Modeling Unobserved Heterogeneity in Social Networks with the Frailty Exponential Random Graph Model

Janet M. Box-Steffensmeier[1], Dino P. Christenson[2] and Jason W. Morgan[3,4]

[1] Vernal Riffe Professor of Political Science, The Ohio State University, Columbus, OH 43210, USA.
Email: steffensmeier.2@osu.edu
[2] Associate Professor, Department of Political Science, Boston University, Boston, MA 02215, USA.
Email: dinopc@bu.edu
[3] Visiting Scholar, Department of Political Science, The Ohio State University, Columbus, OH 43210, USA.
Email: morgan.746@osu.edu
[4] Vice President, Behavioral Intelligence, Wiretap, Columbus, OH 43215, USA

Abstract

In the study of social processes, the presence of unobserved heterogeneity is a regular concern. It should be particularly worrisome for the statistical analysis of networks, given the complex dependencies that shape network formation combined with the restrictive assumptions of related models. In this paper, we demonstrate the importance of explicitly accounting for unobserved heterogeneity in exponential random graph models (ERGM) with a Monte Carlo analysis and two applications that have played an important role in the networks literature. Overall, these analyses show that failing to account for unobserved heterogeneity can have a significant impact on inferences about network formation. The proposed frailty extension to the ERGM (FERGM) generally outperforms the ERGM in these cases, and does so by relatively large margins. Moreover, our novel multilevel estimation strategy has the advantage of avoiding the problem of degeneration that plagues the standard MCMC-MLE approach.

Keywords: networks, heterogeneity, exponential random graph model

The statistical analysis of networks has become increasingly important in the social and behavioral sciences in recent years. Indeed, inferential and predictive statistical models for networks have been applied to a diverse range of social science problems, including international military alliances (Cranmer, Desmarais, and Menninga 2012), international trade networks (Ward, Ahlquist, and Rozenas 2013), segregation in adolescent friendship networks (Moody 2001), communication networks (Eveland and Hively 2009), social and spatial interaction (Lee, Liu, and Lin 2010; Gondal 2011), and interest group networks (Box-Steffensmeier and Christenson 2014, 2015), among many others. These statistical methods are exploding in use as they allow one to relax the conditional independence assumptions required by most other statistical methods. In fact, a key strength of these methods is that they allow complex dependencies—such as reciprocity, homophily, and transitivity—to be explicitly modeled. This allows old theories about social structure and behavior to be analyzed in a new light, while also opening the door for entirely new questions.

Yet, while the statistical analysis of networks continues to attract a great deal of attention from scholars and the broader public, the quantitative study of networks remains in the relatively early stages of development. Only recently have the statistical theory and computational techniques

been developed to rigorously analyze many common types of networks, such as weighted and dynamic networks. Moreover, there remains significant gaps in the statistical tools that limit their usefulness to scientists, among the most prevalent of which is the problem of unexplained heterogeneity in models of network formation. Unexplained heterogeneity is simply variation among the nodes that contributes to link formation but goes unmodeled, because it is either unobserved, unmeasured or unimagined.[1] Our objective is to improve the well-known and widely used exponential random graph model (ERGM), by explicitly modeling unexplained heterogeneity.

The ERGM has been a cornerstone in the revolution of network analysis because it provides a comprehensive and flexible method of incorporating structural network characteristics as well as node and edge traits to explain network formation. However, despite the important advantages of ERGMs over other methods of modeling network data, currently missing is a method of incorporating unobserved heterogeneity into these models. This is an important gap that limits the applicability of ERGMs in many areas of potential interest, given that we expect unobserved heterogeneity to be the rule, rather than the exception. Indeed, one of the two major assumptions of the ERGM is that the model is correctly specified, and coefficient bias or model degeneracy may result from violations of this assumption. In other words, if analysts are unable to collect all the individual-level covariates needed to explicitly model the data generating process, substantive inferences drawn from these models will be suspect. We propose to extend the ERGM to account for this problem through the introduction of a frailty term.

This paper is comprised of three parts. First, we provide a rigorous definition of the proposed frailty ERGM, or FERGM, and compare it to the standard ERGM approach. Second, we perform a series of Monte Carlos to validate the proposed model on networks of various sizes and network characteristics. The objective of the Monte Carlos is to compare the performance of the proposed model to the standard ERGM that does not accommodate unexplained heterogeneity. Beyond the evaluation of the proposed model, the Monte Carlo analysis also provides information about the sensitivity of the standard ERGM to the presence of unexplained heterogeneity. Serious considerations of this nature are absent from the literature and thus this analysis makes a major contribution to the broader literature in its own right.[2]

Third, we demonstrate the applicability and value of the project with applications to two well-known and important social networks: the Lazega (2001) law firm collaboration network and Magnolia High network (Resnick *et al.* 1997). These cases serve as critical tests of the FERGM, both because the networks have been explored in depth and because there is good reason to expect different levels of unexplained heterogeneity in the formation of these networks.

This paper also makes an important contribution to the estimation of ERGM-family models by employing a novel multilevel modeling solution. For the undirected networks, we use a multiple-membership mixed-effects model to derive parameter estimates for the network- and node-level network terms included in the model. This estimation strategy has the important advantage of avoiding the vexing problem of degeneration in estimation that plagues the standard MCMC-MLE approach in all but the simplest of networks while, as the Monte Carlo and substantive results show, not greatly affecting the inferences drawn from these models.

1 The ERGM and Frailty Extensions

The ability to include structural and nodal characteristics sets ERGMs apart from other network methods. As a consequence of this flexibility, ERGMs are an attractive modeling option for

---

1   By unobserved heterogeneity, we are specifically referring to the occasions in which some exogenous characteristic of the actors in a network affects their likelihood to form ties, but goes unmeasured or is otherwise excluded from the model. For instance, the characteristic of *charisma* may influence ties in a friendship network, but because charisma is difficult to measure, it is likely to be a source of unobserved heterogeneity.
2   Section 2 provides some results that point to this being a significant and important gap in scholars' understanding of ERGMs.

behavioral, social, natural networks and their coupling (see, e.g., Saul and Filkov 2007). In this section, we provide a short review of the standard ERGM, along with its limitations, and then present a rigorous description of our proposed extension, the frailty exponential random graph model (FERGM).

1.1 The standard ERGM

Consider a network of n actors for which some dyadic relationship has been recorded. Such actors may be individuals, groups, or nation-states, while the dyadic relationships may be friendship ties, shared group memberships, bilateral treaties, or other forms of ties between the nodes in the network. In the statistics literature, such relations are usually represented in the form of an adjacency matrix, Y. In the binary networks considered here,[3] Y is an $n \times n$ matrix with each element of the matrix, $y_{ij} \in \{0, 1\}$, indicating the existence or absence of a tie between unique actors (nodes) i and j in the network. In undirected networks, $y_{ij} = y_{ji} \ \forall i \neq j$; i.e., the adjacency matrix is symmetric. In directed networks, this symmetry does not necessarily hold. In other words, one member of a dyad may "send" a tie to the other member, but not "receive" a reciprocal tie in return.

The purpose of the ERGM is to model the probability of observing a given network conditional on a set of measures on that network. Mathematically, the ERGM is defined as

$$\Pr(Y = y \mid \mathbf{x}, \theta, \mathcal{Y}) = \frac{\exp\{\theta^T \Gamma(y, \mathbf{x})\}}{\sum_{\forall \tilde{y} \in \mathcal{Y}} \exp\{\theta^T \Gamma(\tilde{y}, \mathbf{x})\}}. \tag{1}$$

Here, $\mathcal{Y}$ is the support of y, $\Gamma(y, \mathbf{x})$ is a vector of network statistics calculated on y and on (exogenous) nodal characteristics $\mathbf{x}$, and θ is a vector of model coefficients. Γ can include any number of measures of the structural characteristics of the network, such as measures of density and transitivity. The denominator in (1), abbreviated below by κ, represents all possible networks that could be formed given the fixed set of nodes in the network.

As Cranmer and Desmarais (2011) have recently pointed out, the ERGM is something of a peculiar model when compared to models typically seen in social science. This is the case for a couple reasons. First, unlike the more familiar statistical models, ERGMs operate on a single realization of a network; i.e., they model the probability of observing the given network, y, with what is essentially a single observation. Second, complex dependencies are allowed in the model. In other words, there is no assumption of independence between observations. These traits are important to fully understand the model and they further highlight the importance of the proposed frailty extension.

This family of models has become well utilized in the social sciences, where they are more generally referred to as $p*$ models. While the modern ERGM is largely attributed to Wasserman and Pattison (1996), the foundations were laid much earlier in spatial statistics (Besag 1974), and further developed by work on the network distributions, the class of models, and estimation techniques (Feinberg and Wasserman 1981; Holland and Leinhardt 1981; Frank and Strauss 1986). Currently, likelihood-based inference is provided by Markov chain Monte Carlo Maximum Likelihood (MCMC-MLE) algorithms (Corander, Dahmström, and Dahmström 1998; Crouch, Wasserman, and Trachtenberg 1998; Besag 2000; Handcock 2000; Snijders 2002) as well as pseudolikelihood procedures (Strauss and Ikeda 1990; Wasserman and Pattison 1996), though there are important objections to the latter (Besag 2000; Snijders 2002). Important to the analysis of social processes, many extensions to the ERGM have allowed for role analysis (Salter-Townshend and Brendan Murphy 2015) and the consideration of egocentrically sampled network data (Krivitsky, Handcock, and Morris 2011). In addition, this family of models is

---

3 In other words, we do not consider weighted networks.

developing to consider longitudinal or dynamic networks in the form of temporal ERGMs (TERGMs) (Hanneke and Xing 2007; Hanneke, Fu, and Xing 2010; Desmarais and Cranmer 2010; Cranmer and Desmarais 2011) and separable temporal ERGMs (STERGMs) (Krivitsky and Handcock 2014; Christenson and Box-Steffensmeier 2016).

As an aside, it is important to recognize that the ERGM is only one of a handful of related models used to understand data with complex interdependencies. The $p*$ model has been extended to identify between- and within-group ties for groups with stochastic block models (Wang and Wong 1987). When group membership is unobserved, Nowicki and Snijders (2001) provide a model to describe latent class membership. The agent-based approach of Snijders (2001) posits conditionally independent dyads on latent class membership, which can be used to understand features of social network dynamics (see also Snijders, Van de Bunt, and Steglich 2010). Relatedly, latent space models, which generally reduce higher-order network dependencies down to lower dimensional spaces, have grown out of work in spatial statistics (Hoff, Raftery, and Handcock 2002; Hoff and Ward 2004; Franzese and Hays 2006), and have paid particular attention to exploring network dynamics. For example, Ward and Hoff (2007) model ties at each time period as a function of latent covariates (see also Ward, Ahlquist, and Rozenas 2013; Durante and Dunson 2014). More recently, Hoff *et al.* (2015) provides a method that combines the dependence representation of the agent-based approach with the node heterogeneity of latent space approaches (see also Minhas, Hoff, and Ward 2016).

Our objective in this paper is to improve on the well-used, powerful and flexible ERGM. It is not to argue that the ERGM should be used to model all networks. One or more of the approaches above may be more appropriate for a particular research question or data set.[4] As stated by Cranmer and Desmarais (2011, 68), the best approach for modeling complex interdependencies will be "based on substance and the quest for a universally best method is quixotic." Instead, we offer here an important amendment to the ERGM, a model that is already extremely popular across disciplines and is likely to continue to be so. In providing a frailty extension to the ERGM we make it even more widely applicable, since we suspect that there are many contexts where complex interdependencies come with unobserved heterogeneity. Moreover, our model has the positive side effect of overcoming estimation problems in the ERGM that lead to degeneracy.

1.2   The frailty ERGM

An important assumption for identifying Equation (1) is that $\Gamma(\mathbf{y}, \mathbf{x})$ is a vector of sufficient statistics for the network; i.e., the model includes all network- *and* nodal-level covariates necessary to explain the network. This is a strong assumption. As is the case with standard regression models, excluding covariates that are correlated with both the dependent and independent variables is a threat to inference. In a network model, excluding relevant covariates is likely to be even more damaging to inference given the inherent correlations between network statistics. For instance, in a friendship network, we may believe that the formation of friendships (ties) between individuals is driven by many factors that are measurable; e.g., social class, race, age, etc. However, we may suspect that there are other, intangible factors specific to each individual that are difficult if not impossible to measure, such as "friendliness" or "charisma," that are also related to network structure (people that are friendlier are likely to have more friends, increasing the centrality of friendly individuals above what we would expect given their other, known attributes). In other words, the observed and measured characteristics are not sufficient for explaining the network we observe. Further, because these unobserved characteristics may be correlated with both the outcome (network structure) and the other explanatory variables, there is the potential for mistaken inferences when such heterogeneity is not accounted for.

---

4   For a summary of the limitations of ERGMs, see Cranmer and Desmarais (2011).

The approach we propose to accommodate such unobserved heterogeneity is directly analogous to the use of frailty terms in other applied methods, such as in event history models. In that context, frailty terms—modeled as individual-level random effects—are added to the model to account for unobserved heterogeneity in the units, where it is meant to capture variation in individuals' susceptibility to disease or death (Vaupel, Manton, and Stallard 1979). In our case, we introduce individual (or group) random effects into Equation (1) to model unobserved heterogeneity in the propensity for individuals in the network to form ties. To be more concrete, suppose each individual has some level of *sociality*—i.e., friendliness or charisma in the friendships network example above—which we will call s_i, then Equation (1) could be augmented as such:

$$\Pr(Y = y \mid \cdot) = \frac{1}{\kappa} \exp \left\{ \theta^T \Gamma(\mathbf{y}, \mathbf{x}) + \sum s_i \right\}. \tag{2}$$

Here $\sum s_i$ is the sum of all of the individual frailty components.[5] Recall that the ERGM estimates a single probability for the graph as a whole. Thus, the frailties are added as a set to the numerator and denominator (i.e., κ). For identification, they are allowed to vary according to a distribution centered at 0. For the purposes of this paper, we use a normal distribution with mean 0 and a variance, σ_s^2, to be estimated from the data; i.e., $s_i \sim N(0, \sigma_s^2)$. To model a greater degree of variability, distributions with more mass in the tails could certainly be used. Notice that, because $E[s_i] = 0$, modeling the frailty term as being centered at zero has a key advantage of leaving the probability of the graph fixed in expectation, though the variance increases.

Our proposed extension is directly applicable to directed networks, though in this case we would include *sender* and *receiver* effects—one each for each individual (or group)—in the network. In this case, the objective is to model heterogeneity in the probability that actors in the network send ties (the sender effect) and the probability they accept ties (the receiver effect). We model these effects as having a joint Gaussian distribution centered at zero,

$$\begin{pmatrix} s_i \\ r_i \end{pmatrix} \sim N \left[\mathbf{0}, \begin{pmatrix} \sigma_s^2 & \sigma_{sr} \\ \sigma_{sr} & \sigma_r^2 \end{pmatrix} \right] \tag{3}$$

where σ_s^2 and σ_r^2 are the variances of the sender and receiver effects, respectively, and σ_{sr} is their covariance, each to be estimated from the data. In terms of the ERGM, including these sender and receiver effects would result in the following model to be estimated:

$$\Pr(Y = y \mid \cdot) = \frac{1}{\kappa} \exp \left\{ \theta^T \Gamma(\mathbf{y}, \mathbf{x}) + \sum s_i + \sum r_i \right\}. \tag{4}$$

As in Equation (2) and the sociality effects, the sender and receiver effects are included as additive terms in the numerator and denominator. For the remainder of this paper, we focus on the undirected frailty model of Equation (2), though it should be clear that the directed case is a relatively straightforward extension.

1.3 Estimation

Estimation of the proposed FERGM was performed using a novel application of a (logistic) multiple-membership mixed-effects model, which is a model developed by education researchers for the purpose of modeling student performance in cases where students may have moved between classrooms (or other groups) during a period being studied (Goldstein 2003, ch. 13). In that context, researchers are often interested in the performance of students on standardized tests

---

5 This specification is analogous to the way in which Krivitsky *et al.* (2009) introduce random effects terms to latent space network models.

and wish to account for the hierarchical structure of the data generating process, where students are nested in classrooms, which are themselves nested in schools, etc. If students never move classrooms (schools), then a standard hierarchical model is easily applied. However, the situation is complicated when students are assigned to multiple classrooms. In that case, researchers need to adjust for the fact that students can spend a portion of their time in multiple classrooms. Multiple-membership models are one tool that has been developed for just such a situation.

In the case of the proposed network model, instead of modeling the performance of students, we model the occurrence of ties between two nodes. In our application of the model, each (potential) tie, y_{ij}, is estimated as being associated with two groups—one being the group associated with node i and the other with node j. More concretely, in the undirected case each tie in this model is estimated as follows

$$\Pr(y_{ij} = 1) = \text{logit}^{-1} \left\{ \theta^T \Delta\Gamma(\mathbf{y}, \mathbf{x}) + s_i + s_j \right\}. \tag{5}$$

Here, s_i and s_j are the random effects associated with nodes i and j, $\Delta\Gamma(\mathbf{y}, \mathbf{x})$ is a vector of change statistics (those corresponding with the sufficient statistics of Equation (1)) for when the tie between i and j is toggled, while θ is the coefficient on these change statistics. In other words, if you were to exclude the random effects, you would recover the maximum pseudolikelihood model for the standard ERGM.

Estimating the FERGM in this way has both advantages and disadvantages. On the one hand, the FERGM can be estimated with many standard mixed-effects modeling packages (we will use R and Stan in the Monte Carlo and examples below), estimated much more quickly for larger networks, and is less susceptible to issues of degeneracy when compared to the MCMC-MLE approach. On the other hand, estimating the model in a way that extends the maximum pseudolikelihood approach potentially opens up issues of efficiency and excessively optimistic inferences, which are a weakness of that approach. However, our Monte Carlos as well as the below applications suggest that the robustness of our estimation strategy to issues of degeneracy should outweigh those concerns for most analysts.

2 Monte Carlo Analysis

In this section, we present the results of an extensive Monte Carlo analysis. The objective of this analysis is to compare, under controlled conditions, the performance of the proposed FERGM to the standard ERGM when unobserved heterogeneity is present in the network formation process. To foreshadow our results, we find that the performance of the FERGM, as measured by the root-mean-squared error of the errors, significantly outperforms the standard ERGM in a number of contexts.

2.1 Network simulation

The first step of the Monte Carlo analysis was to simulate a set of networks of various sizes and specifications.[6] Undirected networks of $n \in \{25, 50, 100\}$ nodes were specified. To approximate the type of networks that may be encountered in applied research, while still keeping the data generating process relatively simple, the model used to generate these networks included: the overall network density, measured by the number of edges (ties) in the network; a transitivity effect, measured by the geometrically weighted edgewise shared partner distribution (gwesp; Snijders *et al.* 2006);[7] and two node-level homophily effects. For the first nodal covariate, group$_i$, each node was randomly assigned with equal probability to one of two groups. For the second

---

6 Replication materials for the Monte Carlo analysis and the applications in Section 3 are available on the Harvard Dataverse (Box-Steffensmeier, Christenson, and Morgan 2017).

7 Our simulation model, as well as our applications below, contain a third-order effect term in the form of GWESP. We assume here that the random nodal effects are not correlated with GWESP.

Table 1. Summary of network densities for the networks generated in the Monte Carlo study.

	25 nodes			50 nodes			100 nodes		
γ	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
0.00	0.017	0.081	0.237	0.214	0.258	0.299	0.225	0.239	0.256
-0.25	0.003	0.042	0.100	0.064	0.158	0.228	0.189	0.205	0.222
-0.50	0.003	0.030	0.077	0.020	0.049	0.113	0.129	0.160	0.188
-0.75	0.003	0.025	0.073	0.011	0.031	0.061	0.065	0.109	0.144
-1.00	0.003	0.020	0.057	0.008	0.023	0.050	0.025	0.068	0.115

Note: 1008 undirected, binary networks were drawn for each specification. The values for γ refer to the coefficient on the individual-level unobserved heterogeneity term, which was drawn from a Gaussian distribution with $\mu = 0$ and $\sigma = 1$. Full model specifications are found in Equation (6).

nodal covariate, a continuous value for each node was drawn from a standard normal distribution, $\sigma_i \sim N(0, 1)$.

In notation, the data generating process used for the simulated networks was as follows:[8]

$$\Pr(Y = y \mid \cdot) = \frac{1}{\kappa} \exp \left\{ -3.25 \times \text{edges}_y + 0.75 \times \text{gwesp}_y \right.$$

$$\left. + 0.25 \times \sum_{i,j \in y} \mathbf{1}(\text{group}_i = \text{group}_j) + \gamma \times \sum_{i,j \in y} |\sigma_i - \sigma_j| \right\}. \tag{6}$$

Here $\mathbf{1}(\cdot)$ is the indicator function and is equal to 1 if i and j belong to the same group and 0 otherwise. In other words, this defines a dummy variable that indicates shared group membership and will have the effect of increasing the likelihood of a tie between nodes in the same group.

As indicated in (6), the vector of coefficients was set to $\theta = (-3.25, 0.75, 0.25, \gamma)$. The coefficient on the continuous node-level homophily effect, γ, varied in increments of 0.25 from 0.0 to -1.0; i.e., $\gamma \in \{0, -0.25, -0.50, -0.75, -1.0\}$. By varying the coefficient on the unobserved heterogeneity in this way, we are able to test the performance of the ERGM and FERGM on networks for which the importance of the unobserved heterogeneity in the data generating process varies.[9]

Overall, the above model produced 15 different data generating processes (three network sizes, five levels of heterogeneity). For each of these, random networks were generated following a three-step procedure. First, an undirected basis network was generated with a default density of 0.15.[10] Second, nodal covariates were assigned as specified above in Equation (6). Third, using the modified basis network, 1008 networks were simulated for the given network coefficient specification.[11] A summary of model specifications and average network densities is presented in Table 1 and six selected networks are show in Figure 1. As reported, network densities remain relatively stable—they do not display any tendency toward 0 or 1—thus we were confident the chosen specification does not produce degenerate networks across the range of network sizes used.[12]

---

8 The data generating process was specified to resemble the Lazega (2001) law firm network analyzed by van Duijn, Gile, and Handcock (2009) in their landmark network simulation studies.

9 The decay strength for the geometrically weighted shared partner measure was fixed at 0.75, which was the value used in van Duijn, Gile, and Handcock (2009). For a description of this measure, see Hunter and Handcock (2006), Hunter (2007).

10 The default density was chosen to limit the chance of creating degenerate networks. Overall, only five degenerate networks were generated. All simulations were performed in R (version 3.0.3) using the network (version 1.9.0) and ergm (version 3.1.2) packages.

11 Simulations were done with a burn-in of 50,000, saving every 100 networks.

12 As expected, the more important the unobserved heterogeneity was to the network generating process, the lower the average network density.
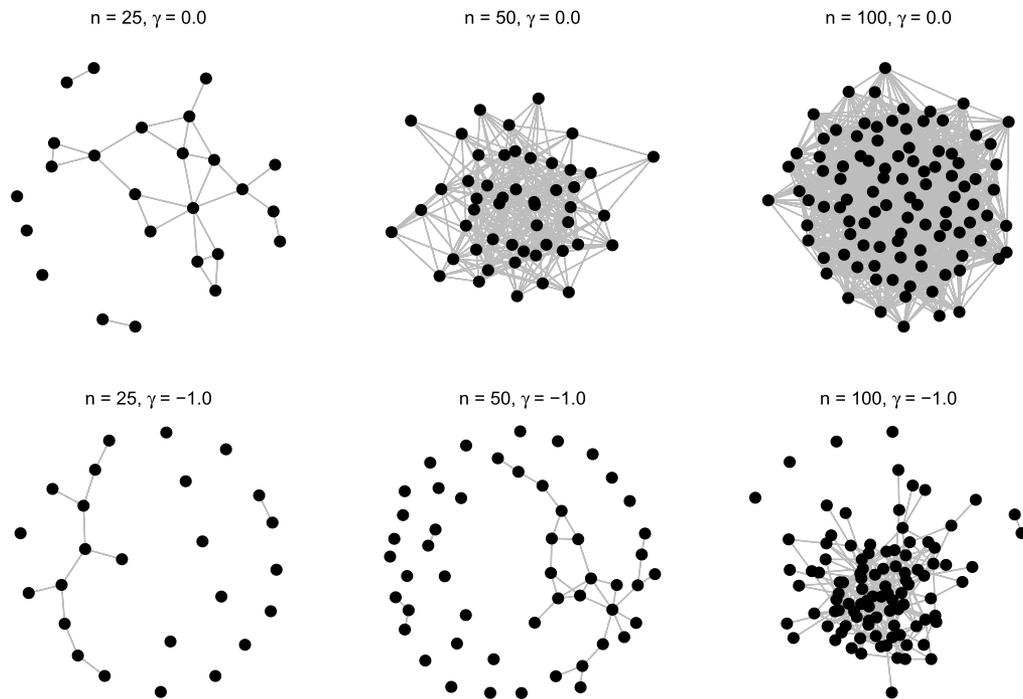
Figure 1. Example networks.

2.2 Model estimation and evaluation

The second task in our preliminary Monte Carlo assessment of the proposed FERGM is to estimate the models using the networks simulated as described in the previous section. Since we are interested in the comparative performance of the ERGM and FERGM in the presence of unobserved heterogeneity, we purposefully excluded from the estimation of the models the node-level continuous homophily term included in the network simulations; i.e., the last term involving γ in Equation (6).

The standard ERGM was estimated through the state-of-the-art Markov Chain Monte Carlo MLE (MCMC-MLE) methodology. In the MCMC-MLE approach, the denominator of Equation (1) is approximated with a large sample from the universe of possible networks given the previous iteration's estimates of θ. After many iterations, estimates of θ are guaranteed to converge. However, while the MCMC-MLE methodology has better theoretical properties than the alternative maximum pseudolikelihood (MPLE) approach, (see, e.g., Robins *et al.* 2007), it can suffer from issues of degeneracy (Snijders 2002; Snijders *et al.* 2006).[13] For this reason, we took a multistep approach to avoid issues of degeneracy and to ensure convergence. In the first step, initial values for θ were generated by taking (up to) four iterations of the algorithm.[14] In the second step, the algorithm was restarted using the estimates from the first step as the initial values. In this step, the algorithm was run for up to 20 iterations, or until convergence (as determined by a lack of a change in the log-likelihood). For situations where the model failed to converge during the second stage, another 10 more attempts were made (with 20 iterations of the algorithm included during each attempt), with the initial values set to those from the previous attempt. If, after all of this, the model still failed to converge, the network was discarded for nonconvergence.[15]

---

13 The most common form of degeneracy is when the sampling procedure puts all the mass on networks with density of zero or one.

14 The algorithm terminates in the event that the model converged in fewer than four iterations.

15 Overall, this occurred for 528 networks, or about 3.5% of the simulated networks.
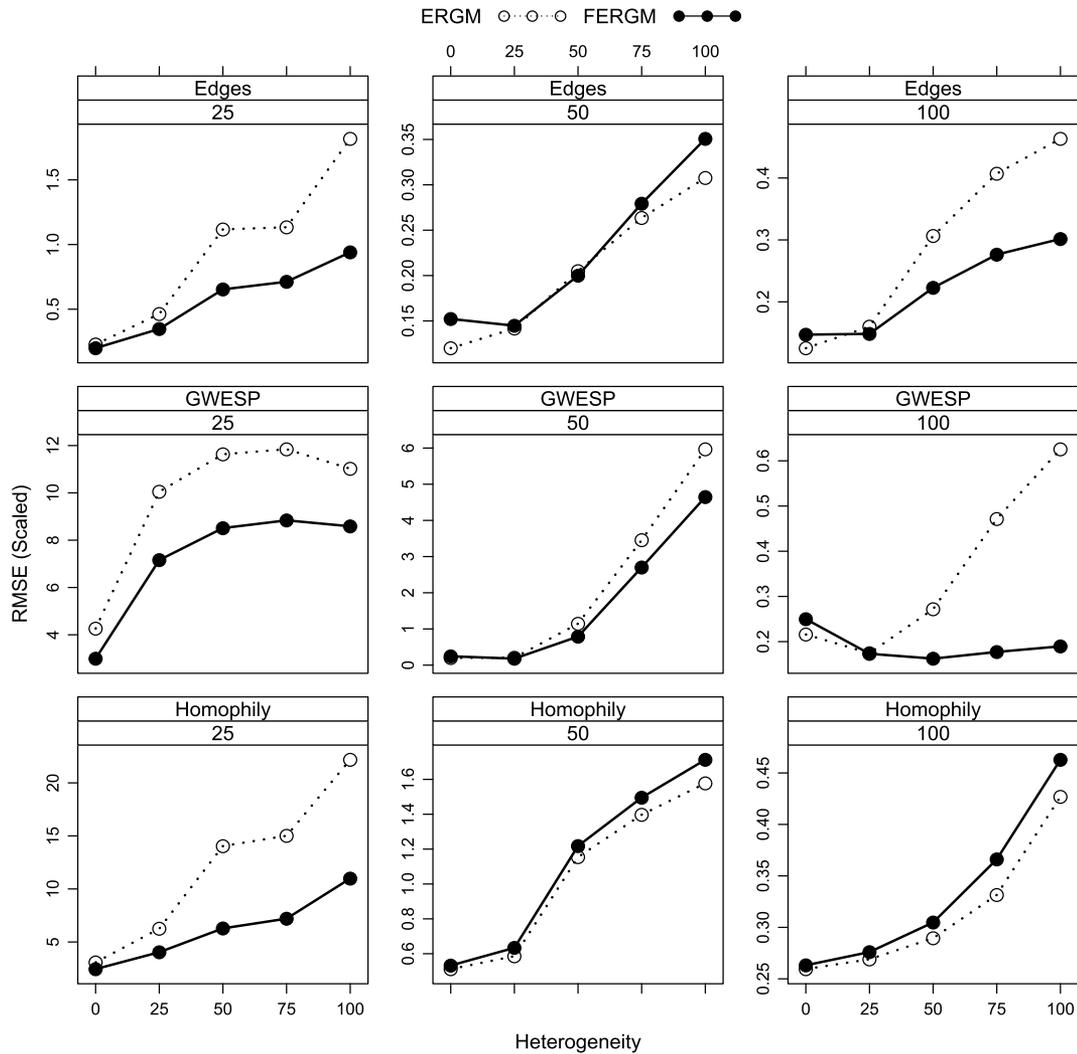
Figure 2. Monte Carlo results.

Estimation of the FERGM was more straightforward. The FERGM was estimated as a multiple-membership mixed-effects model, as described in Section 1.3, in Stan (Stan Development Team 2015). As the Monte Carlos focus on undirected networks, each node's random component was modeled as drawn from a single normal distribution, centered at zero with variance σ_w, with the assumption that σ_w was drawn from a Cauchy distribution centered at zero with a dispersion of 2. The parameters for the endogenous structural parameter and the exogenous covariates were given diffuse normal priors centered at zero with a standard deviation of 10.[16]

2.3 Results

Overall results for the Monte Carlo study are presented in Figure 2. Each of the panels shows the (scaled) root-mean-squared error for coefficient estimates[17] on the vertical axis and the level of the unobserved heterogeneity on the horizontal axis. The first column of panels shows networks with 25 nodes, the center column 50-node networks, and the right column 100-node networks.

---

16 The model was run for 1250 iterations for each network, saving the last 250. This ensured that the Gelman–Rubin (Gelman and Rubin 1992) statistic for convergence, $\hat{R}$, was below 1.05 for all parameters and that the effective number of samples from the posterior was sufficient.

17 To make them more comparable, RMSE was scaled by the magnitude of the true coefficient values.

Each rows reports the results for a different covariate. Results for the ERGM are reported as dotted lines and open circles, while the FERGM results are shown as solid lines and closed circles.

The general takeaway from these results is that, on average, the FERGM performs better than the standard ERGM in the presence of unobserved heterogeneity. In six of nine panels, the FERGM reports lower RMSE than the standard ERGM at any level of unobserved heterogeneity. In the other three cases, the differences in performance are slight. It is important to note that the scales of the y-axes in these three are quite small. Furthermore, two of the cases in which the ERGM seems to outperform the FERGM are for the coefficient on the exogenous homophily covariate for the 50- and 100-node networks (bottom row, right two columns). Since the unobserved heterogeneity was, by construction, not correlated with this covariate, it should not be surprising that the ERGM performed well.

On the other hand, when the ERGM is wrong, it is often significantly so. For instance, take the GWESP structural term for the 100-node networks. Scholars are often interested in this term, as such a parameterization of transitivity tends to explain much about how a network is structured (see, e.g., Goodreau, Kitts, and Morris 2009). In this case, the presence of unobserved heterogeneity leads to RMSEs that are more than 3 times the size of those produced by the FERGM. That is, when the FERGM does better, it generally does so by a huge margin relative to the scale of the coefficient. Overall then, the results are strongly in favor of the FERGM. The ERGM and FERGM perform similarly in only a few cases: for the mid- and large-sized networks the ERGM performs relatively well on the exogenous covariate; ERGM also does well on the endogenous terms for the mid-sized network. However, it fails otherwise, and when it fails, it does so by large margins, which makes inferences based on the ERGM more problematic.

Finally, it should be noted that these results should be considered as a best-case scenario for the standard ERGM. A great deal of effort went into estimating each of these networks and assuring that the model had converged (see Section 2.2). When there was some question as to the convergence status, the network was dropped. This means the Monte Carlo study was designed conservatively, in favor of the standard ERGM. Because convergence in the ERGM may be hampered by heterogeneity, evidence of some of the worst performing ERGMs are not included in the results. Even with these precautions in place, these results suggest the standard ERGM is very fragile in the face of unobserved heterogeneity and that our proposed FERGM can play an important role when such heterogeneity exists.[18]

3 Applications

While the previous section demonstrated the potential importance of directly modeling unobserved heterogeneity in nodal covariates in simulated network data, there remains a question as to how the model we propose will function in practical, real-world problems. For this reason, in this section we turn to two networks that have played an important role in the networks literature: the Lazega (2001) law firm collaboration network and the Magnolia High network (Resnick *et al.* 1997).

---

18 As a supplement to the extensive Monte Carlo analysis we present here, we also performed a more modest analysis to capture another common case. In that analysis, three sets of fifty 50-node networks were generated where 3 nodes (6% of the network) had a greater likelihood of forming ties (higher sociality). This was done by using the `sociality` term from the `ergm` package. The three sets of networks were simulated such that the three "social" nodes had sociality levels of 0.50, 1.00, and 2.00, respectively. Otherwise, the data generating process was the same as in the rest of the paper. We then estimated the standard ERGM and the proposed FERGM using the same routines as described above. The results of this analysis were inline with what we present here: both the ERGM and FERGM performed similarly when the unaccounted for heterogeneity was low, but in networks with higher heterogeneity, ERGM's performance suffered. In fact, in networks with three nodes having sociality of 2.00, only 27 of the 50 ERGMs converged using the routine we described above. And in cases where the standard ERGM did converge, estimated RMSEs were noticeably higher for the ERGM than for FERGM. Full results of this analysis is available from the authors.
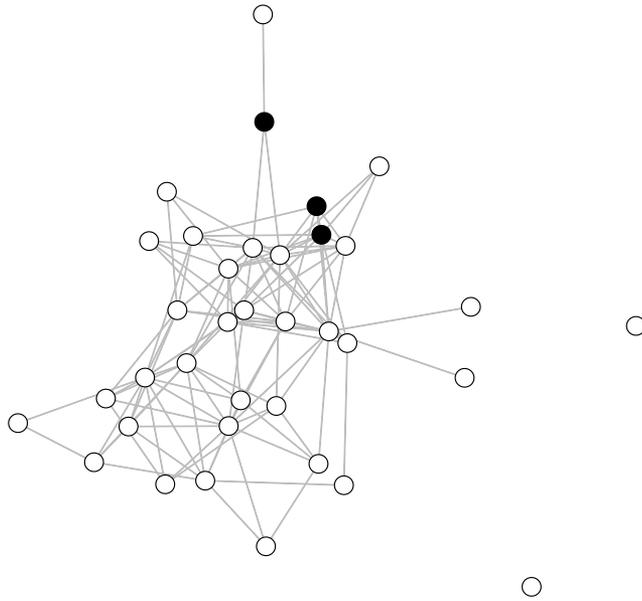
Figure 3. Lazega (2001) Law Firm Collaboration Network. This is an undirected network recording collaborations between the 36 attorneys. The three female attorneys are highlighted in black.

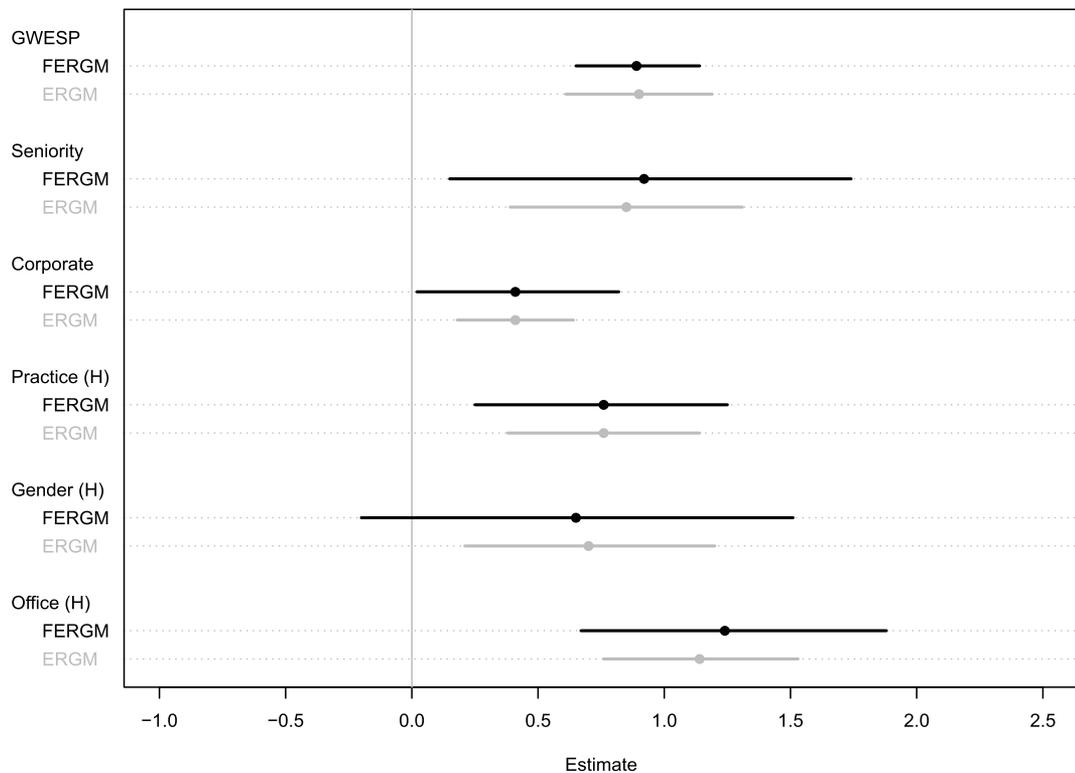## 3.1 Lazega law firm collaboration network

The first example we turn to is the Lazega (2001) law firm collaboration network. This undirected network consists of 36 attorneys (nodes) in three New England offices: 22 attorneys were located in Boston, 13 in Hartford, and 1 in Providence. There were three female attorneys in the data set, all of which were located in the Boston office. In addition to each attorney's office and gender, the network includes data on attorneys' seniority, and whether they were in litigation or corporate practice. Overall, there are 115 ties (density of 18.3%). The network is shown in Figure 3, with the female attorneys in red.

The Lazega network has been used in a number of studies investigating the properties of ERGMs. In addition to acting as a template for the simulated networks explored in the Monte Carlo analysis, van Duijn, Gile, and Handcock (2009) used the network in an important study comparing the performance of the MPLE and MCMC-MLE approaches to estimating ERGMs, while Hunter and Handcock (2006) used the network as an example in their development of curved exponential family models.[19] In other words, the Lazega network is considered to be well understood and it quite amenable to analysis with ERGMs. This makes it an excellent example to gauge the performance of the proposed FERGM.

For this analysis, we have adopted an identical model specification to the one used in van Duijn, Gile, and Handcock (2009). The ERGM and FERGM included an edge term, homophily terms for practice, gender, and office; nodal covariates for seniority and practice; and, to account for the network structure, the model included a geometrically weighted edgewise shared partner term (with decay equal to 0.7781, as in van Duijn, Gile, and Handcock (2009)). The ERGM was estimated with MCMC-MLE, as is current standard practice in the network sciences, while the FERGM was estimated as a Bayesian multiple-membership model, as in the Monte Carlos in the previous section.[20]

---

19 Also see Lazega and Pattison (1999) for another investigation of this network in the literature.

20 The ERGM was estimated with an MCMC burn-in of 250,000 with every 500 samples retained for a total sample size of 40,000. For the FERGM, which was estimated in Stan, four chains of 100 warm up iterations and 500 posterior samples was taken. For this small network, this was enough to get the Gelman–Rubin diagnostic (Gelman and Rubin 1992) of convergence $\hat{R}$ to be at or below 1.01 for all estimated coefficients.

Figure 4. ERGM and FERGM results for the Lazega law firm collaboration network.

The coefficients estimates along with 95% confidence (credible) intervals for these estimated models are reported in Figure 4.[21] Estimated from the standard ERGM are reported in black, while the FERGM results are in red. As expected, the coefficient estimates for the ERGM agree with those published by van Duijn, Gile, and Handcock (2009), while each of these coefficients is also statistically significant at the 95% confidence level (as shown by the confidence intervals excluding zero). The posterior means for the FERGM also largely agree with the standard ERGM results.

The difference between these models emerges when we turn to comparing the confidence intervals of the ERGM with the credible intervals of the FERGM. On average, the FERGM credible intervals are wider than the ERGM confidence intervals. In most cases this does not make a large difference to the inferences we can draw from the model; however, in one important case it does. Notice that the credible interval for the gender homophily term increases significantly in the FERGM when compared to the standard ERGM. In the case of the FERGM, the credible interval includes zero, suggesting that the effect of gender homophily is not statistically different from zero.

Looking more carefully at the network we see that the results for the FERGM are more reasonable than those of the standard ERGM. Recall that there are only three (of 36) women in the network. Of the potential 630 ties in the network, only three of them represent potential gender homophily effects between women and only one of these is a tie. In other words, there is very little data from which we can make the claim that gender homophily is important to the formation of this network; i.e., the results from the standard ERGM seem unreasonably certain about this effect. On the other hand, the proposed FERGM, by incorporating the potential for other unobserved nodal factors, is more conservative.

As an additional check on the relative performance of the two models, we also performed a straightforward predictive analysis through simulation. For the estimated ERGM, we simulated

---

21 The coefficient estimates for the edges term was omitted as it is not typically of particular interest to analysts.
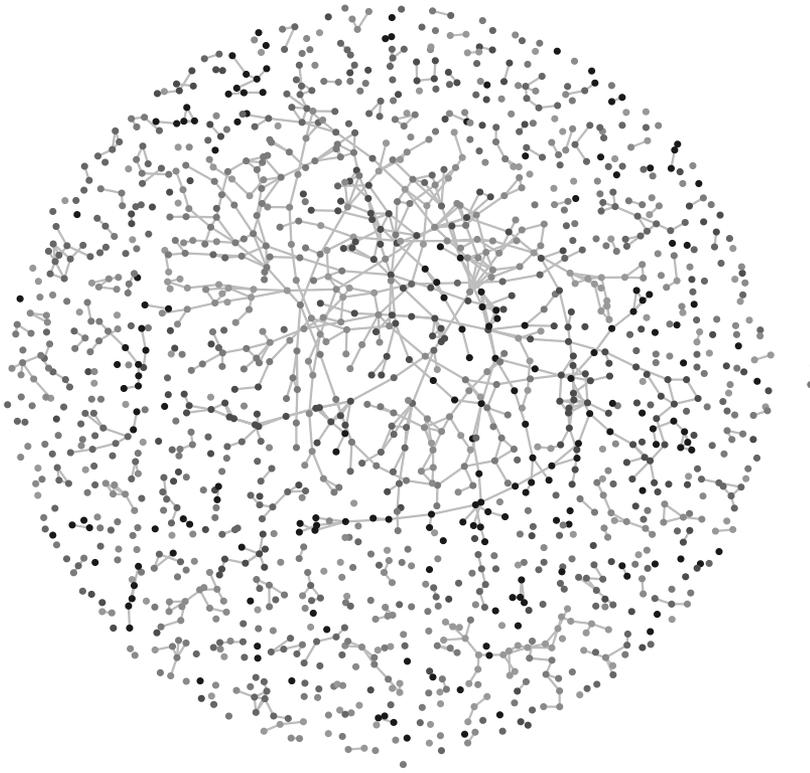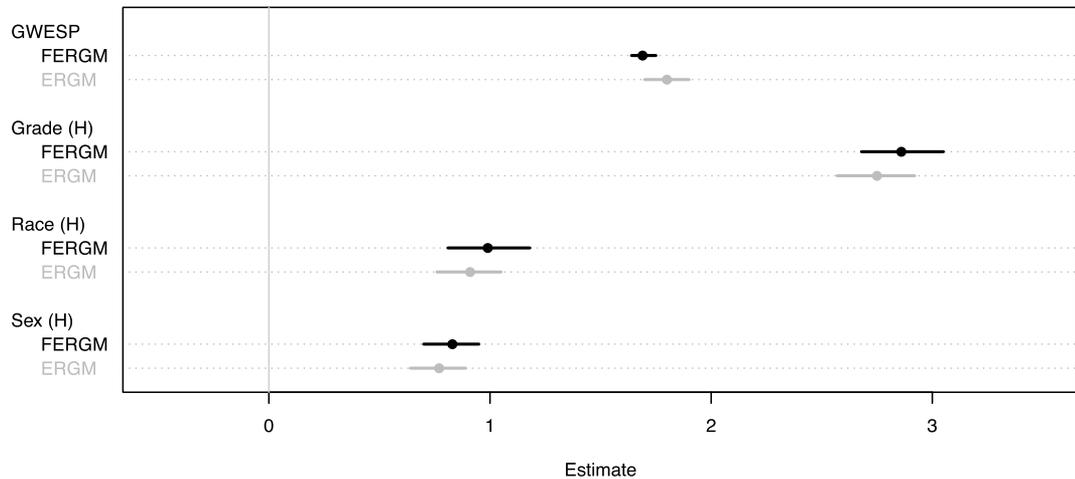
Figure 5. Magnolia high friendship network.

500 networks, extracted the adjacency matrix, and then calculated the percentage of ties that were correctly predicted in each of these. For the FERGM, we calculated the predicted probability of a tie for each of the 500 posterior draws of the estimated parameters, used these predicted probabilities to take a single draw from a Bernoulli distribution for each dyad in the network, then calculated the percent correctly predicted. Calculating the mean percentage correctly predicted for each of these sets of simulations, we found that, on average, the FERGM outperformed the standard ERGM by approximately 5% (79.7% for the FERGM versus 74.6% for the ERGM).

3.2 Magnolia high friendship network

The second network we investigate is the Magnolia High friendship network, which was first reported in Resnick *et al.* (1997). This network is generated from a model of adolescent friendship networks in a large high school in the Southern U.S. Because it is constructed from a well-fitting model of the observed network, the data generating process is known. Thus, the network allows us to explore the behavior of the FERGM in another large, real-world example, but one in which we should expect little to no unobserved heterogeneity. This example, then, gives us an idea of the effect of using the FERGM in a case where a standard ERGM would be sufficient. Should the FERGM differ significantly from the ERGM then analysts should be cautious in applying the FERGM. Alternatively, should the FERGM do no harm, then analysts risk little in applying the FERGM whenever there is reason to suspect unobserved heterogeneity.

Overall, the Magnolia High network has 1461 students across five grades (7th to the 12th). The network includes 974 undirected ties for a density of 0.09%. The network includes data on each node's grade, race, and sex. The full network is presented in Figure 5, with nodes colored by their grade.

For estimating the ERGM and FERGM for the Magnolia High network, we have adopted the same model specified to generate the model. This included a term for the number of edges,

Figure 6. ERGM and FERGM results for the magnolia high friendship network.

a geometrically weighted edgewise shared partner term (decay fixed at 0.25), and homophily terms on grade, race, and sex.[22]

The results for the ERGM and FERGM estimates are shown in Figure 6. As before, point estimates and confidence intervals for the ERGM are shown in black, while the FERGM posterior means and credible intervals are shown in red. Unlike in the case of the Lazega network, there is not much of interest to report; i.e., the ERGM and FERGM results are in agreement on all of the covariates. In other words, in this large network with a known data generating process and little suspicion of unobserved heterogeneity, the FERGM does no harm.[23]

4 Discussion

Network data, virtually regardless of the application area, are likely to exhibit unobserved heterogeneity. It is, at the very least, difficult to rule out *a priori*. Our simulations show that given heterogeneity, ERGMs often fall short. Under these circumstances, a modeling strategy that is robust to heterogeneity is desirable. Here the FERGM looks promising. FERGMs directly estimate the effects of heterogeneity, which make for more reliable estimates of network structure. Moreover, when the FERGM outperforms the ERGM it generally does so by large margins.

In addition, our novel estimation approach, using a particular form of mixed-effects models, opens up numerous avenues for future research. In simulated and substantive networks, we have shown that using a mixed-effects model to estimate the FERGM leads to similar inferences, even when we expect little to no unobserved heterogeneity. Our model has the important advantage of avoiding the issues of degeneracy that present serious obstacles to the estimation of ERGMs via MCMC-MLE in all but the simplest networks. Further exploration of the application of well-known mixed-effects modeling techniques to exponential family network models could lead to advances in modeling more complex network generating processes without making scholars dependent on fragile estimation methods.

We suspect that advances in the development and application of network modeling, such as this, will be useful throughout the social sciences and a host of other areas that are interested in

---

22 The ERGM was estimated with an MCMC burn-in of 65,536 with every 4,096 samples retained for a total sample size of 32,768. For the FERGM estimated in Stan, four chains of 400 warm up iterations and 125 samples were taken. For this small network, this was enough to get the Gelman–Rubin diagnostic (Gelman and Rubin 1992) of convergence $\hat{R}$ to be at or below 1.05 for all estimated coefficients.

23 This should not be taken to mean that FERGM could *never* cause a problem. There may be some case (that we are unaware of) where the data generating process behind the unobserved heterogeneity is not accommodated by our rather straightforward implementation. In such a case the estimates of endogenous and exogenous parameters could be adversely affected.

modeling complex interdependencies. In addition, the methodology may help practitioners in the public policy and national security arenas who rely on the evaluation of coalition strategies and relationships more accurately assess the role of policies and conditions. Further, at least one of the applications appears to advance its own substantive literature, reducing the value of gender in the generation of ties between lawyers.

There are, however, important caveats to this work. The simulations and substantive applications do not comprise an exhaustive set of networks. Thus, while we believe that this paper provides an invaluable extension to modeling network complexity where there is likely to be unobserved heterogeneity, further work is necessary to fully understand the properties of the FERGM and the tradeoffs with the ERGM. This is especially important given the myriad of ERGM terms and several ways in which unobserved heterogeneity may exist in a network. Moreover, the frailty extension is not a substitute for appropriately modeling link formation with the observed data in the context of the ERGM. First and foremost, scholars need to embark on the nontrivial task of specifying their models with the appropriate endogenous and exogenous terms. Provided they have met with success there, employing the FERGM when there is believed to be remaining unobserved heterogeneity would be a logical next step.

References

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* 26(2):192–236.

Besag, J. 2000. Markov chain Monte Carlo for statistical inference. Technical report, University of Washington, Center for Statistics and the Social Sciences.

Box-Steffensmeier, J. M., and D. P. Christenson. 2014. The evolution and formation of amicus curiae networks. *Social Networks* 36:82–96.

Box-Steffensmeier, J. M., and D. P. Christenson. 2015. Comparing membership interest group networks across space and time, size, issue and industry. *Network Science* 3(1):78–97.

Box-Steffensmeier, J. M., D. P. Christenson, and J. W. Morgan. 2017. Replication data for: Modeling unobserved heterogeneity in social networks with the frailty exponential random graph model. Harvard Dataverse, doi:10.7910/DVN/K3D1M2.

Christenson, D. P., and J. M. Box-Steffensmeier. 2016. Why amicus curiae cosigners come and go: A dynamic model of interest group networks. In *Complex networks & their applications V: Proceedings of the 5th international workshop on complex networks and their applications*, ed. H. Cherifi, S. Gaito, W. Quattrociocchi, and A. Sala. Studies in Computational Intelligence. New York: Springer International Publishing, pp. 349–360.

Corander, J., K. Dahmström, and P. Dahmström. 1998. *Maximum likelihood estimation for Markov graphs*. Stockholm: University of Stockholm, Department of Statistics.

Cranmer, S. J., and B. A. Desmarais. 2011. Inferential network analysis with exponential random graph models. *Political Analysis* 19:66–86.

Cranmer, S. J., B. A. Desmarais, and E. J. Menninga. 2012. Complex dependencies in the alliance network. *Conflict Management and Peace Science* 29(3):279–313.

Crouch, B., S. Wasserman, and F. Trachtenberg. 1998. Markov chain Monte Carlo maximum likelihood estimation for p^* social network models. Paper Presented at the XVIII International Sunbelt Social Network Conference, Sitges, Spain.

Desmarais, B. A., and S. J. Cranmer. 2010. Analyzing longitudinal networks: The temporal exponential random graph model. Presented at the Annual Meeting of the Midwest Political Science Association.

Durante, D., and D. B. Dunson. 2014. Nonparametric bayes dynamic modelling of relational data. *Biometrika* 101(4):883–898.

Eveland, W. P., and M. H. Hively. 2009. Political discussion frequency, network size, and "heterogeneity" of discussion as predictors of political knowledge and participation. *Journal of Communication* 59(2):205–224.

Feinberg, S. E., and S. Wasserman. 1981. An exponential family of probability distributions for directed graphs: Comment. *Journal of the American Statistical Association* 76(373):54–57.

Frank, O., and D. Strauss. 1986. Markov graphs. *Journal of the American Statistical Association* 81(395):832–842.

Franzese, R., and J. Hays. 2006. Spatiotemporal models for political-science panel and time-series-cross-section data. Presented at the 23rd Meeting of the Society for Political Methodology, Davis, California.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4):457–472.

Goldstein, H. 2003. *Multilevel statistics models*. Oxford: Oxford University Press.

Gondal, N. 2011. The local and global structure of knowledge production in an emergent research field: An exponential random graph analysis. *Social Networks* 33(1):20–30.

Goodreau, S. M., J. A. Kitts, and M. Morris. 2009. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* 46(1):103–125.

Handcock, M. 2000. Progress in statistical modeling of drug user and sexual networks. Technical report, University of Washington, Center for Statistics and the Social Sciences.

Hanneke, S., W. Fu, and E. P. Xing. 2010. Discrete temporal models of social networks. *Electronic Journal of Statistics* 4:585–605.

Hanneke, S., and E. P. Xing. 2007. Discrete temporal models of social networks. In *Statistical network analysis: models, issues, and new directions: ICML 2006 workshop on statistical network analysis, Pittsburgh, PA, USA, June 29, 2006, Revised Selected Papers*, ed. E. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. X. Zheng. Berlin: Springer, pp. 115–125.

Hoff, P. D. 2015. Multilinear tensor regression for longitudinal relational data. *The Annals of Applied Statistics* 9(3):1169–1193.

Hoff, P. D., A. E. Raftery, and M. S. Handcock. 2002. Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97(460):1090–1098.

Hoff, P. D., and M. D. Ward. 2004. Modeling dependencies in international relations networks. *Political Analysis* 12:160–175.

Holland, P. W., and S. Leinhardt. 1981. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* 76(373):33–50.

Hunter, D. R. 2007. Curved exponential family models for social networks. *Social Networks* 29(2):216–230.

Hunter, D. R., and M. S. Handcock. 2006. Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics* 15(3):565–583.

Krivitsky, P. N., and M. S. Handcock. 2014. A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1):29–46.

Krivitsky, P. N., M. S. Handcock, and M. Morris. 2011. Adjusting for network size and composition effects in exponential-family random graph models. *Statistical methodology* 8(4):319–339.

Krivitsky, P. N., M. S. Handcock, A. E. Raftery, and P. D. Hoff. 2009. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks* 31:204–213.

Lazega, E. 2001. *The collegial phenomenon: The socail mechanisms of cooperation among peers in a corporate law partnership*. Oxford, UK: Oxford University Press.

Lazega, E., and P. E. Pattison. 1999. Multiplexity, generalized exchange and cooperation in organizations: A case study. *Social Networks* 21:67–90.

Lee, L.-F., X. Liu, and X. Lin. 2010. Specification and estimation of social interaction models with network structures. *The Econometrics Journal* 13(2):145–176.

Minhas, S., P. D. Hoff, and M. D. Ward. 2016. A new approach to analyzing coevolving longitudinal networks in international relations. *Journal of Peace Research* 53(3):1–15.

Moody, J. 2001. Race, school integration, and friendship segregation in America. *American Journal of Sociology* 107(3):679–716.

Nowicki, K., and T. A. B. Snijders. 2001. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96(455):1077–1087.

Resnick, M. D., P. S. Bearman, R. W. Blum, K. E. Bauman, K. M. Harris, J. Jones, J. Tabor, T. Beuhring, R. E. Sieving, and M. Shew. 1997. Protecting adolescents from harm: Findings from the national longitudinal study on adolescent health. *Journal of the American Medical Association* 278(10):823–832.

Robins, G., P. Pattison, Y. Kalish, and D. Lusher. 2007. An introduction to exponential random graph (p*) models for social networks. *Social Networks* 29(2):173–191.

Salter-Townshend, M., and T. Brendan Murphy. 2015. Role analysis in networks using mixtures of exponential random graph models. *Journal of Computational and Graphical Statistics* 24(2):520–538.

Saul, Z. M., and V. Filkov. 2007. Exploring biological network structure using exponential random graph models. *Bioinformatics* 23(19):2604–2611.

Snijders, T. A. 2001. The statistical evaluation of social network dynamics. *Sociological methodology* 31(1):361–395.

Snijders, T. A. B. 2002. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure* 3(2):1–40.

Snijders, T. A., P. E. Pattison, G. L. Robins, and M. S. Handcock. 2006. New specifications for exponential random graph models. *Sociological Methodology* 36(1):99–153.

Snijders, T. A., G. G. Van de Bunt, and C. E. Steglich. 2010. Introduction to stochastic actor-based models for network dynamics. *Social Networks* 32(1):44–60.

Stan Development Team. 2015. Version 2.9.0.

Strauss, D., and M. Ikeda. 1990. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association* 85(409):204–212.

van Duijn, M. A. J., K. J. Gile, and M. S. Handcock. 2009. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks* 31:52–62.

Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3):439–454.

Wang, Y. J., and G. Y. Wong. 1987. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* 82(397):8–19.

Ward, M. D., J. S. Ahlquist, and A. Rozenas. 2013. Gravity's rainbow: A dynamic latent space network for the world trade network. *Network Science* 1(1):95–118.

Ward, M. D., and P. D. Hoff. 2007. Persistent patterns of international commerce. *Journal of Peace Research* 44(2):157–175.

Wasserman, S., and P. Pattison. 1996. Logit models and logistic regressions for social networks: I. An introduction to markov graphs and *p*. *Psychometrika* 61(3):401–425.