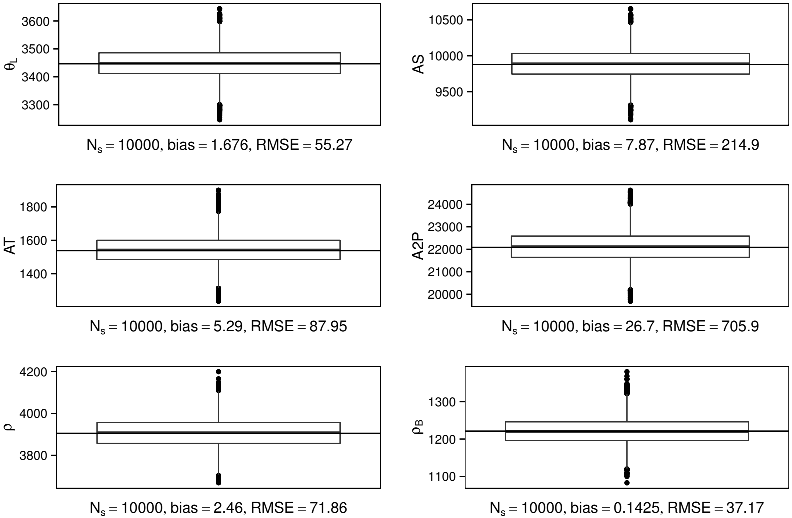
Our estimates are unbiased in the mean-value parameter space

- We can measure bias in the mean-value parameter space by simulating networks from the estimated canonical parameters, and comparing their sufficient statistics to those of the original observed network.
- This was done in van Duijn, Gile, & Handcock (2009), but only for small networks (faster computers and new algorithms mean we can now estimate networks not practical in 2009), and not using the alternating k-star parameter (or any kind of star parameter).
- Repeating this on our estimates (PNet and, for large networks, EE), we see they are unbiased in the mean-value parameter space.
- And for this very reason, are biased in the natural parameter space.

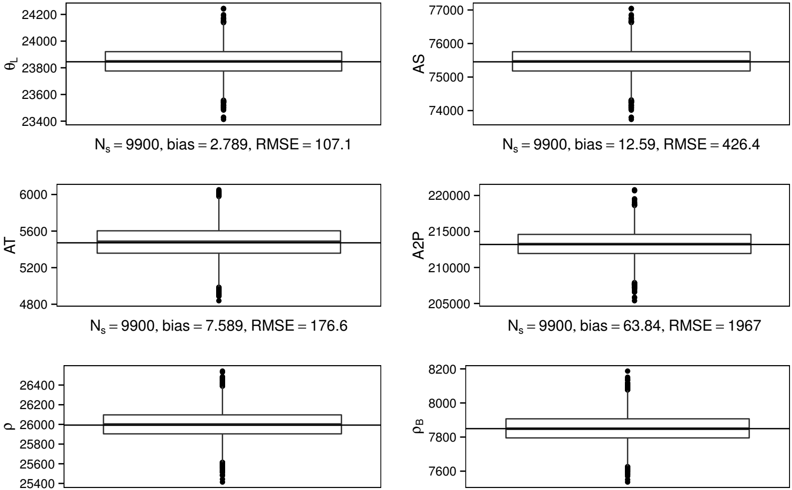
PNet estimates of mean-value parameters for 1000 node network



EE estimates of mean-value parameters for 1000 node network



EE estimates of mean-value parameters for 5000 node network



$N_s = 9900$, bias = 2.789, RMSE = 107.1

$N_s = 9900$, bias = 12.59, RMSE = 426.4

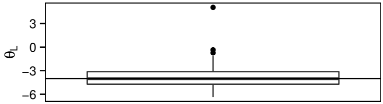
$N_s = 9900$, bias = 7.589, RMSE = 176.6

$N_s = 9900$, bias = 63.84, RMSE = 1967

$N_s = 9900$, bias = 5.159, RMSE = 146.9

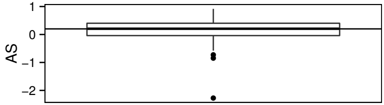
$N_s = 9900$, bias = 1.143, RMSE = 82.69

PNet estimates of canonical parameters for 1000 node network



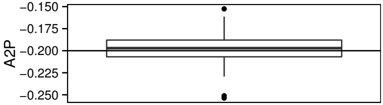
$N_c = 100$, bias = 0.1497, RMSE = 1.483

% In CI = 91, FNR% = 14



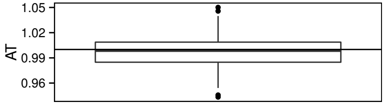
$N_c = 100$, bias = −0.04465, RMSE = 0.4167

% In CI = 91, FNR% = 87



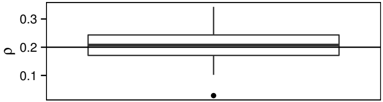
$N_c = 100$, bias = 0.001783, RMSE = 0.01713

% In CI = 95, FNR% = 0



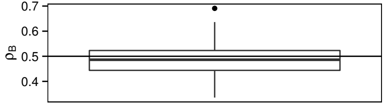
$N_c = 100$, bias = −0.002521, RMSE = 0.02059

% In CI = 94, FNR% = 0



$N_c = 100$, bias = 0.006651, RMSE = 0.05339

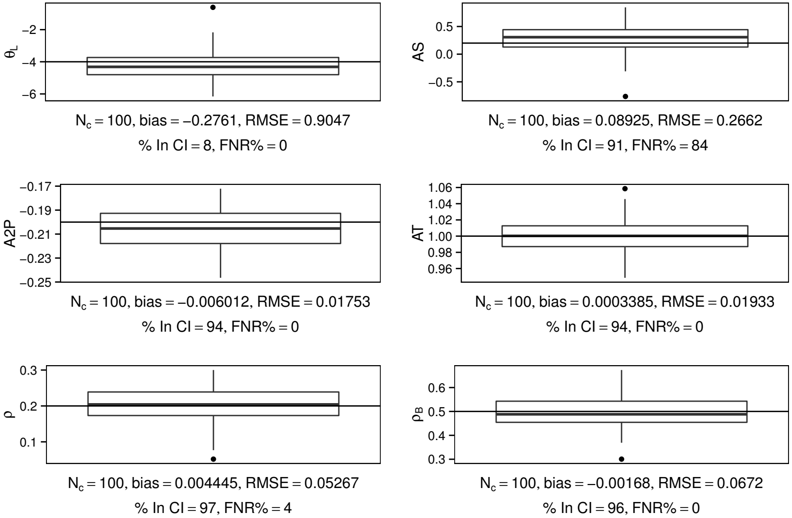
% In CI = 97, FNR% = 2



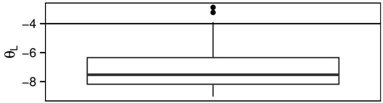
$N_c = 100$, bias = −0.01073, RMSE = 0.0653

% In CI = 99, FNR% = 0

EE estimates of canonical parameters for 1000 node network

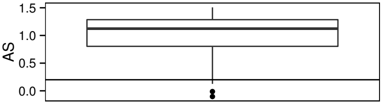


EE estimates of canonical parameters for 5000 node network



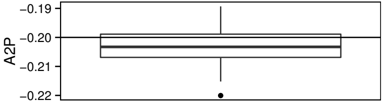
$N_c = 99$, bias = −3.162, RMSE = 3.435

% In CI = 1.01, FNR% = 0



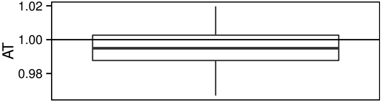
$N_c = 99$, bias = 0.8153, RMSE = 0.8871

% In CI = 64.65, FNR% = 47.47



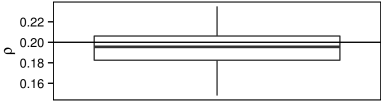
$N_c = 99$, bias = −0.003319, RMSE = 0.006637

% In CI = 95.96, FNR% = 0



$N_c = 99$, bias = −0.005265, RMSE = 0.01161

% In CI = 87.88, FNR% = 0



$N_c = 99$, bias = −0.005508, RMSE = 0.01878

% In CI = 96.97, FNR% = 0

$N_c = 99$, bias = 0.001679, RMSE = 0.0231

% In CI = 98.99, FNR% = 0

What to do about it?

It is already useful just to know to expect a very low inferential power, and, for large enough networks, very large bias on the AS parameter. But why is it only such a problem for Edge and AS? And can we fix it?

- Bias correction, such as that described for maximum pseudo-likelihood estimation in van Duijn, Gile, & Handcock (2009).
- ("Bias-adjusted maximum likelihood estimation" has been suggested by Hummel & Hunter following this idea, but seems not to be published).
- "Normalized degree distribution" instead of Edge and alternating *k*-star parameters (suggested by Maksym Byshkin; work in progress).

Acknowledgments

- Co-authors: Maksym Byshkin, Garry Robins.
- Prof. Tom Snijders.
- Dr David Rolls.
- University of Melbourne ITS High Performance Computing.
- This research was supported by Melbourne Bioinformatics at the University of Melbourne, grant number VR0261.
- This work was funded by PASC project "Snowball sampling and conditional estimation for exponential random graph models for large networks in high performance computing" and was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID c09.

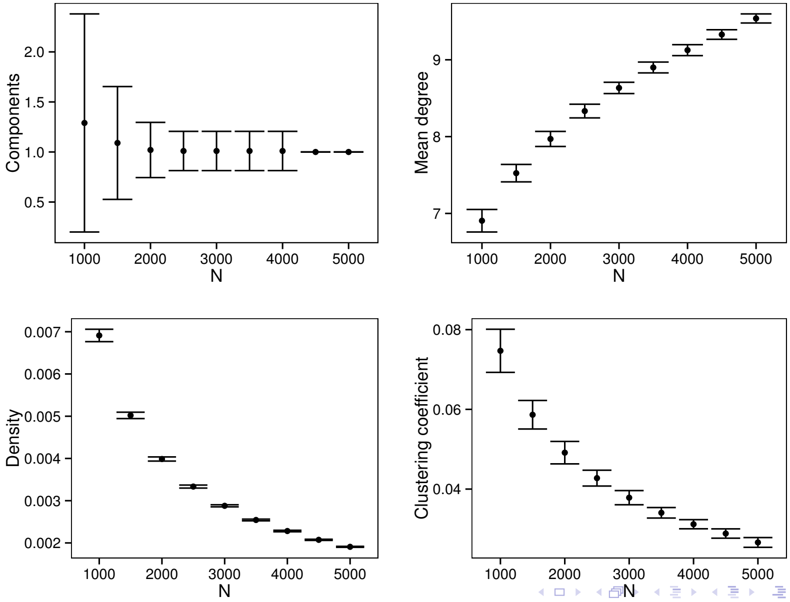
Hidden bonus slides

Simulated network parameters

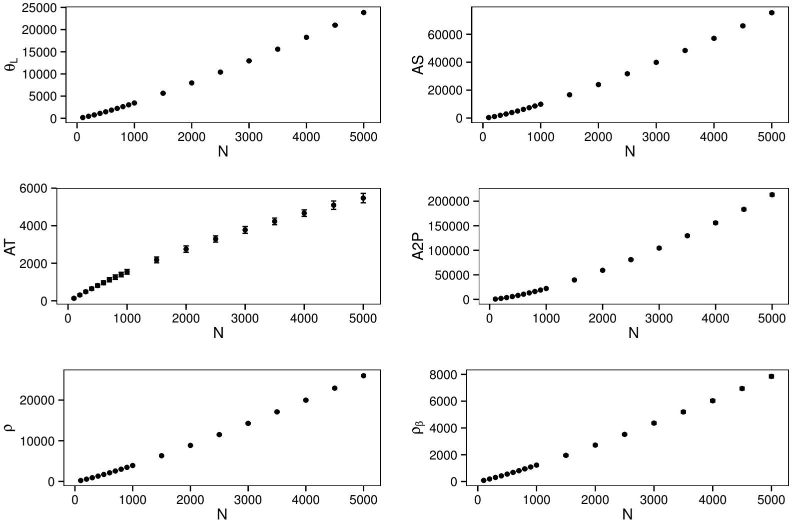
Attributes	Edge (θ_L)	AS	AT	A2P	ρ	ρ_B
50/50	-4.00	0.20	1.00	-0.20	0.20	0.50

Parameters of the simulated networks. The value in the Attributes column shows the percentage of nodes which have, respectively, the values of True and False for their binary attribute. For networks with attributes, the ρ and ρ_B columns then show, respectively, the activity and interaction parameter values.

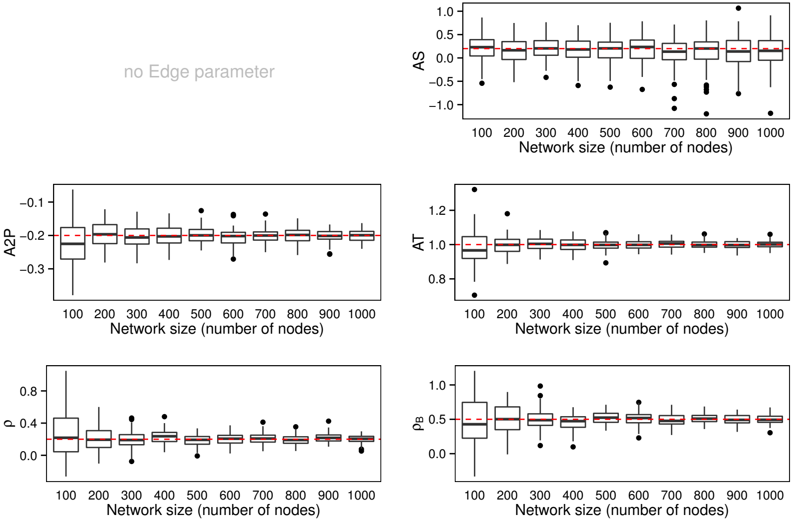
Simulated network descriptive statistics



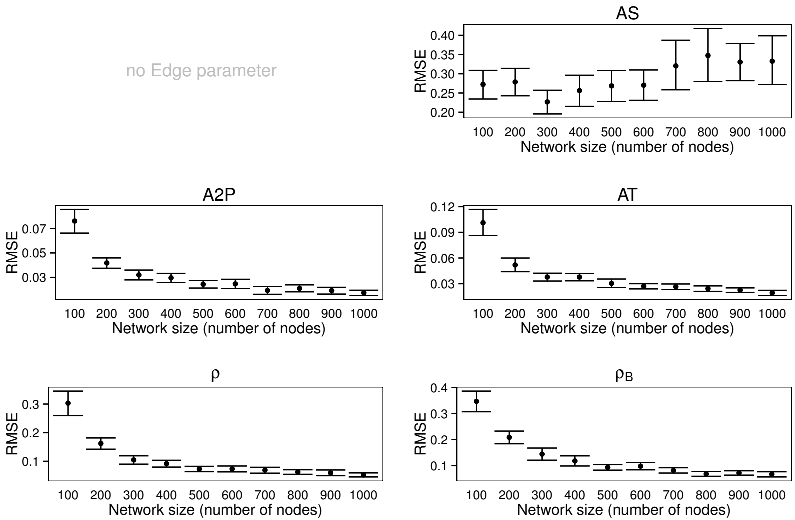
Simulated network sufficient statistics



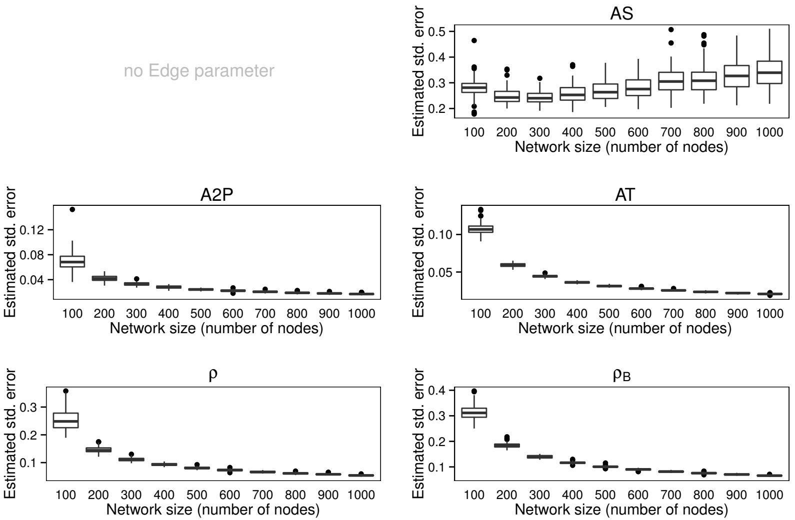
PNet fixed density estimates of canonical parameters for different sizes



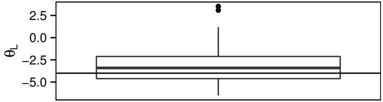
RMSE on PNet fixed density estimates



Estimated standard error of PNet fixed density estimates

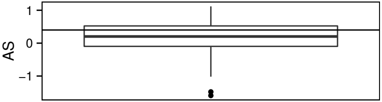


PNet estimates of canonical parameters for 500 node network (AS = 0.4)



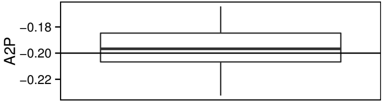
$N_c = 100$, bias = 0.8309, RMSE = 2.036

% In CI = 98, FNR% = 51



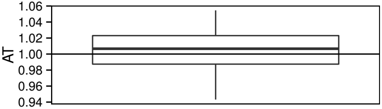
$N_c = 100$, bias = −0.2264, RMSE = 0.5492

% In CI = 97, FNR% = 86



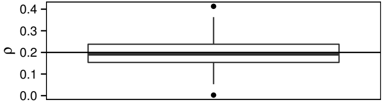
$N_c = 100$, bias = 0.003611, RMSE = 0.01622

% In CI = 96, FNR% = 0



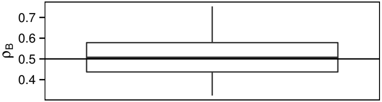
$N_c = 100$, bias = 0.005987, RMSE = 0.02578

% In CI = 98, FNR% = 0



$N_c = 100$, bias = −0.00226, RMSE = 0.07357

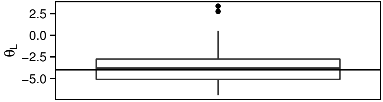
% In CI = 95, FNR% = 17



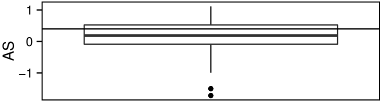
$N_c = 100$, bias = 0.004843, RMSE = 0.09491

% In CI = 91, FNR% = 0

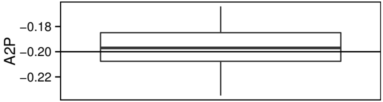
Statnet MCMLE estimates of canonical parameters for 500 node network (AS = 0.4)



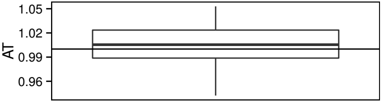
$N_c = 100$, bias = 0.3648, RMSE = 1.943
% In CI = 97, FNR% = 49



$N_c = 100$, bias = −0.236, RMSE = 0.5669
% In CI = 98, FNR% = 87

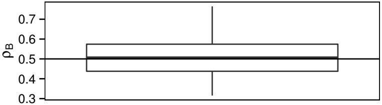


$N_c = 100$, bias = 0.003501, RMSE = 0.01635
% In CI = 96, FNR% = 0



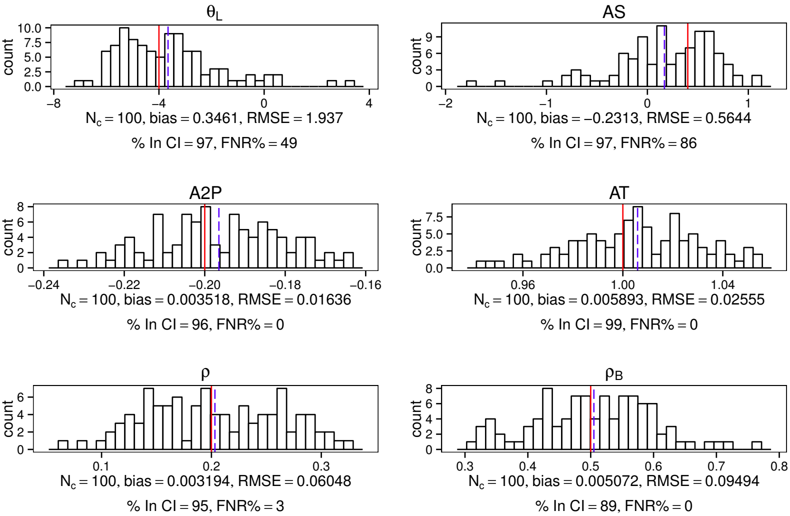
$N_c = 100$, bias = 0.005855, RMSE = 0.0255
% In CI = 99, FNR% = 0

$N_c = 100$, bias = 0.003397, RMSE = 0.06054
% In CI = 95, FNR% = 3



$N_c = 100$, bias = 0.005152, RMSE = 0.09505
% In CI = 90, FNR% = 0

Statnet "Stepping" estimates of canonical parameters for 500 node network (AS = 0.4)



Normalized degree distribution (Maksym Byshkin)

- Let D_k be the number of nodes with degree k.
- Density L is a function of degree distribution: $L = \frac{1}{2} \sum D_k k$
- Normalized degree distribution $D'_k = \frac{D_k}{L} = \frac{2D_k}{\sum D_k k}$.
- And we can use the geometrically weighted degree statistic, but with D'_k instead of D_k, and not include the density parameter in the model.

Mapping from canonical to mean value parameterization

In cases where $\mathfrak{P}$ is full and $\{P_\theta : \theta \in \Theta\}$ is minimal, τ will denote the mapping defined on int Θ by

$$\tau(\theta) = E_\theta t$$

and $\mathfrak{T}$ will stand for $\tau(\text{int}\,\Theta)$. The mapping τ is a one-to-one, both ways continuously differentiable mapping between the two open, connected sets int Θ and $\mathfrak{T}$; moreover, τ is strictly increasing in the sense that

$$(\theta - \tilde{\theta})(\tau(\theta) - \tau(\tilde{\theta})) > 0 \tag{28}$$

for every pair of points $\theta, \tilde{\theta}$ in int Θ. It follows, in particular, that if Θ is open (i.e. $\mathfrak{P}$ is regular) then $\mathfrak{P}$ may be parametrized by the mapping $\tau \to P_\theta$ (where $\tau = \tau(\theta)$). Such a parametrization is called a *mean value parametrization.*

Besides canonical and mean value parametrizations of regular exponential families, also parametrizations which are, so to speak, a mixture of the two have interest. Consider a partition $(\theta^{(1)}, \theta^{(2)})$ of θ and the similar partition $(\tau^{(1)}, \tau^{(2)})$ of τ. Observing that $\tau = D\kappa$ and applying Theorem 5.34 to κ one finds that the mapping

O. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. John Wiley & Sons, Chichester, UK, 2014.