Social Science Research 59 (2016) 107–119

Contents lists available at ScienceDirect

Social Science Research

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

Fitting ERGMs on big networks

Weihua An

Departments of Statistics and Sociology, Indiana University Bloomington, 752 Ballantine Hall, 1020 East Kirkwood Avenue, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form 12 March 2016
Accepted 21 April 2016
Available online 27 April 2016

Keywords:
Big networks
ERGMs
MCMLE
PMLE
Meta network analysis
Link tracing

ABSTRACT

The exponential random graph model (ERGM) has become a valuable tool for modeling social networks. In particular, ERGM provides great flexibility to account for both covariates effects on tie formations and endogenous network formation processes. However, there are both conceptual and computational issues for fitting ERGMs on big networks. This paper describes a framework and a series of methods (based on existent algorithms) to address these issues. It also outlines the advantages and disadvantages of the methods and the conditions to which they are most applicable. Selected methods are illustrated through examples.

1. Introduction

Social network analysis fundamentally changes quantitative analysis in that it shifts the research focus from individuals to their relations and interactions. A social network is typically expressed by its adjacency matrix in which a cell is coded as one if there is a tie between two nodes and zero if the tie is absent. One approach to modeling the network is to vectorize the adjacency matrix and use a logit model to model the occurrence of ties.

$$\text{logit}(P(w_{ij}=1)) = \theta' S(w,X),$$

where w_{ij} indicates the tie from i to j, S(w,X) the model terms, and θ the parameters. The P1 model (Holland and Leinhardt, 1981) can account for mutuality in tie formations and estimates the model by assuming dyadic independence. The P^* model (Frank and Strauss, 1986; Wasserman and Pattison, 1996) can account for Markov dependence (i.e., any two ties sharing a common node) in tie formations. Recent development in ERGMs (Snijders et al., 2006; Hunter, 2007; Robins et al., 2007b, 2009) can account for even higher orders of tie dependence, for example, social circuit dependence (Snijders et al., 2006; Robins et al., 2009) where two ties can be dependent even without sharing common nodes.

ERGMs are typically estimated by the Monte Carlo Maximum Likelihood Estimation (MCMLE) (Hunter and Handcock, 2006). For ERGMs that can account for complex network dependence, MCMLE relies on Monte Carlo Markov Chains (MCMCs) to draw samples of networks for estimations. The sampling process can be time-consuming, especially when the network is big, as the MCMCs will take a long time to reach the target distribution.

E-mail address: weihuaan@indiana.edu.

http://dx.doi.org/10.1016/j.ssresearch.2016.04.019
0049-089X/

Big networks also impose other computational and conceptual challenges for estimating ERGMs. First, there may be computer hardware and software issues. To load big network data and carry out the estimation often requires massive computer memory which is beyond the limit of usual desktops. In addition, the current software for network analysis mostly cannot process long vectors often used for storing big network information. Thus, it becomes extremely challenging to fit ERGMs on big networks. Second and probably more problematically, as the network size grows, the assumption imposed by MCMLE becomes unrealistic that actors have full knowledge of the network when deciding whom to connect to.

This paper provides a framework and outlines a series of methods for addressing the aforementioned issues. Unlike in small networks where speed is less of a concern, when modeling big networks, the key is to strike a good balance between accuracy/consistency on one hand and speed/stability on the other. To clarify, this paper does not propose any new algorithm for estimating EGRMs. Rather, it focuses on proposing new and more effective ways of using existent algorithms. In particular, this paper introduces meta network analysis for modeling big networks with a block structure and shows why it is valuable in these cases. Also to clarify, the main purpose of this paper is not to benchmark the different estimation methods against MCMLE, because in big networks, MCMLE is hardly available and also MCMLE may not be the right standard due to its unrealistic assumption. Those who are interested in comparing MCMLE against other methods (in small networks) should consult Lubbers and Snijders (2007) and van Duijn et al. (2009).

Overall, the paper describes two broad approaches and seven specific methods for fitting ERGMs on big networks. Each method has its own advantages and disadvantages. None works universally well. Researchers should choose the best available method according to the network size and structure and the hardware and software constraints they might have. If possible, they should try to use multiple methods and triangulate the evidence.

The paper proceeds as follows. In Section 2, I introduce ERGMs and highlight the computational challenges for fitting ERGMs on big networks. In Section 3, I present a framework and a series of methods for addressing the challenges. In Section 4, I use examples to illustrate selected methods. Last, I summarize and discuss the implications of this study.

2. ERGM

An ERGM (Wasserman and Pattison, 1996; Hunter and Handcock, 2006; Snijders et al., 2006; Robins et al., 2007a, 2007b, 2009) models the probability of observing a network w as follows.

$$P(W = w|\theta, X) = \frac{\exp\{\theta' S(w, X)\}}{K(\theta, \mathbf{W})}, \tag{1}$$

where S(w,X) are network statistics that typically include covariate effects on tie formations and network endogenous formation processes such as mutuality and transitivity. θ is a vector of parameters. K(θ,**W**) is a normalizing factor which ensures the probabilities sum to one.

$$K(\theta, \mathbf{W}) = \sum_{z \in \mathbf{W}} \exp\{\theta' S(z, X)\}, \tag{2}$$

where **W** represents all possible networks formed by the n nodes. The challenge of estimating ERGMs lies in that the normalizing factor is usually unknown.

It can be shown that an ERGM is equivalent to a conditional logit model (Hunter et al., 2008). The log odds of observing a tie conditional on the rest of the network (w_{ij}^c) is determined by changes in network statistics due to the presence of the tie (i.e., δ^{ij}(w,X)). Hence, the estimated coefficients in ERGMs can be interpreted as (conditional) log odds

$$\text{logit}\left[P\left(w_{ij} = 1 \middle| w_{ij}^c\right)\right] = \theta' \delta^{ij}(w, X) \tag{3}$$

Given an ERGM, the log likelihood of observing a network is as follows.

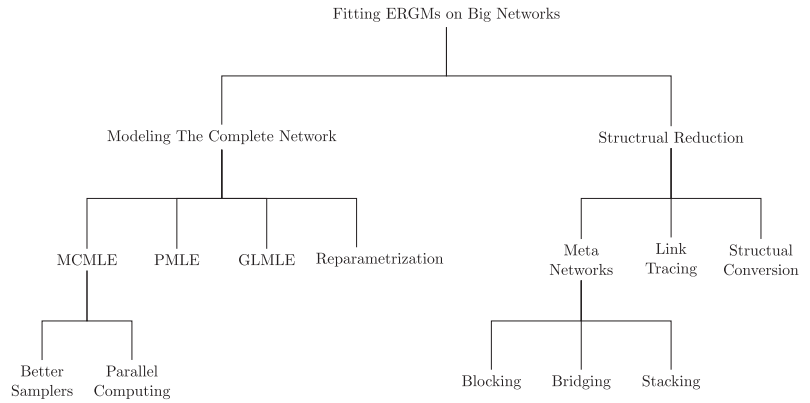
$$\ell(\theta) = \theta' S(w) - \log K(\theta, \mathbf{W}) \tag{4}$$

Now suppose we know θ_0. Consider the following log likelihood ratio.

$$\ell(\theta) - \ell(\theta_0) = (\theta - \theta_0)' S(w) - \log \frac{K(\theta, \mathbf{W})}{K(\theta_0, \mathbf{W})} \tag{5}$$

We can show the following.

$$\frac{K(\theta, \mathbf{W})}{K(\theta_0, \mathbf{W})} = \frac{\sum_{W \in \mathbf{W}} \exp\{\theta' S(W, X)\}}{K(\theta_0, \mathbf{W})} \tag{6}$$



Note: MCMLE refers to Monte Carlo maximum likelihood estimation, PMLE refers to Pseudo maximum likelihood estimation, and GLMLE refers to Graph limit maximum likelihood estimation.

Fig. 1. Methods for Fitting ERGMs on Big Networks.

$$= \sum_{W \in \mathbf{W}} \frac{\exp\{\theta_0' S(W,X)\}}{K(\theta_0, \mathbf{W})} \frac{\exp\{\theta' S(W,X))\}}{\exp\{\theta_0' S(W,X))\}} \tag{7}$$

$$= E_{\theta_0} \exp\{(\theta - \theta_0)' S(W,X)\} \tag{8}$$

Suppose we can sample networks (W_1,W_2,…,W_m) based on θ_0. Then,

$$\ell(\theta) - \ell(\theta_0) \approx (\theta - \theta_0)' S(w,X) - \log\left[\frac{1}{m}\sum_{i=1}^{m} \exp\{(\theta - \theta_0)' S(W_i,X)\}\right]. \tag{9}$$

Maximizing the above equation with respect to θ leads to the Monte Carlo Maximum Likelihood Estimation (MCMLE) of ERGMs (Hunter and Handcock, 2006). Sometimes, several versions of θ_0 have to be tried to help MCMLE converge.

To successfully use MCMLE we must be able to sample networks given a set of ERGM parameter values. The fact that an ERGM is equivalent to a conditional logit model suggests a Gibbs sampling scheme for sampling networks given θ_0.

1) At each step, randomly pick dyad (i,j).
2) Edge w_{ij} is set to one (or zero) according to equation (3).
3) Repeat steps (1) and (2).

Essentially, Gibbs sampling iteratively updates the ties according to the conditional probability $P(w_{ij} = 1|w^c)$ in order to produce samples for the joint probability $P(W = w)$. The process generates a Monte Carlo Markov Chain (MCMC) and with a sufficient number of updates, the chain will reach to the target distribution, from which samples can be drawn.[1]

One key feature of the MCMC-based sampling methods is that the subsequent state of a chain depends on its previous state. Hence, for large networks it will take a long time for the sampling process to reach the target distribution. This creates problems not only for model estimations, but also for model selections and assessment, as each task will take a long time (weeks or months depending on the size of the network) to conduct. Another problem for fitting ERGMs on big networks is that it typically requires big computer random-access memory (RAM) to load the data and carry out the estimations. This problem sometimes can also affect the strategies used to model big networks.

3. Fitting ERGMs on big networks

Fig. 1 lists two broad approaches to fitting ERGMs on big networks. One is modeling the complete network. The other is structural reduction, (i.e., partition the network in order to reduce the complexity and facilitate analyses). Generally speaking, the first approach will require more computer RAM while the second approach is easier for parallel computation and thus is faster. However, each approach includes several specific methods and each method may be more effective in different situations. Below I will outline the basics of these methods and discuss their advantages and disadvantages, respectively.

[1] Similarly spirited Metropolis and Metropolis-Hastings algorithms are also available (Hunter et al., 2008).

3.1. Modeling the complete network

3.1.1. Improving the MCMLE

One strategy to speed up the MCMLE is to design better proposals for updating the MCMC. For example, rather than picking a dyad randomly for toggling the tie status, a better proposal may be to pick the empty dyad with a small probability, like 1/2. Social networks typically have a low density. Picking dyads randomly often leads to flipping an empty dyad to a connected dyad. As a result, the MCMC usually stays in the null state for quite a long time. By reducing the probability of picking an empty dyad, the MCMC reduces its idle time and can quickly move to the connected parts of the network. This is the so-called "TNT" (tie-no tie) sampler used by default in "statnet" (Morris et al., 2008). The probability of picking an empty dyad may be further reduced to improve computation. But this method does not solve the central problem of updating a long MCMC.

Given the attractiveness of distributive computing and the wide use of multi-core computers, it is attempting to use parallel computation to facilitate sampling networks. Parallel computation can help reduce the sampling time once the MCMC reaches to the target distribution. This can be done by running multiple MCMCs with each chain sampling fewer number of networks. But the problem is that each chain will still need to go through a long burn-in period before reaching to the target distribution. Thus conventional parallel computation will not help much. What is really needed is parallel computation for speeding up one long MCMC. This is difficult to do because of the Markovian nature of the MCMC. But see Calderhead (2014) for some recent explorations in this regard.

Another method that has the potential to harness the power of parallel computing is to partition the big network into small subnetworks. Parallel MCMCs can be used to generate the subnetworks very quickly and then merge them into a big network. This way, it will be easier to sample big networks given ERGM parameter values. One limitation with this method is that the sampled big networks may differ somewhat from the ones produced holistically. But considering the gain in computation speed it may be worth sacrificing some accuracy.[2] This method is related to the meta network analysis to be described subsequently. Both of them divide a big network into parts to facilitate computation. But the former does so in the process of sampling networks while the later does so in the post-estimation stage where ERGM results from the subnetworks are combined.

3.1.2. Pseudo maximum likelihood estimation

Pseudo Maximum Likelihood Estimation (PMLE) (Strauss and Ikeda, 1990; Wasserman and Pattison, 1996) models the probability of a tie between two nodes i and j as a logit regression.[3]

$$\text{logit}\left[P\left(w_{ij}=1\middle|w_{ij}^{l}\right)\right]=\theta'\delta^{ij}(w,X), \tag{10}$$

where $\delta^{ij}(w,X)$ represents the changes in network statistics due to the presence of a tie. PMLE assumes ties are independent conditioning on local network structures w_{ij}^{l} while MCMLE conditions on the entire network except tie w_{ij}. Hence, PMLE can be viewed as a local approximation to MCMLE.

PMLE offers great computational convenience and speed for modeling large networks. However, if MCMLE estimates the true model, PMLE may under-estimate endogenous network formation processes and provide more conservative inferences (i.e., wider confidence intervals) on covariate effects (van Duijn et al., 2009).

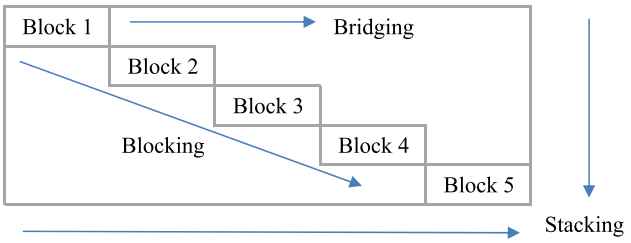
One approach to adjusting the seemingly incorrect standard errors in PMLE is to combine it with the quadratic assignment procedure (QAP) (Krackhardt, 1988). Basically, QAP is a nonparametric permutation procedure that randomly permutes the rows and columns of the adjacency matrix of the observed network. It conducts PMLE for each permutation and uses the resulting distribution of the estimates to test whether the estimates based on the original data are extreme or not. Note that this procedure is a testing procedure but not an inferential procedure. This is because the permutation-based P values condition on the null hypothesis that ties are randomly formed in the network. Hence, they may not correctly reflect the variability in re-sampling the network. Nonetheless, to be able to test the null hypothesis of random tie formation is something useful for big network analysis, especially when other methods do no work.

3.1.3. Graph limit maximum likelihood estimation

Recent research has also started to use graph limits to help estimate ERGMs (Chatterjee and Diaconis, 2013; He and Zheng, 2015). Basically, a big network can be viewed as a random network in which each tie is generated independently according to a probability defined by the graph limit of the network. For an ERGM with given parameter values, there is a graph limit obtainable. Once the graph limit is known, the normalizing factor can be approximated. Then the probability of the observed network is known. Varying the parameter values (e.g., by a grid search) will lead to different graph limits and different estimates of the probability. The set of parameter values that maximize the probability of the observed network is the graph limit based maximum likelihood estimate (GLMLE) of the ERGM.

[2] The idea of this method is presented only conceptually in this paper. To implement and evaluate the method will require a separate study.

[3] "PMLE" seems to be a better acronym than the conventionally used "MPLE", because it is similar to other MLEs like MCMLE and GLMLE.



Note: A big network is divided into blocks. Blocking fits ERGMs on the network in each block. Bridging fits ERGMs on ties across each pair of blocks. Stacking combines the estimates from the blocking and the bridging steps.

Fig. 2. Meta Network Analysis for Big Networks.

Interesting as it is, the GLMLE faces several constraints. First, it is mainly applicable to dense and symmetric networks. But social networks tend to be sparse and include a significant portion of asymmetric ties.[4] Second, currently GLMLE is only applicable to cases where graph limits exist and a certain range of ERGM parameter values (e.g., a nonnegative coefficient on the triangle count). Third, the transformation from ERGM network statistics to their graph limit representations is not intuitive. In particular, the method needs to be extended to more general ERGMs such as those including covariates effects and high orders of network dependence. Last, the variance of the GLMLE needs to account for the uncertainty in approximating the normalizing factor.

3.1.4. Reparametrization

It is possible to reparametrize ERGMs or use some alternative and simpler models to study network formation. For example, traditional regressions may be used to examine the correlations between covariates and selected network measures like indegree, outdegree, and transitivity. The regression models are simple to understand and estimate, but they face two constraints. One is that they usually do not account for endogenous network formation processes. One way to remedy this is to include related network measures in the regressions. For example, when predicting indegree centrality, outdegree, mutuality (the number of mutual ties a node is in), and transitivity (the number of triangles a node is in) can be added to the regression model in order to account for these mechanisms. The other constraint is that simple regressions do not account for correlations in the outcomes across nodes. Possible remedies are to use permutation *P* values or bootstrapped standard errors.

Chandrasekhar and Jackson (2014) develop a new class of network formation models, Statistical Exponential Random Graph Models (SERGMs) in order to simplify the ERGM estimation. SERGMs formulate networks in terms of sufficient statistics instead of networks. More formally, they model $P(S = s)$ rather than $P(W = w)$, where S are vectors of sufficient network statistics. SERGMs have lower dimensionality and so are easier to estimate. Since the method is still work-in-progress, I only list the reference here for interested readers without going into details.

3.2. Structural reduction

3.2.1. Meta network analysis

When networks display a block structure where the majority of ties occurring within groups than across groups, meta network analysis (a "divide and conquer" strategy) can be very useful. The method proceeds with three steps. See Fig. 2 for an illustration.

1) Blocking. An ERGM is fitted on the network in each block. Network statistics can be included to account for the effects of ties formed outside of the block.
2) Bridging. An ERGM is fitted on ties between two blocks. Network statistics can be included to account for the effects of ties formed inside the blocks.
3) Stacking. The ERGM results from the blocking and bridging steps are combined via meta analysis (Snijders and Baerveldt, 2003; Gasparrini et al., 2012; An, 2015).

The approach of meta network analysis has several advantages. First, it is easier for parallel computing due to partition of the data. As a result, it will be much faster. Second, meta network analysis assumes within-block dependence (i.e., ties within block can be dependent in various forms) and weak cross-block dependence (i.e., ties across blocks are only weakly dependent). The assumption is in between the Markov dependence assumption by PMLE and the full network dependence by MCMLE. Therefore, there is no significant loss in the modeling of network dependence. Third, this approach is useful for

[4] See Bollobás and Riordan (2011) for some explorations of the method for sparse networks.

assessing the effect homogeneity assumption in ERGMs namely, the ERGM parameter values are assumed to be the same in all parts of a network (Holland and Leinhardt, 1981; Lusher et al., 2013). This is problematic when the network is big. By fitting ERGMs on subnetworks, meta network analysis provides us with one way to examine the variability of the estimates in different parts of the network.

3.2.2. Link tracing

Many big networks tend to exhibit a core-periphery or polycentric structure. In these situations, it makes more sense to model the core part of the network rather than the overall network To do so, we need to select a set of central nodes, as most of the connections are from/to the central nodes. There is no strict way to select central nodes. Sometimes ERGMs will fit well when just the isolates are removed (Goodreau et al., 2008). Other times, it works if nodes with centrality (e.g., indegree) above a certain threshold are selected. Still other times, nodes with a certain attribute (e.g., papers by star scientists in citation networks) may be selected. Regardless how central nodes are selected, the foremost consideration in this strategy is that the selected central nodes should make it easier to fit ERGMs while the features of the original network are preserved as much as possible.

Once central nodes are selected, subnetworks composed of these central nodes and their neighbors can be constructed. This is the so-called link tracing. If only the central nodes and their neighbors are included, it is the 1st degree link tracing. If the central nodes' neighbors' neighbors are also included, it is the 2nd degree link tracing, etc. There are two ways to analyze the resulting subnetworks. One is to combine the subnetworks into a single network and fit ERGMs on this core network. For very large networks, sometimes the core network itself can be too large to fit ERGMs. Then ERGMs may be fitted on each of the subnetworks that is centered on one central node and the results can be combined via meta analysis. Since in very large networks, the subnetworks are less likely to overlap with one another, the method should work reasonably well in many cases.

Prior research (Shalizi and Rinaldo, 2013) has argued that ERGM estimates based on sampled networks may not be consistent. This may be true in theory. But Handcock and Gile (2010) show that in link tracing designs, as the sample size increases, the information contained in the samples approaches that of the complete network and only a modest amount of information is lost. They also show that the ERGM estimates in the samples are approximately unbiased and the efficiency losses are small. They also notice that in small samples, there is a great amount of variability in the ERGM estimates across samples. Overall, link tracing helps provide us with a good understanding of tie formations in the core part of the network.

3.2.3. Structural conversion

Structural conversion aims to convert a big network to a smaller one by aggregating the nodes to higher levels or by converting the relationships to ones that are manageable. For example, school friendship ties may be aggregated to class level so as to study friendship connections across classes. Paper citation networks may be converted to journal citation networks or co-author networks. Since the number of journals or authors are usually much smaller than the number of papers, ERGMs are easier to fit on the converted networks. As it can be seen, structural conversion usually changes the original research question and so may not be desired in some cases.

4. Examples

Since some of the methods covered in this paper like GLMLE and SERGMs are still under development and some other methods like traditional regressions are last-resort-methods, I focus on demonstrating the more mature methods such as MCMLE, PMLE, meta network analysis, and link tracing. The first two examples are based on large friendship networks in China and the U.S. In these cases, it is still possible to fit ERGMs on the complete networks through MCMLE, although the process is very slow. The next two examples are based on super-large networks. One is a citation network in a scientific field composed of over 30 thousand papers and about 10 thousand citations. The other is the Facebook network (mostly a thought example). In these cases, it is generally impossible or undesirable to estimate ERGMs through MCMLE due to various concerns.

4.1. A school friendship network in China

The first dataset is drawn from a middle school in China. The school has three grades (7, 8, and 9), 18 classes (six for each grade), and 783 students. The class size ranges between 32 and 54 with a mean of 44. The friendship network data comes from a survey conducted by the author in 2010, in which the students are asked to report up to ten of their closest friends in the school. The school friendship network displays a clear block structure. On average, the density of within-class friendship networks is 11% while the density of cross-class friendship networks is only 1% (see Table A1 in the appendix). The survey also collects a list of covariates of the students, including sex (boys = 1; girls = 0), academic ranking (top ten in class = 1; otherwise = 0), smoking status (yes = 1; no = 0), and family economic condition (good = 1; otherwise = 0).

In total, I fit four models on the friendship network. Model 1 is fitted on the complete school friendship network and estimated by MCMLE. It serves as the baseline model.[5] Receiver effect in the model indicates the effect of a covariate on

[5] I also tried other models, for example, by including GWDSP (Geographically weighted dyad-wise shared partners) to model whether ties are more likely to run across two-paths without closing a triangle and GWODEDREE (Geographically weighted outdegree distribution). The MCMLE fit poorly in these models.

receiving friend nominations. Sender effect indicates the effect of a covariate on making friend nominations. Homophily effect indicates the tendency for ties to form between students with the same characteristics as compared to those with different characteristics. The model includes receiver effects and sender effects for sex and smoking status, receiver effect for academic ranking, sender effect for family economic condition, and homophily effects for the four covariates and whether two students are from the same class. Additionally, the model also includes several variables to account for typical endogenous network formation processes. Mutuality indicates whether reciprocal ties are more likely to form. GWESP (Geographically weighted edge-wise shared partners) indicates whether transitive ties are more likely to form (i.e., friends of friends are also friends). GWIDEGRE (Geographically weighted in-degree distribution) is used as a measure of degree assortativity (Hunter, 2007). A positive coefficient indicates a tie is more likely to occur, especially between students with low indegrees than between students with high degrees. A negative coefficient indicates a tie is more likely to be absent, especially between students with low indegrees than between students with high degrees.[6] Hence, a negative coefficient often indicates the network exhibits a core-periphery structure (Lusher et al., 2013). The last term "edges" accounts for the baseline connectivity in the network.

The second model includes the same variables as in model 1, but is estimated by PMLE. Besides conventional *P* values, I also provide *P* values based on 1000 permutations of the adjacency matrix of the school friendship network.

Model 3 is based on meta network analysis. In the "blocking" step, I fit the baseline ERGM on the friendship network in each of the 18 classes. To account for the influence of external friendships, the number of friend nominations received by each student from other classes is added to the baseline model. In the "bridging" step, I fit the baseline model on the friendship network between each pair of the 18 classes. There are $\binom{18}{2} = 153$ such networks in total. To account for the influence of within class friendships, I add the number of friend nominations received by each student from their own class to the model. To account for the fact that only cross-class friendships are allowed in these networks, I offset the coefficient for within-class ties to be negative infinity. To account for the overall sparsity of the networks (namely, not many students have cross-class friendships), I also include the number of isolates as a parameter in the model. In the "stacking" step, I combine the results from the previous two steps through meta analysis (Snijders and Baerveldt, 2003), where each estimated coefficient is weighted in proportion to its estimated variance. For conciseness, the results in the stacking step is shown in the main text while the results in the blocking and bridging steps are shown in the appendix (Table A3).

Model 4 is to model the core part of the school friendship network. I first select as the central nodes the top five students who received the most friend nominations ($\geq$24 friends) in the school. Then I conduct a 1st degree link tracing by also including these students' friends. This leads to 141 students and 709 friendship ties sampled. In other words, 18% of the students and 11% of the friendship ties are selected into the core network for analysis. The baseline model is fitted on this core network.

All models are estimated by (partly) using "statnet" (Handcock et al., 2003) in R. To facilitate model estimations, the decay parameters in GWESP and GWIDEGREE have been fixed at 0.2.[7] To account for the cap on maximal friend nominations (i.e., $\leq$10), the MCMCs are constrained to sample only networks meeting that requirement. All model estimations have been checked for convergence. The MCMC in model 1 has mixed reasonably well but the final estimates seem not to be able to produce networks that are statistically similar to the observed one. This is true even after tuning the MCMC algorithms through multiple ways. The (slow and painful) experience illustrates the general difficulties in fitting ERGMs on big networks using MCMLE. Among the ERGMs fitted on the 18 within-class friendship networks, 11 of them converged. However, among the ERGMs fitted on the 153 cross-class friendship networks, only 13 of them converged. This may be because most of the cross-class friendship networks are extremely sparse. For conducting meta analysis, I only include the ERGMs that have converged. More results of model diagnostics are available upon request.

Table 1 presents the results of the models. The shown coefficients are log odds ratios. For example, Model 1 shows that the odds of a tie directed to boys is 1.42 ($e^{0.35} = 1.42$, $P < 0.01$) times that of a tie directed to girls. But the odds of a tie sent by boys is only 0.63 ($e^{-0.47} = 0.63$, $P < 0.01$) times that of a tie sent by girls. Students with a high academic ranking receive more friendship nominations than those with a low ranking (odds ratio: $e^{0.26} = 1.3$, $P < 0.01$). Smokers receive more nominations (odds ratio: $e^{0.3} = 1.35$, $P < 0.01$) but make fewer nominations (odds ratio: $e^{-0.31} = 0.73$, $P < 0.01$). There are significant homophily effects on sex, ranking, smoking, and class. There is also significant evidence for friendships to be mutual and transitive. Roughly speaking, the odds of a mutual tie occurring is about seven times ($e^{1.95} = 7.03$, $P < 0.01$) that of an asymmetric tie. A tie that closes a triangle is over eight times as likely to occur as a tie that does not close a triangle ($e^{2.14} = 8.50$, $P < 0.01$).

To facilitate comparisons of the estimates across models, Fig. 3 shows the estimated coefficients for selected variables and their 95% confidence intervals (CIs). The estimates of MCMLE, PMLE, meta network analysis (stacking), and link tracing are colored in red, brown, purple, and green, respectively. Several patterns are clear. First, the covariate effects (including both main effects and homophily effects) are mostly statistically indistinguishable across the models. The point estimates are close and the 95% CIs overlap quite a bit. The notable difference is that the link tracing estimates generally have wider 95% CIs, probably because they rely on partial samples of the network.

[6] This statistics is also viewed as a measure of indegree heterogeneity (Papachristos et al., 2013).

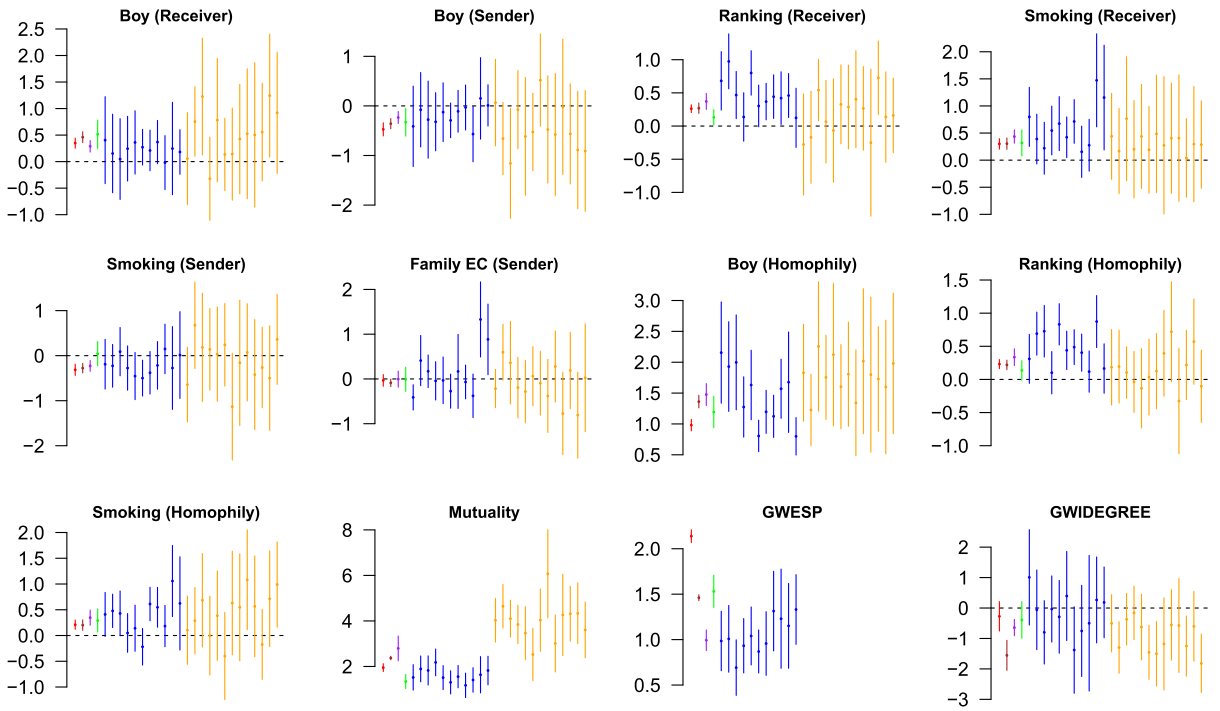
[7] The decay parameter controls the contributions of additional counts of the same network structure to the corresponding network statistics. Small decay parameters help model convergence (Goodreau et al., 2008).

Table 1
ERGM results for the school friendship network in China.

	A. Modeling the complete network							B. Structural reduction						
	Model 1			Model 2				Model 3				Model 4		
	MCMLE			PMLE				Meta network analysis				Link tracing		
	Est.	SE	P	Est.	SE	P	MP	Est.	SE	P	Q	Est.	SE	P
Main effects														
Boy (Receiver Effect)	0.35	0.05	0.00	0.46	0.05	0.00	0.00	0.29	0.05	0.00	0.89	0.52	0.13	0.00
Boy (Sender Effect)	−0.47	0.07	0.00	−0.36	0.05	0.00	0.00	−0.23	0.06	0.00	0.78	−0.33	0.14	0.02
Ranking (Receiver Effect)	0.26	0.03	0.00	0.27	0.04	0.00	0.00	0.37	0.06	0.00	0.07	0.13	0.06	0.02
Smoking (Receiver Effect)	0.30	0.05	0.00	0.31	0.05	0.00	0.00	0.44	0.06	0.00	0.63	0.32	0.12	0.01
Smoking (Sender Effect)	−0.31	0.06	0.00	−0.28	0.05	0.00	0.00	−0.23	0.06	0.00	0.83	0.04	0.14	0.75
Family EC (Sender Effect)	−0.03	0.07	0.62	−0.09	0.04	0.04	0.15	0.00	0.09	0.96	0.00	−0.01	0.14	0.95
Homophily														
Boy	0.98	0.05	0.00	1.36	0.05	0.00	0.00	1.48	0.09	0.00	0.00	1.19	0.13	0.00
Ranking	0.23	0.03	0.00	0.22	0.04	0.00	0.00	0.34	0.06	0.00	0.00	0.14	0.07	0.07
Smoking	0.21	0.05	0.00	0.20	0.05	0.00	0.00	0.35	0.07	0.00	0.03	0.29	0.11	0.01
Family Econ Condition	0.07	0.04	0.05	0.08	0.04	0.03	0.04	0.00	0.06	0.94	0.01	0.09	0.10	0.37
Class	1.02	0.03	0.00	1.45	0.03	0.00	0.00					0.97	0.06	0.00
Network structures														
Mutuality	1.95	0.09	0.00	2.37	0.04	0.00	0.00	2.79	0.27	0.00	0.00	1.34	0.15	0.00
GWESP	2.14	0.04	0.00	1.46	0.02	0.00	0.00	0.99	0.06	0.00	0.31	1.53	0.09	0.00
GWIDEGREE	−0.27	0.24	0.26	−1.55	0.25	0.00	0.00	−0.64	0.13	0.00	0.13	−0.39	0.30	0.19
Edges	−6.96	0.08	0.00	−7.31	0.08	0.00	0.00	−5.97	0.20	0.00	0.00	−6.28	0.19	0.00
Total time (seconds)	45,247			3.95			3426	3668				197		
Average time (seconds)	45,247			3.95			3.426	77.89				197		

Note: The "Q" column shows the *P* values for testing effect homogeneity across the networks.

Second, the estimated endogenous network formation processes are also similar, although with some variations across the models. It appears that PMLE and meta network analysis overestimate mutuality while link tracing under-estimate it. Overall, however, all the estimates show statistically significant mutuality in the friendship ties. Also, in comparison to MCMLE, PMLE,



Note: Estimates of MCMLE, PMLE, stacking, link tracing, blocking and bridging are colored in red, brown, purple, green, blue, and orange, respectively.

Fig. 3. Comparing Estimates on Selected Variables in the ERGMs for the School Friendship Network in China.

meta network analysis, and link-tracing under-estimate transitivity (GWESP). This is expectable. PMLE is a local approximation to MCMLE and so naturally under-estimates transitivity. Meta network analysis and link-tracing use the partial network to estimate the parameters and so may have lost some information on transitivity too. However, all the estimates agree three is statistically significant transitivity in the friendship ties. In term of GWIDEGREE, all the estimates suggest there is negative degree assortativity, except that the PMLE estimate is more negative and turns out to be statistically significant.

Third and probably more importantly for the purpose of this study, there are big differences in the time used for estimating the models (shown in the bottom of Table 1). MCMLE takes the longest time (45,247 seconds) to fit, which is over 10 thousand times longer than PMLE (less than four seconds). Obtaining permutation-based *P* values for PMLE takes more time, requiring about an hour to complete. Meta network analysis also takes about an hour to fit. Link tracing takes about three minutes to fit. The permutation PMLE and meta network analysis are easily adaptable to parallel computation. If paralleled, roughly speaking, the average time is 3.43 seconds for the former and about 80 seconds for the latter. So these two methods strike a good balance between speed and accuracy (without losing much dependence in tie formations).

Last, meta network analysis offers distinct information that can be used to assess effect homogeneity in the network. The "Q" column in Table 1 shows the *P* values testing whether the estimated coefficients are the same across the different subnetworks. About half of the estimated coefficients are statistically indistinguishable across the subnetworks while the other half (notably homophily effects and mutuality) vary significantly across the subnetworks. The last two sets of estimates in Fig. 1 present the estimated coefficients and their 95% CIs in the "blocking" and "bridging" steps (colored in blue and orange, respectively). Such information helps provide further assessment of effect homogeneity for the subnetworks within or across classes. In addition, Table A3 in the appendix shows that one additional outside-class tie is associated with an increase of within-class ties by about 1% ($e^{0.01} = 1.01$, $P < 0.01$) and one additional within-class tie is associated with an increase of outside-class ties by about 2% ($e^{0.02} = 1.02$, $P < 0.01$). Thus, having these terms in the ERGMs helps account for cross-block tie dependence.

4.2. A school friendship network in the U.S.

The second example is based on a school friendship network from the AddHealth Study (Resnick et al., 1997; Handcock et al., 2003; Goodreau et al., 2008). The students are in junior or senior high schools in the same community in the southern U.S. Only mutual friendship ties are available in the data. The network includes 1461 students and 974 mutual friendship ties. Available student covariates include sex (male or female), race (for simplicity recoded as Black, White, and Others), and grade (7–12).

The school friendship network in the U.S. differs somewhat from that in China. In the U.S., students from different classes can take courses together and interact with one another actively whereas in China students mostly learn and interact with other students in the same classroom. That said, the school friendship network in the U.S. also exhibits a strong block

Table 2
ERGM results for the school friendship network in the U.S.

	A. Modeling the complete network						B. Structural reduction						
	Model 1			Model 2			Model 3				Model 4		
	MCMLE			PMLE			Meta network analysis				Link tracing		
	Est.	SE	P	Est.	SE	P	Est.	SE	P	Q	Est.	SE	P
Main effects													
Boy	−0.12	0.05	0.02	−0.05	0.04	0.20	−0.11	0.08	0.25	0.10	−0.07	0.13	0.58
Race (Black)													
Others	0.21	0.10	0.04	0.35	0.10	0.00	0.17	0.11	0.17	0.33	1.34	0.52	0.01
White	−0.22	0.06	0.00	−0.11	0.07	0.11	−0.36	0.04	0.00	0.95	1.08	0.57	0.06
Grade (7)													
8	0.08	0.07	0.25	0.10	0.07	0.16					−0.12	0.15	0.40
9	−0.23	0.07	0.00	−0.25	0.07	0.00					−0.40	0.23	0.08
10	−0.17	0.07	0.02	−0.12	0.07	0.08					−0.34	0.18	0.06
11	−0.08	0.07	0.28	−0.07	0.07	0.28					−0.17	0.15	0.26
12	−0.05	0.07	0.48	−0.01	0.08	0.94					−0.26	0.20	0.19
Homophily													
Sex	0.72	0.08	0.00	0.80	0.07	0.00	0.81	0.06	0.00	0.53	0.17	0.17	0.31
Race	1.23	0.10	0.00	1.22	0.10	0.00	1.30	0.07	0.00	0.69	0.64	0.43	0.14
Grade	2.76	0.10	0.00	2.86	0.09	0.00					2.53	0.26	0.00
Network structures													
GWESP	1.71	0.09	0.00	1.62	0.03	0.00	1.45	0.10	0.00	0.21	1.51	0.18	0.00
GWDSP	−0.31	0.05	0.00	−0.35	0.03	0.00	−0.56	0.06	0.00	0.13	−0.58	0.09	0.00
GWDEGREE	−1.01	0.16	0.00	−1.07	0.09	0.00	−0.72	0.33	0.07	0.05	0.16	0.39	0.68
Edges	−7.79	0.29	0.00	−7.97	0.22	0.00	−5.51	0.33	0.00	0.11	−6.68	1.11	0.00
Total time (seconds)	1728			2.27			9120				78.63		
Average time (seconds)	1728			2.27			1566				78.63		

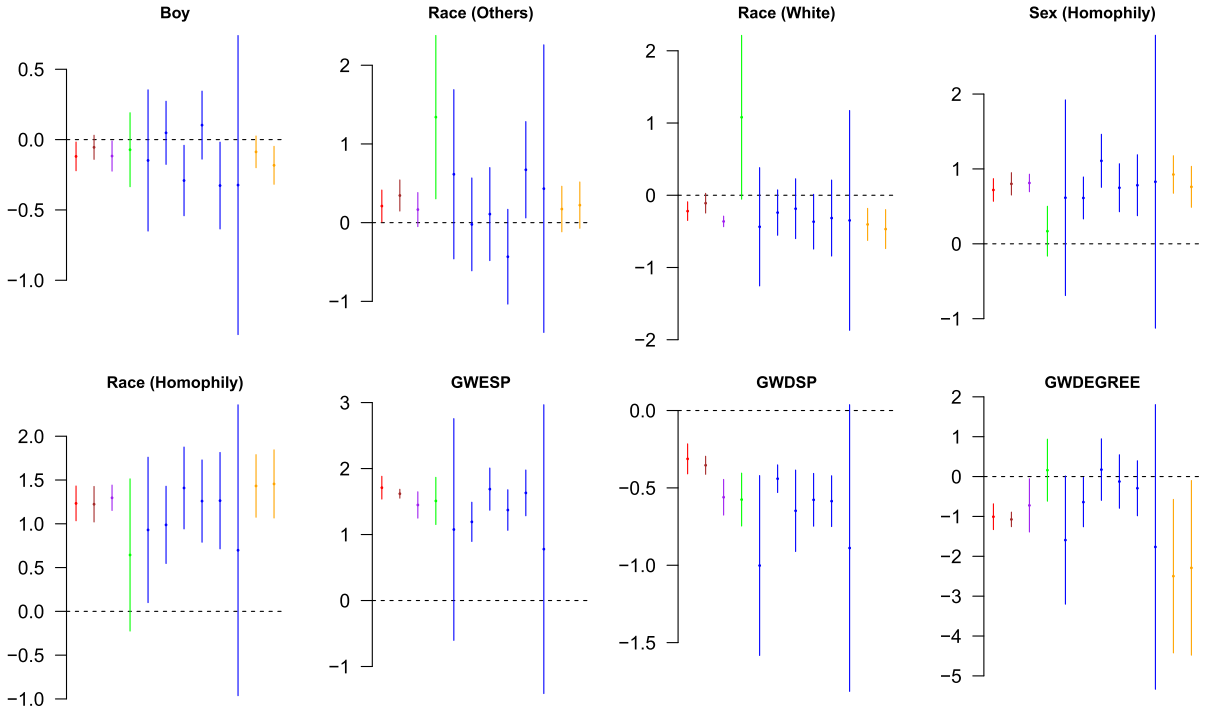
Note: The "Q" column shows the *P* values for testing effect homogeneity across the networks.

structure, although at the grade level instead of at the class level. As shown in Table A2 in the appendix, the densities of the within-grade friendship networks range between 0.15% and 0.38% with a mean of about 0.24%, whereas the maximal density of the friendship networks across grades is only 0.06% and most of the densities are close to zero. Like in the case of the school friendship work in China, this feature facilitates the use of such methods as meta network analysis and link-tracing.

Table 2 presents the results of four models that are fitted on the friendship network, including MCMLE, PMLE, meta network analysis, and link tracing. In meta network analysis, the model is fitted on each of the friendship networks in the six grades and two of the cross-grade friendship networks and the results are combined via meta analysis.[8] In link tracing, students with at least 8 friends (the 95% percentile of degree centrality) and their friends are selected to construct the core network.

For easier comparisons of the results, Fig. 4 presents the estimated coefficients and their 95% CIs. The estimates of PMLE, meta network analysis, and link tracing are statistically indistinguishable from those of MCMLE, except two differences. First, meta network analysis provides a significantly lower estimate of GWDSP than MCMLE, namely, ties forming open triangles are less likely to occur. This is expectable because meta network analysis in this case mostly relies on within-class friendship networks where ties are presumably less likely to form open triangles. Despite the slight difference in the estimates, both meta network analysis and MCMLE agree that friendship ties in general are less likely to run across two paths without closing a triangle. Second, the estimates from link tracing generally have wider confidence intervals. This is probably also because link tracing relies on the partial network. Also, as compared to MCMLE, the estimates of link tracing suggest there is less homophily in sex and less degree assortativity (as indicated by the coefficient for GWDEGREE) in friendship formations. Both differences are statistically significant at the 5% significance level. The differences suggest that the patterns of friendship formations in the core network may differ from those in the complete network in these two dimensions.

In this example, the computational time also varies quite a bit by the estimation methods. The MCMLE takes about half an hour to complete while the PMLE and the link tracing only take about 2.27 seconds and 78.63 seconds to finish, respectively. Meta network analysis takes the longest to complete, about 2.5 hours. Thus meta network analysis seems not providing any computational advantage. This is probably because some of the subnetworks have peculiar features that slow down the ERGM



Note: Estimates of MCMLE, PMLE, stacking, link tracing, blocking and bridging are colored in red, brown, purple, green, blue, and orange, respectively.

Fig. 4. Comparing Estimates on Selected Variables in the ERGMs for the School Friendship Network in the U.S.

[8] For the six grades, there are 15 cross-grade friendship networks. But most of these friendship networks are so sparse that no ERGMs can be stably fitted.

estimations and also because we stop the ERGM estimation on the complete network early (once the specified number of iterations are exhausted). However, if properly paralleled, the time of meta network analysis can be reduced to less than half an hour, which is comparable to MCMLE on the complete network. In addition, as before, meta network analysis provides unique information that can be used to assess effect homogeneity across the parts of the network. In this case, we find there is no much variation in the estimates across the subnetworks, as indicated by the large P values shown in the "Q" column in Table 2. We also find positive and statistically significant correlations in friendship ties within and cross grades (Table A4). Further such information can be obtained, for example, by examining the estimates in the blocking and bridging steps as shown in Fig. 4.

4.3. Examples based on super-large networks

When fitting ERGMs on large networks through MCMLE, it can take a long time to complete the estimation. Most analyses will also require many trials to select appropriate model specifications and assess the model convergence. The whole process can take hours, days, or weeks to complete depending on the size of the network and the model specifications. When the network size is super-large, in general it is infeasible or undesirable to fit ERGMs through MCMLE. One issue is related to hardware constraints. To load big network data and carry out estimations through MCMLE usually requires big computer memory (typically over several hundred gigabytes) that most personal computers cannot afford (typically between 16 and 64 gigabytes). One way to address the memory issue is to use super computers. As such computational resource is scarce, it often takes a long time to locate appropriate supercomputers and get the required work scheduled and conducted.

Another issue with super-large networks is related to software constraints. "statnet" in R has been the standard software for fitting ERGMs. R utilizes a special form of long vectors to store big network information. However, two base languages C and Fortran that "statnet" relies on for big network analysis do not accept such "long vectors". Thus even if data can be loaded into "statnet", estimations of ERGMs on super-large networks often fault for this reason.

Third, even if there are no hardware or software issues, to estimate ERGMs on super-large networks usually takes weeks or months to finish. Whether this is desirable or not depends on the specific research purpose. If the research purpose is to get the most accurate estimates, then probably it is worth the time. Otherwise, other methods presented in this paper can be useful to provide reasonable estimates in a much faster fashion.

An and Ding (2016) use ERGMs to study the citation network in the field of causal inference. The citation network is composed of 33,306 publications and 8442 citations in the field. To simply load the data into software requires at least 250 gigabytes of computer RAM. The authors use supercomputers provided by their institution to solve the computer memory problem. However, a long vector representing no forward referencing (i.e., papers published earlier cannot cite those published later) cannot be passed from R to Fortran and so prevents the estimations from proceeding normally. Finally, the authors are able to fit a simple ERGM with only covariate effects (without the aforementioned long vector) through PMLE on the complete network. The estimation takes about an hour to complete. The same ERGM is fitted on the core citation network (excluding 28,537 isolates) and the estimation takes four minutes to complete. The results are broadly similar across the two sets of estimates. Also with the reduced network size, a more complex ERGM that accounts for endogenous network formation processes can be fitted on the core network. The estimation is feasible through PMLE and takes about half an hour to finish. Thus by combining structural reduction and PMLE, the authors are able to provide insights into the citation patterns in the field of causal inference, although the authors also caution that the results from the core network may not be immediately applicable to the full network.

When the network size gets even bigger, even supercomputers (usually with up to 500 gigabytes of RAM) may not be able to load the data and carry out the estimations. In these cases, currently the most feasible way is to use some of the reduction methods to model the network. For example, suppose the interest is to model the friendship network of American college students on Facebook. The network is super large, composed of maybe 20 million nodes and millions of friendship ties. Given our understanding of friendship formations, it is reasonably to assume that friendship ties are much more likely to form among students at the same institution than those at different institutions. Hence, the Facebook network may likely follow a block structure too. As before, we can use meta network analysis to divide the big network into small parts (i.e., friendship networks within and across institutions) and fit ERGMs on the subnetworks and combine the results via meta analysis. This strategy will allow us to have an overall understanding of friendship patterns on Facebook and also more nuanced understandings of different patterns of friendship formations across institutions. Another possible approach is through link tracing. We can sample central students from (selected) institutions and construct core networks or egocentric networks surrounding these central nodes. Then we fit ERGMs on these subnetworks and combine them via meta analysis. In short, when the network size gets so big that it becomes infeasible to fit ERGMs on the complete network, methods that utilizes structural reduction may come to the rescue.

5. Summary and discussion

There are both conceptual and computational problems for fitting ERGMs on big networks. First, the full information assumption held by MCMLE becomes problematic in big networks. In those cases it is unrealistic to require actors to know the connections among all others when they are tying to make any connections. Second, there are possible

Table 3
Comparing methods for fitting ERGMs on big networks.

Methods	Features	Advantages	Disadvantages	Ratings
A. Model the complete network				
1. Monte Carlo maximum likelihood estimation (MCMLE)	Sample networks using MCMC and optimize the likelihood to obtain MLE.	Standard approach to obtain the MLE for the complete network.	Slow and often fail to converge. Better samplers and parallel computation may help.	B
2. Pseudo maximum likelihood estimation (PMLE)	Assume ties are conditionally independent. Local approximation to MCMLE.	Fast, stable, and parallelable.	May under-represent higher orders of tie dependence.	A
3. Graph limit maximum likelihood estimation (GLMLE)	Use graph limits to approximate the normalizing factor and then optimize.	Relatively fast and stable.	Currently only applicable to dense networks and particular ERGMs.	B+
4. Reparametrization	Model network statistics instead of the network.	Fast and stable.	Model network features rather than networks. Under-represent tie dependence.	B+
B. Structural reduction				
5. Meta network analysis	Fit ERGMs on subnetworks and combine the results via meta analysis.	Fast and stable. Help examine effect homogeneity.	May under-represent higher orders of tie dependence.	A
6. Link tracing	Fit ERGMs on the core network or sampled networks.	Fast and usually stable.	Inferences are limited to the partial network. Under-represent tie dependence.	A−
7. Structural conversion	Convert ties to different levels or types to simplify the network.	Fast and stable.	Often change the original research question.	B

hardware and software constraints. To load big networks and carry out the estimations requires big computer RAM that usual desktops cannot afford. Even if the data can be loaded, the current software for network analysis may not be able to process long vectors used to store big network information. Third, due to its reliance on MCMC, the MCMLE that has been typically used for estimating ERGMs can be notoriously slow in big networks.These issues make it challenging to fit ERGMs on big networks.

The general solution is to strike a balance between accuracy and speed. MCMLE usually provides a good approximation to the likelihood of observing a network. But it is very slow in big networks. In comparison, PMLE, GLMLE, meta network analysis, and link tracing provide faster estimations of ERGMs. When accuracy is of primary interest and MCMLE is feasible, then MCMLE should be used, maybe along with PMLE and other methods so that the evidence can be triangulated for robust inference. When MCMLE is infeasible or speed is of primary concern, then PMLE or a combination of structural reduction and MCMLE/PMLE may be preferred. Among the structural reduction methods, meta network analysis is applicable to networks with a clear block structure. By assuming within-block dependence and weak cross-block dependence, meta network analysis imposes an assumption that is in-between the Markov dependence by PMLE and the full network dependence by MCMLE. Also, meta network analysis can provide information useful for examining variations in the ERGM estimates across parts of the network. This is valuable, because it is unrealistic to assume that a single set of estimates are applicable to the whole network, especially when the network is big. When the network follows a core-periphery or polycentric structure, link-tracing can be a useful method for learning the features in the main part of the network. For more comparisons of the methods, please see Table 3, where I have listed the advantages and disadvantages of the different methods.

The empirical examples in this paper show that PMLE, meta network analysis, and link tracing are able to estimate covariate effects in a way consistent with that by MCMLE. The 95% confidence intervals of the estimated covariate effects overlap quite a bit across these methods. In addition, the statistical inferences on the endogenous network formation processes are qualitatively similar across these methods. The results also help provide a sense of potential biases in the estimates. For example, PMLE and meta network analysis may under-estimate transitivity and overestimate mutuality and degree assortativity, as compared to MCMLE. It is worth pointing out again that the estimation time can differ greatly across these methods. Sometimes MCMLE can be over ten thousand times slower.

In short, the paper outlines a framework and a series of methods for fitting ERGMs on big networks. Future research may focus on developing faster MCMC samplers for the MCMLE, designing smart parallel algorithms for mixing a long Markov chain, generalizing GLMLE, etc. More work in this area (both simulative and empirical) is valuable.

Acknowledgments

The author thanks the Graduate School of Arts and Sciences, the Multidisciplinary Program in Inequality and Social Policy, the Fairbank Center for Chinese Studies, and the Institute for Quantitative Social Science, all at Harvard University, for the financial support for collecting the data used in this paper. The author also thanks Professor Stephanie Moller and anonymous reviewers for their helpful comments.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ssresearch.2016.04.019.

References

An, Weihua, 2015. Multilevel meta network analysis with application to studying network dynamics of network interventions. Soc. Netw. 43, 48–56.
An, Weihua, Ding, Ying, 2016. The landscape of causal inference: perspectives from the citation networks. Working Paper. Indiana University.
Bollobás, Béla, Riordan, Oliver, 2011. Sparse graphs: metrics and random models. Random Struct. Algorithms 39 (1), 1–38.
Calderhead, Ben, 2014. A general construction for parallelizing metropolishastings algorithms. Proc. Natl. Acad. Sci. U. S. A. 111 (49), 17408–17413.
Chandrasekhar, Arun G., Jackson, Matthew O., 2014. Tractable and consistent random graph models. NBER Working Paper 20276. http://www.nber.org/papers/w20276.
Chatterjee, Sourav, Diaconis, Persi, 2013. Estimating and understanding exponential random graph models. Ann. Statistics 41 (5), 24282461.
Frank, Ove, Strauss, David, 1986. Markov graphs. J. Am. Stat. Assoc. 81, 832–842.
Gasparrini, A., Armstrong, B., Kenward, M.G., 2012. Multivariate meta-analysis for non-linear and other multi-parameter associations. Statistics Med. 31 (29), 3821–3839.
Goodreau, Steven M., Handcock, Mark S., Hunter, David R., Butts, Carter T., Morris, Martina, 2008. A statnet tutorial. J. Stat. Softw. 24 (9), 1–26.
Handcock, Mark S., Gile, Krista J., 2010. Modeling social networks from sampled data. Ann. Appl. Statistics 4 (1), 525.
Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., Morris, M., 2003. Statnet: Software tools for the statistical modeling of network data. http://statnetproject.org.
He, Ran, Zheng, Tian, 2015. GLMLE: graph-limit enabled fast computation for fitting exponential random graph models to large social networks. Soc. Netw. Anal. Min. 5 (8), 1–19.
Holland, Paul W., Leinhardt, Samuel, 1981. An exponential family of probability distributions for directed graphs. J. Am. Stat. Assoc. 76, 33–65.
Hunter, David R., 2007. Curved exponential family models for social networks. Soc. Netw. 29, 216–230.
Hunter, David R., Handcock, Mark S., 2006. Inference in curved exponential family models for networks. J. Comput. Graph. Statistics 15 (3), 565–583.
Hunter, David R., Handcock, Mark S., Butts, Carter T., 2008. ergm: a package to fit, simulate and diagnose exponential-family models for networks. J. Stat. Softw. 24 (3), 1–29.
Krackhardt, David, 1988. Predicting with networks: nonparametric multiple regression analysis of dyadic data. Soc. Netw. 10, 359–381.
Lubbers, Miranda J., Snijders, Tom A.B., 2007. A comparison of various approaches to the exponential random graph model: a reanalysis of 102 student networks in school classes. Soc. Netw. 29, 489–507.
Lusher, Dean, Koskinen, Johan, Robins, Garry, 2013. Exponential random graph models for social networks: theory, methods, and applications. Cambridge University Press, New York, NY.
Morris, Martina, Handcock, Mark S., Hunter, David R., 2008. Specification of exponential-family random graph models: terms and computational aspects. J. Stat. Softw. 24 (4), 1–24.
Papachristos, Andrew V., Hureau, David, Braga, Anthony A., 2013. The corner and the crew: the influence of geography and social networks on gang violence. Am. Sociol. Rev. 78 (3), 417–447.
Resnick, M.D., Bearman, P.S., Blum, R.W., Bauman, K.E., Harris, K.M., Jones, J., Tabor, J., Beuhring, T., Sieving, R.E., Shew, M., Ireland, M., Bearinger, L.H., Udry, J. R., 1997. Protecting adolescents from harm: findings from the national longitudinal study on adolescent health. J. Am. Med. Assoc. 278, 823–832.
Robins, G., Pattison, P., Kalish, Y., Lusher, D., 2007a. An introduction to exponential random graph (p*) models for social networks. Soc. Netw. 29 (2), 173–191.
Robins, G., Pattison, P., Wang, P., 2009. Closure, connectivity and degrees: new specifications for exponential random graph (p*) models for directed social networks. Soc. Netw. 31 (2), 105–117.
Robins, G., Snijders, T., Wang, P., Handcock, M., Pattison, P., 2007b. Recent developments in exponential random graph (p*) models for social networks. Soc. Netw. 29 (2), 192–215.
Shalizi, Cosma R., Rinaldo, Alessandro, 2013. Consistency under sampling of exponential random graph models. Ann. Appl. Statistics 41 (2), 508535.
Snijders, T.A.B., Baerveldt, C., 2003. A multilevel network study of the effects of delinquent behavior on friendship evolution. J. Math. Sociol. 27, 123–151.
Snijders, Tom A.B., Pattison, P.E., Robins, Garry L., Handcock, Mark S., 2006. New specifications for exponential random graph models. Sociol. Methodol. 36, 99–153.
Strauss, David, Ikeda, Michael, 1990. Pseudolikelihood estimation for social networks. J. Am. Stat. Assoc. 85 (409), 204–212.
van Duijn, Marijtje A.J., Gile, Krista J., Handcock, Mark S., 2009. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Soc. Netw. 31, 52–62.
Wasserman, S., Pattison, P.E., 1996. Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p*. Psychometrika 61 (3), 401–425.