Exponential Random Graph Models with Big Networks: Maximum Pseudolikelihood Estimation and the Parametric Bootstrap

Christian S. Schmid and Bruce A. Desmarais

The Pennsylvania State University
State College, USA

Abstract. With the growth of interest in network data across fields, the Exponential Random Graph Model (ERGM) has emerged as the leading approach to the statistical analysis of network data. ERGM parameter estimation requires the approximation of an intractable normalizing constant. Simulation methods represent the state-of-the-art approach to approximating the normalizing constant, leading to estimation by Monte Carlo maximum likelihood (MCMLE). MCMLE is accurate when a large sample of networks is used to approximate the normalizing constant. However, MCMLE is computationally expensive, and may be prohibitively so if the size of the network is on the order of 1,000 nodes (i.e., one million potential ties) or greater. When the network is large, one option is maximum pseudolikelihood estimation (MPLE). The standard MPLE is simple and fast, but generally underestimates standard errors. We show that a resampling method—the parametric bootstrap—results in accurate coverage probabilities for confidence intervals. We find that bootstrapped MPLE can be run in 1/5th the time of MCMLE. We study the relative performance of MCMLE and MPLE with simulation studies, and illustrate the two different approaches by applying them to a network of bills introduced in the United State Senate.

1 Introduction

The field of network science faces a double-edge sword when it comes to computationally intensive research. First, the availability of digital source data has led the growth in network science to be synonymous with the growth in research on big data. Second, analytical methods are growing more sophisticated, increasingly involving iterative and/or simulation-based optimization, rather than simple descriptive calculations [27]. Closing the gap in terms of the size of the networks to which it is feasible to apply the most sophisticated methods of network modeling requires research into scalable methods of inference. We propose a method of statistical inference for one of the most popular models for networks—the exponential random graph model (ERGM), in which both parameter estimates and confidence intervals are derived, that can require less than half the compute time of currently used methods.

The ERGM is a probabilistic model for networks [21, 30]. They can be used for link prediction [19], simulating network adjacency matrices [10], and testing theories regarding the processes underlying tie formation [9]. The ERGM was first introduced by Holland and Leinhardt (1981) [11]. However, due to the intractable normalizing constant in the likelihood function of the ERGM, it did not see widespread and complete use until the 2000s, following the development of algorithms and software for efficient simulation-based methods for working with ERGM [26]. Training ERGM using simulation-based methods is computationally expensive, and can still be prohibitively burdensome with data on big networks. Approximate methods of estimation, which are much more feasible with large networks, have existed for some time, but these methods perform poorly when it comes to characterizing the uncertainty in parameter estimates, which is necessary when assessing risk in predictions or simulation, or in hypothesis testing.

The ERGM takes the adjacency matrix of an observed network G^{obs}, which is a matrix-valued random variable. This means that a network of N nodes can be defined as a adjacency matrix $G = (g_{ij}) \in \{0,1\}^{(N\times N)}$, where $g_{ij} \in \{0,1\}$ for all $i,j \in \{1,\ldots,N\}$. $g_{ij} = 1$ means that there is an edge between actors i and j, while $g_{ij} = 0$ indicates that these actors are not directly connected. Since the model does not consider loops, one has $g_{ii} = 0$ for all $i \in \{1,\ldots,N\}$. Furthermore, define $\mathcal{G}(N)$ as the set of all possible networks on N nodes without loops. Note that the cardinality of set $\mathcal{G}(N)$ is increasing exponentially for every newly included actor, which results in $2^{N(N-1)/2}$ total elements. For this reason calculating the likelihood function of the ERGM, which requires evaluating a normalizing constant on $\mathcal{G}(N)$ is either extremely time-consuming or with today's technology not achievable. As a consequence, many approximation methods have been provided by the literature, with the most popular method making use of Markov Chain Monte Carlo (MCMC) methods [15], as we will introduce in the next section.

The probability function for the ERGM is defined as

$$\mathbb{P}_\theta(G) = \frac{\exp(\theta^T \cdot \Gamma(G))}{\sum_{G^*\in\mathcal{G}(N)} \exp(\theta^T \cdot \Gamma(G^*))} \tag{1}$$

where $\theta \in \mathbb{R}^q$ is a $q-$dimensional vector of parameters, $\Gamma : \mathcal{G}(N) \to \mathbb{R}^q \ , \ G \mapsto (\Gamma_1(G),\ldots,\Gamma_q(G))^T$ is a q-dimensional function of different network statistics and $c(\theta) := \sum_{G^*\in\mathcal{G}(N)} \exp(\theta^T \cdot \Gamma(G^*))$ is a normalization constant which ensures that (1) defines a probability function on $\mathcal{G}(N)$. The generative processes captured by a model (e.g., density, reciprocity, popularity, clustering) are informed by the decision regarding which network statistics (i.e., $\Gamma(\cdot)$) are incorporated. The flexibility of the ERGM in capturing virtually any network generative process has led to it being applied broadly across several fields, including sociology [24], economics [18], political science [3], ecology [6], and neuroscience [23].

2 Estimation

The first method proposed in the literature for estimating ERGM parameters was maximum pseudolikelihood estimation [28]. Under maximum pseudolikelihood estimation (MPLE), the joint distribution is replaced by the product over conditional distributions [2]. The conditional probability of a tie in ERGM reduces, conveniently, to a logistic regression form given by

$$\mathbb{P}_\theta\left(g_{ij}=1|G_{-ij}\right)=\text{logit}^{-1}\left(\theta^T\cdot\delta(\Gamma(G))\right),$$

where G_{-ij} is the adjacency matrix, excluding element ij, $\delta(\Gamma(G))$ is the "change statistic" given by the difference in the network statistics when the ij element is toggled from 0 to 1 (i.e., $\Gamma(G|g_{ij}=1))-\Gamma(G|g_{ij}=0)$), and $\text{logit}^{-1}(x)=1/(1+\exp(-x))$ [9]. For the ERGM, the pseudolikelihood function can be maximized using logistic regression software, in which the dependent variable is given by the elements of the adjacency matrix, and the covariates are given by the values of the change statistics corresponding to each element of the adjacency matrix.

Despite the computational efficiency underlying the implementation of MPLE, existing methods for assessing uncertainty with respect to the MPLE perform poorly (see van Duijn et al. [29]). Estimating the uncertainty in parameter estimates (e.g., standard errors, confidence intervals), is a critical step in using the results from a statistical model. Estimates of uncertainty are used to test hypotheses about parameters, estimate variance (i.e., risk) in model predictions, and estimate effect sizes. The current conventional approach to estimating θ, introduced by Snijders [26], is based on a Markov Chain Monte Carlo (MCMC) approximation of the MLE. This Monte Carlo maximum likelihood method (MCMLE) is based on a more direct attempt to approximate the log-likelihood function derived from (1). The log-likelihood function is not evaluated directly, rather, the log ratio of the likelihood under a proposed value of the parameters θ, and an initial value of the parameters θ_0, is approximated using L networks simulated from the ERGM with parameter values θ_0. The approximation, detailed in Snijders (2002) [26] is given by

$$\text{loglik}(\theta)-\text{loglik}(\theta_0)\approx-\log\left(\frac{1}{L}\cdot\sum_{i=1}^{L}\exp\left((\theta-\theta_0)^T\cdot\Gamma(G_i)\right)\right) \tag{2}$$

As we demonstrate below, the MCMLE grows more accurate as L increases. Indeed, MCMLE approaches the MLE as the number of networks simulated goes to infinity.

3 Efficiency of MPLE and MCMLE

As mentioned in the previous section the MPLE approaches the MLE as the size of the networks increase and as a consequence, is a consistent estimator (see Lindsay [17], Strauss and Ikeda [28], Hyvarinen [16], Desmarais and Cranmer [4, 5]). This implies that for an increasing number of nodes, the MPLE converges in probability to the MLE, meaning that for large enough networks the MPLE performs as well as or better than MCMLE, and requires less compute time. To illustrate the relative efficiency of MPLE and MCMLE we run a simulation study. Desmarais and Cranmer [4] show the MPLE outperforms the MCMLE if the number of simulated networks used to approximate the likelihood in MCMLE is not large enough. It is even more remarkable that the number of simulated networks needed for the MCMLE, in order to surpass the MPLE increases as the size of the network (i.e., the number of nodes in the network) increases. This means that, for very large networks, it becomes difficult to determine the number of simulated networks required for the MCMLE to outperform the MPLE. In other words, the larger the network, the more computationally intensive it becomes to use MCMLE in a way that out-performs MPLE.

To demonstrate this disadvantage of the MCMLE we conduct a simulation study using Goodreau's [14] Faux Mesa High School data, which represents a simulation of an in-school friendship network among 203 students as well as the Faux Magnolia High School data, representing an in-school friendship network among 1451 students. The data for both networks originates from Resnick et al. [20]. For both networks, we first calculate the MCMLE and treat the estimated coefficients as the network's true values θ. Then, we take the same parametrization, using the number of edges, the nodal attribute for gender, and the geometrically weighted edgewise shared partners (gwesp) distribution (see Hunter [13]) where we fix the decay parameter λ at 0.25. The gwesp statistic is used to model the tendency towards triangles and clustering in a network.

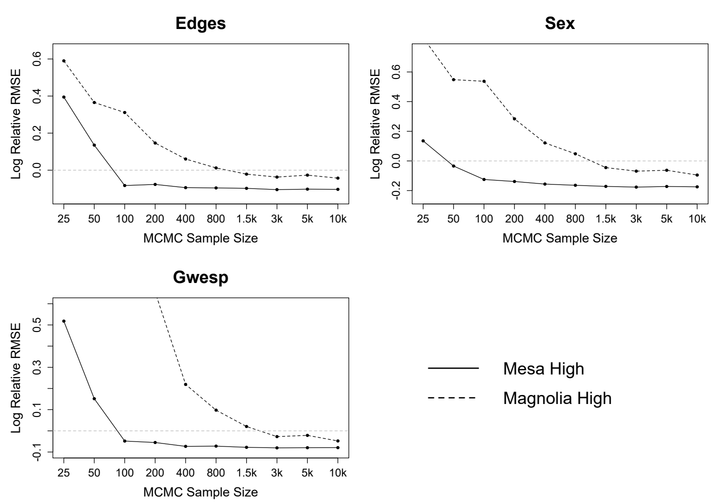

Fig. 1. The log of the ratio of the RMSE for the MCMLE to the MPLE for different sample sizes and two different networks, Faux Mesa High and Faux Magnolia High

We simulate $m = 500$ new networks using the 'true' coefficients and estimate the MPLE as well as the MCMLE of these simulated networks. For every single simulated network the MCMLE calculation is being repeated several times for 25 to $10,000$ simulated networks used in the likelihood approximation. Based on these results, we compute the root mean square error, which is a measure of the accuracy of an estimator, combining both the bias and the variance. Mathematically written, the RMSE for an estimator $\hat{\theta}$ is defined as

$RMSE = \sqrt{\sum_{i=1}^{m}(\theta - \hat{\theta}_i)^2}$ implying that the smaller the RMSE, the more accurate is the estimator. Since the MCMLE has higher efficiency and converges to the MLE, the RMSE decreases as the number of simulated networks used for the likelihood approximation increases. On the other hand, the RMSE of the MPLE is a constant value since no network simulations are required. In order to compare the RMSE of the two estimation techniques, we take the log of the ratio of the MCMLE to the MPLE. As a result, a negative value indicates a better MCMLE performance, while a positive value indicates a better MPLE performance. Figure 1 visualizes the results of the simulation study. The solid line illustrates the results of the log relative RMSE of the Faux Mesa High network, while the dashed line illustrates the corresponding results of the Faux Magnolia High network. The plots support the fact that larger networks require a larger sample size of simulated networks for the MCMLE to outperform the MPLE. While the fairly small Faux Mesa High network only requires a sample size of about $50 - 100$ networks, the larger Faux Magnolia High network requires a sample size of at least 1,500 networks for the MCMLE to surpass the MPLE. For especially large networks (e.g., social media data) the sample size has to be set in order to justify the approximately exact, but computationally expensive and potentially prohibitive MCMLE method.

4 Bootstrapped MPLE

As discussed in the previous section, the MPLE converges to the MLE as the size of the network increases. Moreover, the MPLE is able to outperform the MCMLE if the sample size used in MCMLE is not large enough. The main reason why the MCMLE is still widely preferred is that, in contrast to the MPLE, it does not underestimate the standard errors (van Duijn et al. [29]). By the definition of the ERGM it is obvious that this model is an exponential family distribution where θ is the natural parameter and $\Gamma(G)$ is the sufficient statistic. For exponential family distributions, a covariance matrix can be estimated by the inverse of the negative Hessian matrix $[-H]^{-1}$ of the likelihood function at the MLE. The problem with the MPLE is that calculating $[-H]^{-1}$ by the pseudolikelihood function will underestimate the variance of the MPLE [29], resulting in an underestimate of the width of the confidence intervals. van Duijn et al. show that constructing 95% MPLE confidence intervals can result in intervals that comprise the true value in less than 75% instead of the nominal 95%. In this paper, we are going to refer to the MPLE confidence intervals as *logistic regression confidence intervals* simply because the MPLE is calculated using logistic regression methods that also use the inverse of the negative Hessian matrix as an estimate for the covariance matrix.

Since the MPLE has the advantage of being approximately exact and computationally inexpensive, but has the disadvantage of underestimating corresponding confidence intervals, we apply a technique referred to as bootstrap resampling [7]. Bootstrap resampling refers to constructing a sampling distribution for the parameter estimate by resampling the data with replacement, and re-estimating the model on the resampled data. Under non-parametric bootstrap resampling, the data are resampled directly from the dataset. Under the parametric boostrap, the data is resampled from the estimated model. The idea of using boostrap resampling with MPLE for ERGM was first introduced by Desmarais and Cranmer [4] and provides a consistent estimate of MPLE confidence intervals. However, the methods introduced by Desmarais and Cranmer [4] only applied to cases in which the researcher had a large sample of networks (e.g., a time series of networks). For the case in which there is just a single network to be studied, we propose the use of a parametric bootstrap. Under the parametric bootstrap, the sampling distribution of the MPLE is derived by re-estimating the MPLE on a sample of networks simulated from the MPLE estimated on the observed network. We verify the consistency of the bootstrapped MPLE by conducting a simulation study on the same two networks with the same parametrization as in the previous chapter: The Faux Mesa High friendship network and the Faux Magnolia High friendship network.

For the simulation study, we determine the MPLE for the model and treat these estimates as the networks' 'true' parameter values. We use these parameter values to simulate a sample of 1000 networks from the distribution of G. For each of the 1000 networks, we calculate 95% confidence intervals based on the MCMLE and the logistic regression and examine whether the 'true' parameter values lie in these intervals. In addition, we determine the bootstrapped MPLE confidence intervals by sampling 500 networks for each of the originally sampled 1000 networks, by using the respective MPLE as parameter values. For every newly sampled network, we again determine the MPLE and then take the 2.5th and 97.5th percentile of the 500 MPLE estimates to obtain 95% bootstrap confidence intervals. We verify whether the 'true' parameter value can be found in the bootstrapped confidence interval.

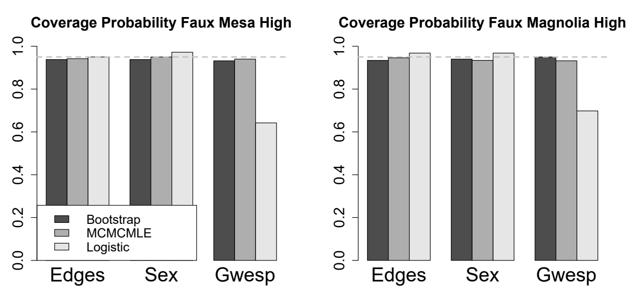

Fig. 2. The Coverage Probability results of the Faux Mesa High network (left) and of the Faux Magnolia High network (right) for bootstrapped MPLE, MCMLE and logistic regression

Figure 2 visualizes the coverage percentages for each of the three methods for both networks. The dashed line is set at 0.95 and represents the optimal value. It is evident that the bootstrapped MPLE performed equally well as the MCMLE, achieving results that obtain the true parameter values in approximately 95% of the cases. Additionally, a difference in the results between the smaller Faux

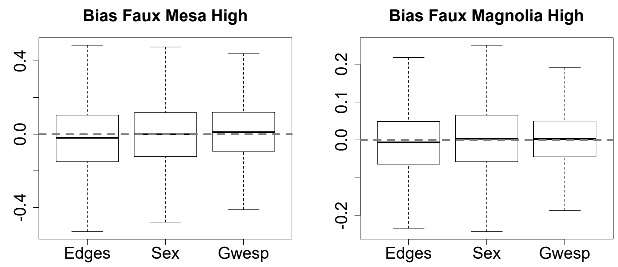

Fig. 3. The boxplots visualize the bias $(\hat{\theta} - \theta)$ over the 500 iterations for the Faux Mesa High network (left) and the Faux Magnolia High network (right)

	MCMLE		Logistic Regression		bootstrapped MPLE	
	Estimate	St. Error	Estimate	St. Error	Lower Bound	Upper Bound
Edges	-5.884	0.065	-5.869	0.015	-6.007	-5.751
Sponsor Party	1.440	0.015	1.440	0.015	1.411	1.467
Alternating k-star	0.124	0.064	0.108	0.006	-0.011	0.2379

Table 1. Estimation results for the Cosponsorship network using MCMLE, logistic regression and bootstrapped MPLE

Mesa High network and larger Faux Magnolia High network is not identifiable. Similar to the results of van Duijn et al. [29] our results for the logistic regression differ distinctively from the anticipated 95%, confirming that the MPLE underestimates the variance of its estimates. Figure 3 illustrates the bias between the 'true' network coefficients θ and the MPLE estimates. The median MPLE estimates approximate the true parameters. It is especially worthwhile to mention that the bias of the larger Faux Magnolia High network is smaller than the bias of the Faux Mesa High network, supporting the fact that the MPLE converges to the MLE as the network size increases.

5 Cosponsorship Network Data

To illustrate the performance of MCMLE relative to that of the bootstrapped MPLE we apply both approaches to the data on cosponsorship of bills in the U.S. House of Representatives for the 108th Congress (2003–2004), developed by Fowler (2006) [8]. The cosponsorship network consists of 2,635 nodes, which we define as pieces of legislation (i.e., bills), considered by the Senate during the 108th Congress. In this undirected network bills are tied together based on the similarity of the sets of legislators who cosponsor them. Specifically, we include an edge between bills i and j if the correlation coefficient between the indicator vectors indicating whether i and j were sponsored each legislator is greater than a random uniform draw. This results in an undirected network with 28060 edges.

We build an ERGM specification that extends the work of Zhang et al [31] in exploring the structure of cosponsorship ties. They find that congressional cosponsorship is primarily characterized by intra-party ties—among Republicans and among Democrats, but few cross-party ties. We test for this party-based clustering (i.e., homophily) in our ERGM. This is done through a term that accounts for the party of the senators who sponsored the two bills in the pair. A positive parameter value for this statistic indicates that ties tend to be formed between bills sponsored by the same political party.

We extend the homophily-based model to account for a network dynamic that is commonly found in the study of networks—that of popularity or preferential attachment [1]. The alternating k-star statistic was introduced by Snijders et al. [25] and modified by Hunter and Handcock [13]. A positive parameter estimate associated with the alternating k-star statistic indicates that tie formation follows a form of preferential attachment [25]. This could arise in a network of bill-to-bill ties if the majority party in power was particularly disciplined at rallying its partisans to pile on to the bills that its members proposes, thus producing a large set of very similar bills. We estimate the ERGM using MCMLE and the bootstrapped MPLE. The MCMLE requires a sample size of at least 1000 networks to converge. The bootstrapped MPLE was estimated by using 500 simulated networks. As we described in the section *Estimation*, only one edge at a time is changed when simulating networks. The results can be found in table 5.

The MPLE estimate is equivalent to the logistic regression estimate, but the bootstrap confidence intervals, especially for the alternating k-star statistic, are much wider than would be calculated using the logistic regression standard errors. An estimate is generally considered statistically different from zero (i.e., statistically significant) if the confidence interval does not contain zero, or if the ratio of the estimate to the standard error exceeds 1.96 in magnitude. This cosponsorship network example perfectly illustrates the inferential problems that can arise with the conventional logistic regression standard errors when using MPLE. All of the parameter estimates are statistically significant according to the logistic regression estimates. However, the alternating k-star statistic is not significant according to either the MCMLE or the bootstrapped MPLE.

6 Parallel Computing with MPLE

The bootstrapped MPLE is not only simple and fast, it is highly parallel. Once the networks on which to estimate the bootstrap replicates are simulated, each re-estimate can be run in parallel. By using multiple cores, the computing time for estimating bootstrapped MPLE confidence intervals can be reduced substantially. Figure 4 illustrates the relative computing time of the bootstrapped MPLE using 500 simulated networks and the MCMLE for the three networks Faux Mesa High (205 nodes), Faux Magnolia High (1461 nodes) and Cosponsorship (2635 nodes) for an increasing number of computing cores. For the small network we simulate 2000 networks using a MCMC interval of 2000 steps, for the medium network we simulate 8000 networks using a MCMC interval of 5000 steps and for the large network we simulate 10000 networks using $30,000$ MCMC steps in order to approximate the likelihood appropriately. The chosen sample sizes and

MCMC steps are necessary to guarantee a good model fit. The small network took 14 seconds, the medium network took 123 seconds and the large network took 986 seconds to run. We define the simulation time of the bootstrapped MPLE as a function of the number of available computing cores x:

$$\text{MPLE time} = \text{network simulation time} + \frac{500 \cdot \text{MPLE estimation time}}{x}$$

Based on this, we define the relative computing time as $\frac{\text{MPLE time}}{\text{MCMLE time}}$. This means that a relative computing time greater than 1 indicates that the MCMLE computing time is shorter, while a relative computing time smaller than 1 indicates that the bootstrapped MPLE provides faster results.

Figure 4 demonstrates that all three networks only require three cores for

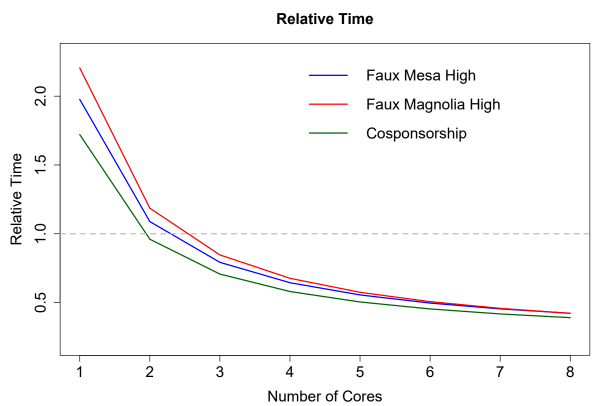

Fig. 4. The y-axis gives the ratio of the bootstrapped MPLE time to that of the MCMLE time. Values below 1 indicate that the bootstrapped MPLE requires a shorter computing time.

the bootstrapped MPLE to outperform the computing time of the MCMLE and that the computing time can further be reduced if more computing cores are available. If exactly 500 computing cores are being used the ratio of the bootstrapped MPLE time to the MCMLE time levels off at 0.20 for the small and large network and 0.17 for the medium network, meaning that the computing time can be quintupled using the bootstrapped MPLE. This figure also depicts that larger network in general require a longer computation time and will benefit more if the bootstrapped MPLE is used.

One of the major disadvantages of MPLE over MCMLE is that degeneracy is not assessed while the model is being estimated. The bootstrapped MPLE, however, allows assessing degenerate models as well since the method requires simulating from the estimated parameters. In order to verify whether a model is degenerate or not, one can take a look at density and trace plots as visualized in figure 5. The trace plots on the left side depict the the attained values via MCMC simulated networks for every single statistic included into the model, centered on the statistic values of the observed network. The plots on the right side visualize the empirical density function of the respective statistic, based on the simulated networks (Hunter and Handcock [13]). For a non-degenerated model the empirical density function should be approximately symmetrical around zero for every included centered statistic, since this corresponds with the expected value of a centered statistic. Otherwise, the values of the simulated networks systematically differ from the corresponding statistics in the observed network, making it unreasonable to assume that the simulated networks originate from the same distribution as the observed network. Furthermore, the trajectories in the trace plot should not indicate a dependence structure. This would be a signal that the constructed stochastic process violates the Markov properties.

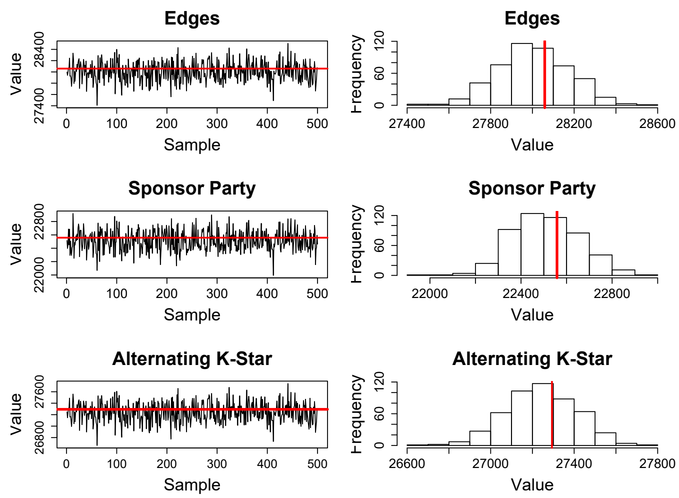

Fig. 5. Network statistics of the 500 booostrap samples for the cosponsorship network. The thick line in both, the traceplots and the histograms, represents the network statistics of the observed network.

7 Conclusion

In this paper we introduced the bootstrapped MPLE as an alternative method of statistical inference for ERGMs and compared the performance to the commonly applied MCMLE. Based on a simulation study we demonstrated that the larger the size of a network is the larger the MCMC sample size has to be in order for the MCMLE to outperform the fast and simple MPLE. However, the big disadvantage of the MPLE is that, even though it is an approximately exact estimator, it underestimates the standard error. For this reason, we propose a parametric bootstrap method of evaluating model uncertainty. On the basis of another simulation study on two different networks, we demonstrate that the bootstrapped MPLE covers the true coefficients just as well as the MCMLE, while the simple MPLE performs clearly poorer. This means that the bootstrapped MPLE combines the advantages of both methods, the MPLE and the MCMLE, because it is still simple and fast, and provides approximately exact results, but also accurately estimates model uncertainty. We conclude that the bootstrapped MPLE should be regarded as an alternative to the MCMLE. It also has the advantage of being parallel, which leads to a rapid speed-up of the calculation if multiple computing cores are used.

8 Acknowledgements

This work was supported in part by NSF grants SES-1558661, SES-1637089, SES-1619644, and CISE-1320219.

References

1. Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
2. Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 259–302, 1986.
3. S. J. Cranmer and B. A. Desmarais. Inferential network analysis with exponential random graph models. *Political Analysis*, 19(1):66–86, 2011.
4. B. A. Desmarais and S. J. Cranmer. Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications*, 391(4):1865–1876, 2012.
5. Bruce A Desmarais and Skyler J Cranmer. Consistent confidence intervals for maximum pseudolikelihood estimators. In *Proceedings of the Neural Information Processing Systems 2010 Workshop on Computational Social Science and the Wisdom of Crowds*, 2010.
6. Cody J. Dey and James S. Quinn. Individual attributes and self-organizational processes affect dominance network structure in pukeko. *Behavioral Ecology*, 25(6):1402–1408, 2014.
7. Bradley Efron. The jackknife, the bootstrap and other resampling plans. In *CBMS-NSF Regional conference series in applied mathematics. 25 cm. 92 p.* Society for industrial and applied mathematics, Philadelphia. US, 1982.
8. James H. Fowler. Connecting the congress: A study of cosponsorship networks. *Political Analysis*, 14(04):456–487, 2006.
9. Steven M Goodreau, James A Kitts, and Martina Morris. Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1):103–125, 2009.
10. Jeremy Keith Hackney and Kay W. Axhausen. An agent model of social network and travel behavior interdependence. In *Conference on Issues in Behavioral Demand Modeling and the Evaluation of Travel Time*, 2006.
11. Paul W. Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, 76(373):33–50, 1981.
12. David R. Hunter, Steven M. Goodreau, and Mark S. Handcock. Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258, 2008.
13. David R. Hunter and Mark S. Handcock. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583, 2006.
14. Hunter , David R., Handcock , Mark S., Butts , Carter T., Goodreau , Steven M., and Martina Morris. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 2008.

15. David R. Hunter, Pavel N. Krivitsky, and Michael Schweinberger. Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, 21(4):856–882, 2012.
16. Aapo Hyvrinen. Consistency of pseudolikelihood estimation of fully visible boltzmann machines. *Neural Computation*, 18:2283–2292, 2006.
17. Bruce G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, 80(1), 1988.
18. Alessandro Lomi and Fabio Fonti. Networks in markets and the propensity of companies to collaborate: An empirical test of three mechanisms. *Economics Letters*, 114(2):216–220, 2012.
19. Zhengdong Lu, Berkant Savas, Wei Tang, and Inderjit S. Dhillon. Supervised link prediction using multiple sources. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 923–928. IEEE, 2010.
20. Michael D. Resnick, Peter S. Bearman, Robert Wm Blum, Karl E. Bauman, Kathleen M. Harris, Jo Jones, Joyce Tabor, Trish Beuhring, Renee E. Sieving, Marcia Shew, Marjorie Ireland, Linda H. Bearinger, and J. Richard Udry. Protecting adolescent's from harm: Findings from the national longitudinal study on adolescent health. *JAMA - Journal of the American Medical Association*, 278(10):823–832, 9 1997.
21. Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191, May 2007.
22. Michael Schweinberger. Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496):1361–1370, 2011.
23. Sean L. Simpson, Satoru Hayasaka, and Paul J. Laurienti. Exponential random graph modeling for complex brain networks. *PloS one*, 6(5):e20039, 2011.
24. Sanne Smith, Frank Van Tubergen, Ineke Maas, and Daniel A. McFarland. Ethnic composition and friendship segregation: Differential effects for adolescent natives and immigrants. *American Journal of Sociology*, 121(4):1223–1272, 2016.
25. Tom A. B. Snijders, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006.
26. Tom A.B. Snijders. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40, 2002.
27. Tom A.B. Snijders. Statistical models for social networks. *Annual Review of Sociology*, 37, 2011.
28. David Strauss and Michael Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85:204–212, 1990.
29. Marijtje A. J. van Duijn, Krista J. Gile, and Mark S. Handcock. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, 31(1):52–62, 1 2009.
30. Stanley Wasserman and Philippa Pattison. Logit models and logistic regression for social networks: An introduction to markov graphs and p-star. *Psychometrica Vol.61*, 1996.
31. Yan Zhang, Andrew J. Friend, Amanda L. Traud, Mason A. Porter, James H. Fowler, and Peter J. Mucha. Community structure in congressional cosponsorship networks. *Physica A: Statistical Mechanics and its Applications*, 387(7):1705–1712, 2008.