Substantive Implications of Unobserved Heterogeneity: Testing the Frailty Approach to Exponential Random Graph Models *

Janet M. Box-Steffensmeier† Benjamin W. Campbell‡
Dino P. Christenson§

Abstract

Exponential Random Graph Models (ERGMs) are an increasingly common tool for inferential network analysis. However, a potential problem for these models is the assumption of correct model specification. Through six substantive applications (Mesa High, Florentine Marriage, Military Alliances, Militarized Interstate Disputes, Regional Planning, Brain Complexity), we illustrate how unobserved heterogeneity and confounding leads to degenerate model specifications, inferential errors, and poor model fit. In addition, we present evidence that a better approach exists in the form of the Frailty Exponential Random Graph Model (FERGM), which extends the ERGM to account for unit or group-level heterogeneity in tie formation. In each case, the ERGM is prone to producing inferential errors and forecasting ties with lower accuracy than the FERGM.

Keywords: Inferential Network Analysis; ERGM; Unobserved Heterogeneity; Frailty Term; Model Fit; Simulated Networks; Florentine Marriage; Military Alliances; Regional Planning; Militarized Disputes; Brain Networks.

*This work was generously supported by the National Science Foundation's Methodology, Measurement, and Statistics Program and Political Science Program Awards #1528739 & #1528705.

†JABS: Vernal Riffe Professor of Political Science and Sociology, The Ohio State University, e: steffensmeier.2@osu.edu

‡BWC: Corresponding Author, Doctoral Candidate in Political Science, The Ohio State University, e: campbell.1721@osu.edu

§DPC: Associate Professor of Political Science, Boston University, e: dinopc@bu.edu

Exponential Random Graph Models (ERGMs) are an increasingly common tool used to draw meaningful inferences from network data. Indeed, the ability to model the effects of nodal, dyadic, subgraph, and network covariates on network generation makes this tool exceptionally powerful. Like other likelihood-based models, using the ERGM to identify a causal generative process requires the assumption that the model is correctly and completely specified. In other words, central to using ERGMs to identify causal effects is the assumption that there is no omitted variable bias. While there exist a plethora of tools to overcome this barrier in generalized linear models or survival models, few tools have been innovated for those interested in employing ERGMs. This problem is particularly acute as incorrectly specified ERGMs are prone to degeneracy, inferential errors, and poor model fit as terms may be highly collinear (Hunter et al., 2008).

Currently, the only approach to overcome this significant barrier is based on an extension of the ERGM to account for unit and group-level heterogeneity in sender or receiver effects with the inclusion of a frailty component, the Frailty Exponential Random Graph Model (FERGM) (Box-Steffensmeier et al., 2017). The FERGM has been found to produce unbiased and consistent effect estimates while generating model estimates that forecast ties with greater accuracy than an ERGM alone.[1] However, little is known about the prevalence and effects of unobserved heterogeneity in practice, particularly in terms of the mid- to large-sized networks that constitute the bulk of ERGM applications.

In this manuscript, we shed light on how significant the problem of unobserved heterogeneity is in ERGMs in applied work, where the goal is drawing meaningful inferences or forecasting out-of-sample. This problem is illustrated through six substantive applications to canonical networks across a range of disciplines, including the Mesa High, Florentine Marriage, Military Alliances, Militarized Interstate Disputes, Regional Planning, and Brain Complexity networks. For each of these applications, the problem of unobserved heterogeneity matters in important ways. In particular, we find that the inferences drawn from model estimates differ, and the FERGM routinely out-predicts the ERGM. These applications demonstrate the dangers and prevalence of unobserved heterogeneity, as well as the great promise of the FERGM for the increasing number of social and natural scientists relying on ERGMs for network inference.

1 The ERGM and Its Limitations

The Exponential Random Graph Model (ERGM) is a flexible statistical framework for jointly modeling the influence of exogenous covariates and endogenous network dependencies on relational outcomes. In recent years, the ERGM has become a tool commonly used for estimating the effect of covariates and complex interdependencies alike (e.g., Wasserman and Pattison, 1996; Snijders, 2002; Snijders et al., 2006; Robins et al., 2007a,b; Cranmer and Desmarais, 2011; Box-Steffensmeier and Christenson, 2014, 2015; Cranmer et al., 2016;

[1] Indeed, the article that introduced the FERGM won the Editors' Choice Article in 2018 (`https://www.cambridge.org/core/journals/political-analysis/editors-choice-articles`), indicating that they see the work as "providing an especially significant contribution to political methodology."

Victor et al., 2016). In this section we briefly review the standard cross-sectional ERGM, highlighting a significant limitation of the model: its failure to account for unobserved heterogeneity.[2] This limitation manifests itself in three observable problems in ERGMs: degeneracy, inferential errors, and model fit. In the following section we revisit the FERGM with particular attention to how the frailty approach addresses unobserved heterogeneity and what that should mean for applied work.

1.1 ERGM and Estimation

Consider a network N that is constituted by n actors and a series of dyadic relationships between those actors, marked n_{ij} for the presence of a tie between nodes n_i and n_j. The types of actors in n or the relationships included, are arbitrary, but ultimately constitute N. N is typically characterized by an adjacency matrix, and in the case of the binary unimodal networks typically utilized within an ERGM, N is an $n \times n$ matrix where each element refers to the existence or non-existence of a tie, $n_{ij} \in \{0, 1\}$.[3] When $n_{ij} = 1$ necessary implies that $n_{ji} = 1$, then the adjacency matrix is symmetric and the network is said to be undirected. If $n_{ij} = 1$ and it is possible for $n_{ji} = 0$, the adjacency matrix is said to be asymmetric and the network is directed. Typically, the n_{ii} diagonals of this matrix are undefined, in other words, loops are typically ignored.

The ERGM explicitly models the probability of observing N conditioned on a set of model terms which is comprised of nodal, dyadic, subgraph, and network covariates.[4] The ERGM is canonically derived as:

$$\mathcal{P}(N \mid \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}'\mathbf{h}(N^*)\}} \tag{1}$$

where $\boldsymbol{\theta}$ refers to the parameters informing the generative model of the network and $\mathbf{h}(N)$ refers to a set of statistics computed on the network. The denominator in the former equation reflects a normalizing term constituted by all possible permutations of network N, $\mathcal{N}$.

The estimation of ERGMs via maximum-likelihood estimation is difficult as the normalizing constant can be computationally intractable. In cases where networks are particularly small, true estimation via maximum likelihood is tractable. However, for relatively large networks traditional maximum likelihood estimation is simply not a computationally tractable option. The logic is relatively simple, the normalizing constant must be recomputed as $\boldsymbol{\theta}$ is updated, and for large networks the number of permutations in $\mathcal{N}$ can increase to incomprehensibly large numbers. As a result, this normalizing constant is typically approximated through Markov chain Monte Carlo maximum likelihood estimation (MCMC-MLE) (Snijders, 2002), although other less commonly used methods exist, including maximum psudolikelihood estimation (MPLE) (Wasserman and Pattison, 1996; Anderson et al., 1999).

[2]This limitation has been addressed in non-ERG network models (Hoff, 2008; Krivitsky et al., 2009; Minhas et al., 2016).

[3]However, it is worth noting that there have been generalizations of this model to different tie classifications (Desmarais and Cranmer, 2012; Wilson et al., 2017).

[4]The ERGM, largely developed by Holland and Leinhardt (1981) and Frank and Strauss (1986), was first fully derived by Wasserman and Pattison (1996).

While a thorough review of ERGM estimation is neither novel or within the scope of this piece, a brief review is necessary to understand the problems associated with ERGM estimation.[5] MCMC-MLE involves the simulation of a distribution of random graphs that are most similar to the network observed. With these simulated networks that best approximate the normalizing constant, values of $\boldsymbol{\theta}$ are updated and refined until there is little change in the likelihood and the parameter estimates appear to have stabilized (Snijders, 2002). MPLE offers a quick and pragmatic alternative to MCMC-MLE, but may ultimately produce biased confidence intervals under strong dependence among observations (Robins et al., 2007a; Van Duijn et al., 2009). MPLE represents a form of change statistic regression wherein the probability of dyadic tie formation is predicted through conventional change statistic implementation of ERGM-based terms (Anderson et al., 1999). When estimating ERGMs, a series of problems may emerge which we will discuss in the following pages.

1.2 Unobserved Heterogeneity

Modeling networks using the Exponential-family approach is prone to a significant problem in statistical inference: unobserved heterogeneity (Hoff, 2005; Thiemichen et al., 2016; Box-Steffensmeier et al., 2017). Those familiar with survival or event history models may be familiar with a parallel problem: some units may be more or less prone to experience an event based upon unobserved unit-level heterogeneity. In the classic example of cardiac health, some individuals may be more or less likely to experience a heart attack based upon factors that are difficult, if not impossible, to observe in observational data: diet, genetic makeup, etc. In network modeling, a similar dynamic plays out, some individuals may be more or less likely to form friendships based upon factors that are difficult, if not impossible, to observe in observational data: charisma, personality, etc. In modeling either heart attacks or friendship ties, this unobserved unit-level variation may manifest itself in a variety of modeling challenges. When using ERGMs to model network formation under unit-level heterogeneity, three particular challenges emerge: degeneracy, omitted variable bias, and inferior model fit. We will discuss each of these briefly in turn, saving an extended discussion for the following section.

1.2.1 Degeneracy

Any analyst estimating ERGMs through MCMC-MLE has likely encountered a degenerate model specification. Degeneracy is an estimation problem that occurs when the specification of a model is so unlikely to have generated the network that an ERGM cannot be estimated (Handcock et al., 2003). When using MCMC-MLE, a poorly specified model will lead the Markov chains to move to extreme ends of the graph – perfectly empty or perfectly full networks – where the chain will stay and produce errors or poorly fitting models. While degeneracy is not particularly a problem as it indicates a poorly fitting model, it may prevent

[5] For further discussion, we would direct the reader to Anderson et al. (1999), Snijders (2002), Robins et al. (2007a), and Van Duijn et al. (2009).

an analyst from considering a desired theoretically-motivated specification, or from fitting a model to particularly dense or sparse networks.

1.2.2 Omitted Variable Bias

Like most statistical models, ERGMs require the assumption that models are correctly and completely specified, that there is no omitted variable bias. This is an untestable assumption, and violating it threatens the credibility of inferences that can be drawn from ERGMs. This latent and unobserved heterogeneity that exists among units in their motivations for tie formation may confound relationships of interest, and as such, lead to Type 1 or Type 2 errors. Given the collinearity inherent in networks between structural model terms, this problem is particularly acute (Hunter et al., 2008).

1.2.3 Inferior Model Fit

The failure of ERGMs to account for unobserved heterogeneity and their risk of creating inferential errors certainly has substantive implications, but it also has significant implications for model fit. If the estimates from model terms are biased from variable omission, then traditional goodness of fit routines used for network models may illustrate poor model fit (Hunter et al., 2008). For example, if certain individuals are more outgoing than others, ERGMs may confuse this unit-level characteristic with a network tendency towards triadic closure and thus, networks simulated from estimated parameters will have more triangles than the observed network. As such, when analysts use ERGMs to forecast ties out-of-sample, they must be aware of the role of unobserved heterogeneity in producing poorly fitting models.

2 The FERGM Solution

As previously mentioned, an essential assumption of the ERGM is that $\mathbf{h}(N)$ reflects a vector of sufficient statistics for the network. In other words, that the model reflects the correct and complete specification of all endogenous dependencies and exogenous covariates to explain the network's generation. This assumption is strong and untestable. Should important confounders, observed or unobserved, be excluded, then the aforementioned problems plaguing ERGMs will likely emerge.

Unobserved heterogeneity may lead to problematic substantive inferences. Take the canonical example of a friendship network, where many observable factors may lead to the formation of friendships between individuals (e.g. common political beliefs, socio-economic status, age, race, etc.). It would also be expected that friendships form between actors with common friends, that there would be a tendency towards triadic closure. However, certain unobserved factors may inform whether two actors become friends, including personality characteristics like friendliness or charisma, or deviant tendencies, like drug use (Box-Steffensmeier et al., 2017). These unobservables may also be related to network structure, as more friendly nodes may create triadic closure among alters. This heterogeneity

may lead to any of the following: degenerate models as the data generating process may be incorrectly described; incorrect inferences about the prevalence of triadic closure as a generative feature of the network or the effect of other exogenous covariates; or poor model fit. We will explore these problems explicitly through a series of analyses, illustrating how these problems become manifest in ERG-family modeling, their substantive implications, and how a frailty-approach can resolve these issues.

To overcome the issue of unobserved confounding, we explore a recent extension to the ERGM, the Frailty Exponential Random Graph Model (FERGM) (Box-Steffensmeier et al., 2017), which includes a frailty term analogous to those used in event history models (Box-Steffensmeier and De Boef, 2006; Box-Steffensmeier et al., 2007).[6] The FERGM introduces individual (or group) level random effects into the $\mathbf{h}(N)$ component of the ERGM to capture the latent factors that inform the propensity for certain individuals or groups to form ties. In other words, a term is added to model the variance in individual-level degree distributions. If one assumes that there are just some actors that are more social s_i, or some actors that are just more popular r_i, and if one assumes these terms are distributed standard normal, the following directed FERGM is identified:

$$\mathcal{P}(N \mid \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{h}(N) + \sum_{i=1}^{N} s_i + \sum_{i=1}^{N} r_i\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}'\mathbf{h}(N^*) + \sum_{i=1}^{N} s_i + \sum_{i=1}^{N} r_i\}} \tag{2}$$

For an undirected network, one simply only needs to exclude the $\sum_{i=1}^{N} r_i$ term. We will focus upon this reduced version for the undirected model in our replications:

$$\mathcal{P}(N \mid \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{h}(N) + \sum_{i=1}^{N} s_i\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}'\mathbf{h}(N^*) + \sum_{i=1}^{N} s_i\}} \tag{3}$$

2.1 Estimation

The FERGM previously proposed can be estimated through a logistic multiple membership mixed-effects model. In this particular case, we can assume there are random effects associated with each node, modeling each potential undirected tie as a function of change statistics computed on the network and the previously discussed random effects:

$$Pr(n_{ij} = 1) = logit^{-1}\{\boldsymbol{\theta}'\mathbf{h}(N) + s_i + s_j\} \tag{4}$$

In the undirected case, s_i and s_j refer to the random effect frailty terms associated with nodes i and j, $\mathbf{h}(N)$ is a vector of change statistics for when a tie n_{ij} is toggled on and off, and $\boldsymbol{\theta}$ is the vector of coefficients associated with these change statistics. For those familiar with ERGM estimation, when excluding s_i and s_j, one is left with the MPLE estimator for the ERGM.

Estimating the FERGM through MPLE in this way offers a variety of advantages. First, and foremost, the FERGM is not susceptible to degeneracy given that it does not rely upon

[6] The related R package for the FERGM (Morgan et al., 2018) is available online at CRAN.

MCMC-MLE estimation. This is a significant benefit to those who might be interested in testing an ideal model specification which might otherwise be degenerate. Second, the FERGM can be estimated using standard mixed-effects modeling packages available in `R` and Stan. Third, given that the technique does not rely upon MCMC sampling, model estimation is scalable and for larger networks occurs at a fraction of the time of conventional ERGMs. One potential criticism of using MPLE is that confidence intervals may be artificially narrow (Van Duijn et al., 2009). However, previous work has suggested that the estimated credible intervals for the FERGM are likely accurate (Box-Steffensmeier et al., 2017).

2.2 Prior Approaches to Accounting for Heterogeneity

With the FERGM introduced, we now turn to distinguishing our approach from others. We should note that there has certainly been work on unit-level heterogeneity within a network context, but much of this work has explicitly failed to account for unit-level heterogeneity as an omitted variable while allowing users to comfortably use the ERGM-based framework they are accustomed to. Fellows and Handcock (2012) introduce an endogeneous model, the Exponential Random Network Model (ERNM), representing selection-diffusion problem, providing an opportunity to disentangle the relationship between social influence and social selection. This approach is distinct as it does not attempt to model unit-level heterogeneity in the sociality of nodes from a random-effects based approach. Others have advocated for a latent factor-based approach to representing unit-level heterogeneity, modeling network dynamics and heterogeneity in degree distributions as low-dimensional nodal attributes (Hoff, 2005, 2009; Fosdick and Hoff, 2015; Minhas et al., 2016). This approach, while accounting for network dependencies and unit-level heterogeneity, do not allow for hypothesis testing within the ERGM-based framework that many are comfortable using.

Other approaches to unit-level heterogeneity have adopted a mixture of data generating processes approach, stepping outside of the ERGM framework. Conventional approaches, such as the Stochastic Block Model or the Mixed-Membership Stochastic Block Model attempt to model heterogeneity in the probability that two actors of two given blocks form a relationship as a function of network position (Holland et al., 1983; Snijders and Nowicki, 1997; Airoldi et al., 2008) and/or covariate data (Gormley and Murphy, 2010; White and Murphy, 2016). More recently, the ego-ERGM and ego-TERGM were introduced to represent cases where actors within a broader network are assumed to have heterogeneous motives for forming relationships or tie generating processes (Salter-Townshend and Brendan Murphy, 2015; Box-Steffensmeier et al., 2018; Campbell, 2018a).

Thiemichen et al. (2016) introduce a fully Bayesian approach to the ERGM, the Bayesian ERGM (BERGM), that can account for unit-level heterogeneity through random effects. While this approach is certainly similar to the FERGM introduced by Box-Steffensmeier et al. (2017) and discussed here, the models differ in important ways. Perhaps most importantly, the BERGM requires the analyst to select a variety of hyperparameters, including the choice of priors for the random effects. This choice is not required by the FERGM as random effects are imposed to be mean zero centered. We direct the reader to Box-Steffensmeier et al. (2017) for further discussion. It should be underscored that the goal of

this piece is not to distinguish between these models or to make the case for one over the other, but to demonstrate the relative importance of accounting for unobserved heterogeneity in the ERGM framework. Other approaches have documented the properties of these random effects-based network models and their asymptotic properties (Yan et al., 2018). Given our experience and ability to remain agnostic to hyperparameter decisions, we choose the FERGM to indicate the importance of accounting for unobserved heterogeneity.

2.3 Improved Model Performance

In this section, we discuss the how the FERGM overcomes the previously articulated problems likely to be plaguing the ERGM across a host of scholarly domains. While Box-Steffensmeier et al. (2017) presented the model and its proof of concept, we extend this discussion by further detailing the benefits of the model, namely its ability to avoid degeneracy and omitted variable bias, and its ability to produce more accurate predictions out-of-sample. We also provide a unique and important contribution in showing the substantive implications of the ERGM's limitations and how the FERGM can fix them. We begin by discussing the issue of degeneracy in ERGMs, and how the FERGM specifically avoids this pitfall common in network modeling. We then move to discussing omitted variable bias in ERG-family modeling, and how the FERGM's frailty terms account for this unobserved heterogeneity. Third, we assess the relative out-of-sample predictive performance of the ERGM and FERGM. In the following section we present a series of replication exercises to illustrate the FERGM's improvements over the ERGM.

2.3.1 Degeneracy

One significant feature of ERGMs estimated through MCMC-MLE is model degeneracy. Degeneracy describes a feature of the MCMC-based estimation procedure of ERGMs wherein algorithms converge towards graphs that are either empty or complete, or do not consistently converge (Handcock et al., 2003). This problem is a particular feature of how the normalizing constant is approximated through MCMC. Those who have attempted to estimate ERGMs have likely experienced this problem and resorted to alternative model specifications in an attempt to find an identifiable model. While this problem may in many cases be a guard against a poorly fitting model (Handcock et al., 2003, 7), for substantive scholars it may reflect an obstacle to estimating a theoretically-motivated model.

The FERGM is not prone to degeneracy as it does not rely upon MCMC to approximate the normalizing constant. As noted, one potential concern is that confidence intervals estimated through MPLE will be artificially narrow (Van Duijn et al., 2009). However, Box-Steffensmeier et al. (2017) have indicated that the confidence intervals estimated for the FERGM are consistent and unbiased. As such, one of the largest drawbacks of ERG-family modeling may be resolved when accounting for unobserved heterogeneity through the use of frailty terms.

2.3.2 Omitted Variable Bias

A core modeling assumption of most likelihood based models, the ERGM included, is that the terms included reflect the full data generating process. This assumption is rarely met in observational data as a variety of factors influencing an outcome may be unobserved, including individual-level factors that may predispose an observation to observing an event. This problem lead to the emergence of random-effects models and frailty terms to measure this unobserved heterogeneity. Traditionally used in survival or event history modeling (Box-Steffensmeier and De Boef, 2006; Box-Steffensmeier et al., 2007), frailty models have been widely used to improve model fit and estimate unbiased effects under the presence of unit-level heterogeneity.

Within the context of network data, this unobserved heterogeneity may manifest itself in influencing whether certain observations are more or less likely to form relationships than others. Unfortunately, failing to model this heterogeneity directly in ERGMs poses significant inferential challenges given the excessive collinearity and subtle differences in network dependencies (Hunter et al., 2008). Through including unit or group-level frailty terms, the FERGM provides a means of accounting for this omitted variable bias and unobserved heterogeneity, improving model fit and the model's inferential accuracy.

2.3.3 Out-of-Sample Model Fit

Unobserved confounding and omitted variable bias may undermine the predictive power and model fit of ERGMs. Biased coefficients and the inability to account for important unobserved features may create serious problems for model fit by creating inaccurate estimates. If these inaccurate coefficients are used to simulate networks and evaluate model fit as described by Hunter et al. (2008), then the model may appear to fit poorly. Alternatively, if the observed estimates are compared to those calculated on alternative realizations of the network, the model may appear to fit poorly or the alternative realizations may appear to be of a distinct generative process. Regardless, the problem of unobserved confounding has significant consequences for evaluating the fit and predictive performance of ERGMs.

The FERGM overcomes these problems by producing more accurate effect estimates and ensuring that any unobserved confounding that may undermine forecasts is accounted for explicitly.[7] By accounting for unit (or group) level heterogeneity, FERGMs should generally produce better fitting models out-of-sample. Indeed, never in our Monte Carlos or the replication analyses presented in the following section, do we find that FERGMs do not out predict their ERGM equivalent. Thus, it appears that FERGMs appear to routinely fit better out-of-sample than ERGMs, all while showing no significant evidence of overfitting, at least to the extent that it would offset any gains made through including a frailty term.

[7] A Monte Carlo study presented in the Supplementary Information (SI) Appendix illustrates that under a variety of very difficult conditions, such as excluding important covariates, the FERGM will better predict tie formation out-of-sample than the ERGM.

3 Replication Studies

To illustrate the improvements made by accounting for unobserved heterogeneity and using the FERGM over the ERGM, we conduct six analyses that attempt to replicate either published analyses or findings drawn from canonical networks. Specifically, we examine the relative changes in substantive inferences drawn from the models and predictive power for ERGMs and FERGMs fit on the same network with the same set of variables. These six networks chosen reflect a great deal of intellectual diversity – two canonical social networks (Mesa High, Padgett's Florentine Marriage), two International Relations networks (Military Alliances and Militarized Interstate Disputes), one from American Politics (Regional Land-Use Planning), and one from Neuroscience (Brain Networks). Once we collected these networks, we estimated a well-fitting ERGM on each of them. In each case we attempted to approximate a published model specification associated with each network, but were typically constrained by data availability, degeneracy, and poor model fit. Once these well-fitting ERGMs were estimated, the corresponding FERGM was estimated using the aforementioned MPLE routine.

When examining substantive implications, we assess how the inferences made from the model estimates changes and how that may influence the theoretical and empirical conclusions drawn. While it is impossible to know what the true effect of a particular variable might be, we hope to underscore how the hypotheses one might conclude supporting might change as a result of accounting for unobserved heterogeneity through the FERGM. From comparing the ERGM and FERGM estimates and their relative robustness, one may not be able to definitively conclude that one better approximates reality than the other. We hope to make the case that the FERGM estimates are more credible through discussing what unobserved confounders may be lurking and the relative predictive performance of the ERGM and the FERGM. When examining predictive performance, we use a tie prediction routine paralleling the one introduced by Box-Steffensmeier et al. (2017). In this section we will begin by discussing each replication in great detail, concluding with a discussion of our general findings. Our findings reveal that accounting for unobserved heterogeneity through using the FERGM has significant implications for the inferences we draw from ERG-family models and for our ability to predict tie formation.

3.1 Applications of the FERGM

The previous section has demonstrated the methodological advantages to using the FERGM over the ERGM, including its ability to model unobserved heterogeneity, overcome the common problem of degeneracy, and produce superior model fit. In this section, we illustrate the real world problem of unobserved heterogeneity and how it impacts model estimation and influences the substantive implications drawn from real world studies.

In particular, we conduct six replication analyses that take a well-known network and attempt to replicate well-known analyses and findings using ERGMs. We then fit FERGMs and compare the fits to assess how the substantive inferences drawn between the two change, and whether one fits better than the other. Again, we must underscore that while we try

to build the case that the FERGM estimates should be more credible by including a frailty component and producing better predictive accuracy, it is impossible to know whether one model would produce inferences that reflect the actual data generating process. We begin by examining the Mesa High network (Hunter et al., 2008). This network is the simulation of an in-school friendship network that reflects a network collected in the AddHealth Study (Resnick et al., 1997). We then examine Padgett's well-known and canonical Florentine Marriage network (Padgett and Ansell, 1993). Our third replication examines whether unobserved confounding impacts our support for the conventional view of military alliance formation through analyzing the defensive commitments network, a well studied network in International Relations (Cranmer et al., 2012a,b). We then move to the network of militarized interstate disputes to examine whether our inferences about a well-known empirical regularity, the democratic peace, are impacted by accounting for network effects and unit-level heterogeneity (Cranmer and Desmarais, 2011; Campbell et al., 2018). For additional diversity, our fifth exercise examines a network that is widely known in the study of American Politics which captures an institutional collective action context in regional land-use planning networks (Gerber et al., 2013). We end with an interesting exercise examining brain networks, an increasingly attractive area for network modeling (Simpson et al., 2012; Stillman et al., 2017).

These applications were chosen for two reasons. First, they reflect a great deal of intellectual diversity. Second, they reflect distinct network generative processes with acknowledged but unmodeled heterogeneity in the motivations for tie formation. In other words, the applications demonstrate the need for a frailty approach to ERG-modeling. For each replication, the substantive inferences drawn from the ERGM and the FERGM and their relative model fit vary significantly, demonstrating how neglected heterogeneity may impact the accuracy of our inferences.

3.1.1 Replication 1: Mesa High

The Mesa High network, illustrated in Figure 1, is a network simulated by Hunter et al. (2008) to mirror an in-school friendship network collected by the AddHealth Study (Resnick et al., 1997). This network, even while simulated, is perfect for our replication study as the properties of the network are well-known. Given that it is a network simulated according to a set of ERGM parameters, there should be little improvement in model fit or changes in substantive implications when using the FERGM as there is almost by definition no unobserved heterogeneity. If there was, then this may actually be cause for concern.

ERGMs (MCMC-MLE estimated) and FERGMs were fit on this network using the following terms: Edges, Alternating K-Stars ($\alpha = 0.6$), GWESP ($\alpha = 0.2$), Race Homophily, Grade Homophily, and Sex Homophily. When comparing model fit, the ERGM and FERGM are nearly indistinguishable, a $< 0.1\%$ decrease in tie forecasting ability, as illustrated in Figure 13. As this is a simulated network with a known generating process, the minuscule difference in predictive model fit when using the FERGM over the ERGM makes sense.

Also to be expected, there are no significant differences between the ERGM and FERGM fits in the great majority of inferences drawn from model estimates. All terms, with the ex-

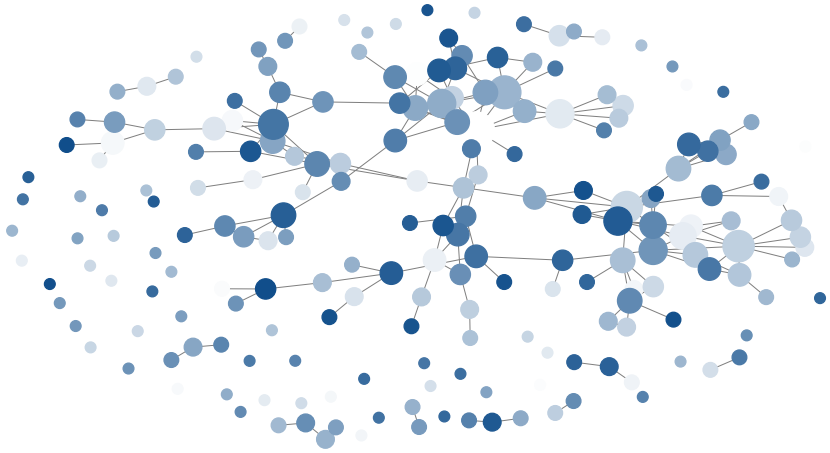

Figure 1: Mesa High Network. Nodes are sized and colored according to increasing degree.

ception of Alternating K-Stars, have similar estimates, standard errors, and significance levels. The difference in effects for this term is not particularly troubling. First, the ERGM-estimated effect for this term is already fairly small with relatively wide confidence intervals. As such, it is possible that even slight changes in model specification may change the significance of this term. Second, it is possible that our frailty term capturing unobserved sender/receiver effects is collinear with the Alternating K-Stars term capturing sociality/popularity effects.

The comparisons made here between ERGM and FERGM fits do not directly highlight the importance of unobserved heterogeneity for model fit and substantive inference. They do, however, highlight that the FERGM does not overfit the data and produce flawed inferences when attempting to account for latent sender/receiver effects.

3.1.2 Replication 2: Florentine Marriages

The Florentine Marriage Network, collected by Padgett and Ansell (1993), is a collection of marriage alliances among Renaissance Florentine families. Visualized in Figure 3, this network is selected given its canonical status, and as such, most readers should be familiar with the network and its properties.

As is with prior cases, the best-fitting ERGM found and its FERGM equivalent were fit on the Florentine Marriage network. This specification includes the following terms: Edges, Triangles, Degree (2), Degree (3), Wealth Absolute Difference, Priorates, and a node's Total Ties. When comparing model fit via tie prediction per Figure 13, the FERGM fairs much better, yielding an improvement in tie prediction greater than 10%. This makes a great deal of sense as, naturally, individuals and families in Renaissance Florence may have heterogeneous reasons or incentives for marrying one another. While the relative wealth of families or their political power may explain marriage, there are a variety of interpersonal and member-variant factors that may attract some members to one another.

A natural question emerging from this improvement in model fit is whether there are

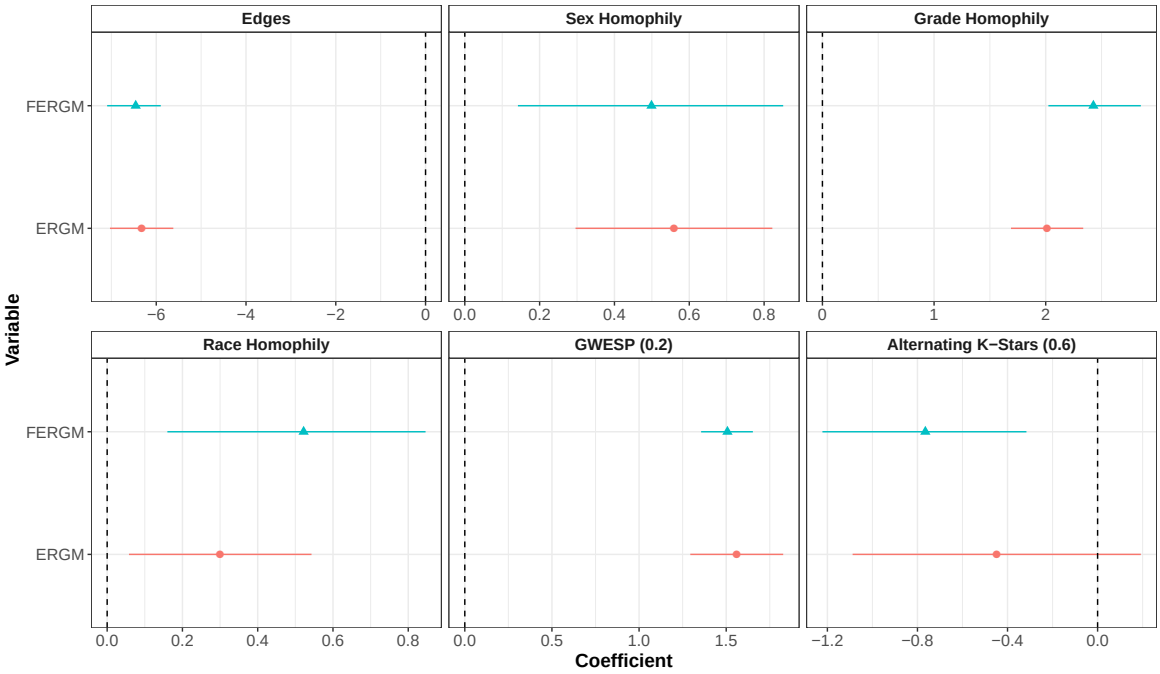

Figure 2: Coefficient Plot, Mesa High Network. 95% confidence intervals presented.

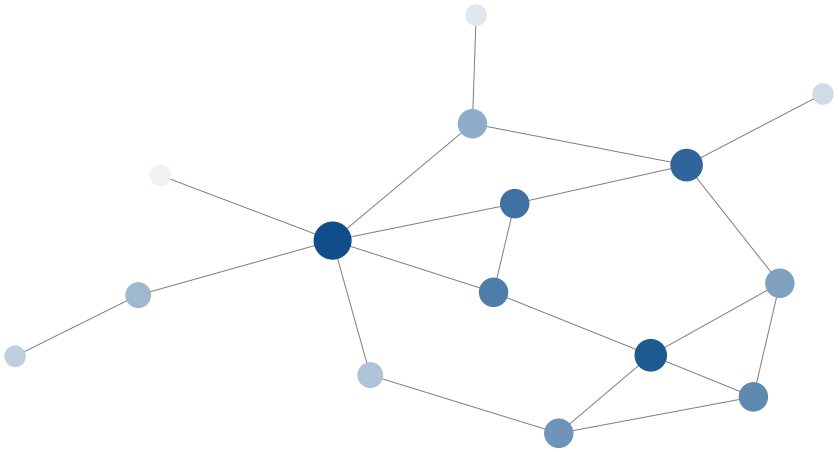

Figure 3: Florentine Marriage Network. Nodes are sized and colored according to increasing degree.

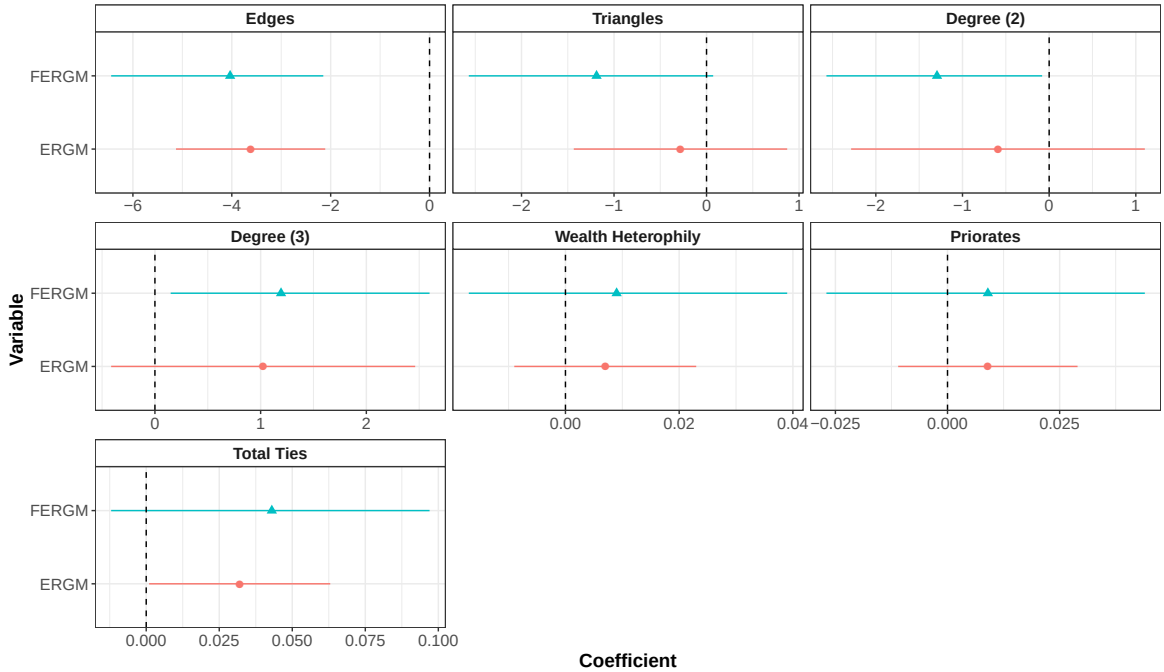

Figure 4: Coefficient Plot, Florentine Marriage Network. 95% confidence intervals presented.

dramatic differences in the substantive inferences drawn from the ERGM and FERGM. The answer is presented in Figure 4. Interestingly enough, while there are changes in the inferences drawn from certain variables, such as Degree (2), Degree (3), and Total Ties, many of the variables of interest such as Edges, Triangles, Wealth Heterophily, and the number of Priorates remain unchanged. This is interesting, as there are a variety of unobserved sender/receiver effects that may likely be collinear with these variables, such as a family's sociality or popularity.

Nevertheless, the relative differences in model fit and inferences drawn underscore the importance of accounting for heterogeneity in actors' tie forming processes. In the following application, a more theoretically motivated discussion is used to illustrate how these differences may influence our knowledge about important international phenomena.

3.1.3 Replication 3: Military Alliance Formation

Conventionally, countries are thought to form defensive military alliances as responses to external threats (Carr, 1946; Morgenthau, 1948; Waltz, 1979; Walt, 1990). With the birth of quantitative studies of alliances, however, this logic has been challenged as scholars begin to embrace the idea that alliance decisions may be a function of power politics (Lake, 2009), trade (Long, 2003; Powers, 2004; Sprecher et al., 2006; Fordham, 2010), and regime change (Pevehouse, 2002). Recently, research has noted that the decisions of states to form alliances with one another may not be independently distributed, noting a strong tendency towards triadic closure (Cranmer et al., 2012a,b).

While the gains made by Cranmer et al. (2012a) and Cranmer et al. (2012b) are cer-

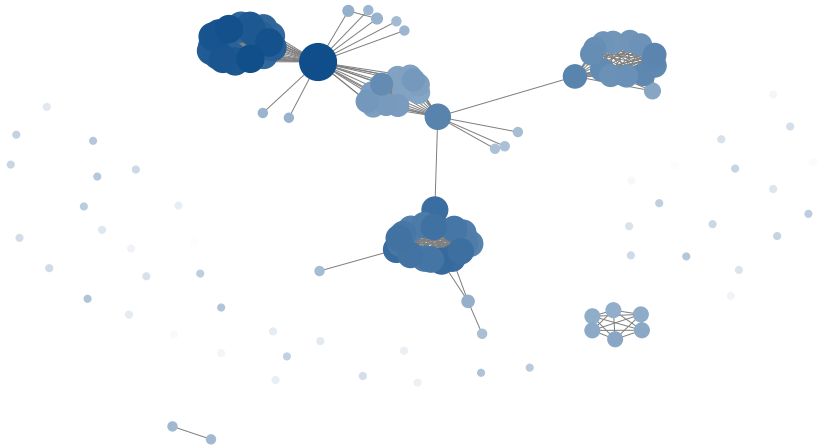

Figure 5: Defensive Pact Network, 1985. Nodes are sized and colored according to increasing degree.

tainly dramatic, one problem that has plagued alliance politics scholars for decades has been the ability to jointly model complex interdependencies in the alliance network and the documented unobserved heterogeneity in the decision of two states to form an alliance (Campbell, 2018b). This heterogeneity may be considered as a function of many unobserved sender or receiver effects, including unobserved asymmetric gains from the alliance (Morrow, 1991), *quid pro quos* (Snyder, 1997), or other idiosyncratic characteristics of leaders, relationships, or strategic environments (Taylor, 1954; Schroeder, 1996; Bridge and Bullen, 2014).

To demonstrate how unobserved heterogeneity may impact the inferences drawn from conventional ERGMs, we apply the FERGM to a cross-sectional model approximating the Cranmer et al. (2012a) (CDK) model specification. In particular, we examine the undirected alliance network for a system reflecting a late stage of the Cold War, 1985. This network is presented in Figure 5 which reveals quite a bit of interesting network structure. For this year we fit a model that includes Edges, Geometrically Weighted Edgewise Shared Partners (GWESP), Militarized Interstate Dispute (MID) History, Common Enemy, and Contiguity. We also measure the difference in regime scores for the states, Regime Difference, and a node-level covariate for state capabilities (CINC).

To get a sense of the difference in effect estimates from the ERGM to the FERGM, we direct the reader to Figure 6. We find three differences between the CDK-inspired ERGM and its FERGM equivalent. First, the state capabilities term that the ERGM finds support for, indicating that great powers are more likely to form alliances, drops off when adding the frailty term. Specifically, for the ERGM a positive and statistically significant result is detected, but for the FERGM a negative effect indistinguishable from zero is uncovered. This increased uncertainty may be a function of unobserved asymmetric gains associated with alliances including great powers (Morrow, 1991). Second, the ERGM would support the conventional logic of states forming alliances to counter common enemies (Carr, 1946; Morgenthau, 1948; Waltz, 1979; Walt, 1990), as evidenced by the positive and statistically significant effect size for the common enemy term. While this effect is relatively small, when

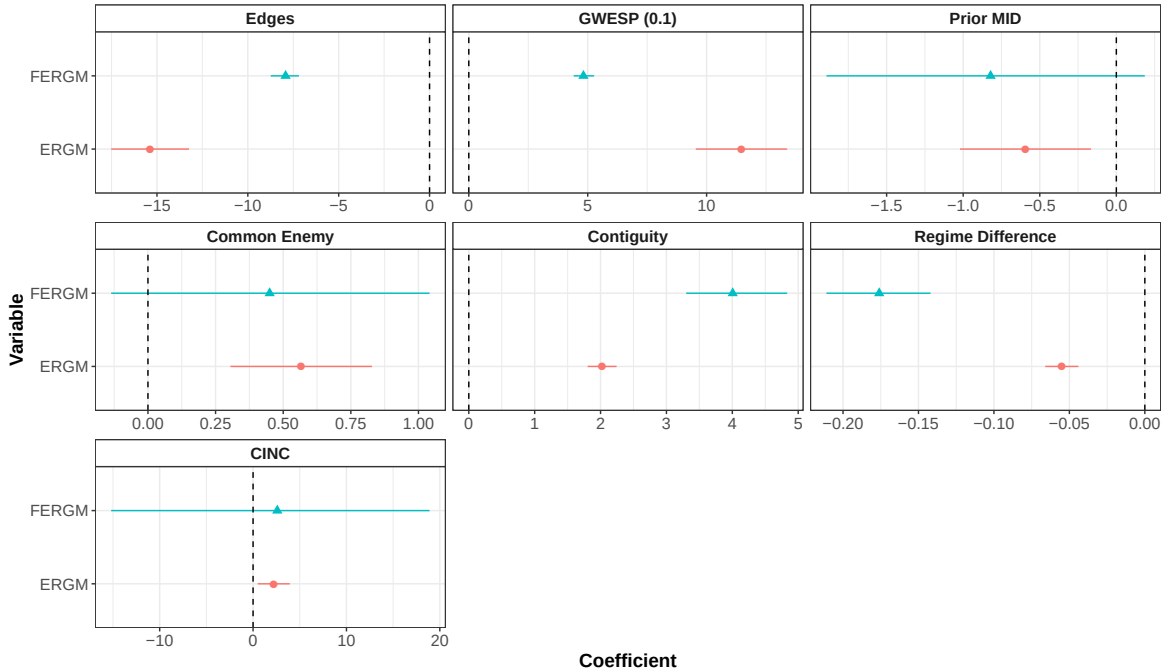

Figure 6: Coefficient Plot, Alliance Network. 95% confidence intervals presented.

using the FERGM, however, we find that this effect is positive, but even closer to zero with wider 95% credible intervals. This further shrinkage towards zero from the ERGM to the FERGM may indicate that the conventional wisdom is problematic. Third, the ERGM supports the conventional wisdom, that two states who have previously engaged in a militarized dispute should not align as they do not have common interests and may not trust one another. We find that once the frailty term is added, the effect estimates has greater uncertainty and the 95% credible interval contains 0. In addition, one may be confident in these results as there is an improvement in the ability of the model to accurately predict ties by 1.4%, as evidenced by Figure 13.

The previous comparison and the improvement in fit between the ERGM and FERGM highlights that unobserved heterogeneity may confound our ability to understand the true effects that variables have upon the generative structure of the alliance network. While the core finding of network analysis within the alliance literature remains, that there is a tendency towards triadic closure, many of the foundational insights from realist and bargaining theories of alliance formation may be problematized.

3.1.4 Replication 4: Militarized Interstate Disputes (MIDs)

The democratic peace, an empirical finding that two democracies are less likely to fight one another, has been so robust that many have referred to it as the first law of international politics (Levy, 1998; Hegre, 2014). Increasingly, there has been attention paid to understanding the mechanisms underlying the democratic peace. For example, early work emphasized normative or institutional explanations (Maoz and Russett, 1993; Russett, 1994), only to turn to

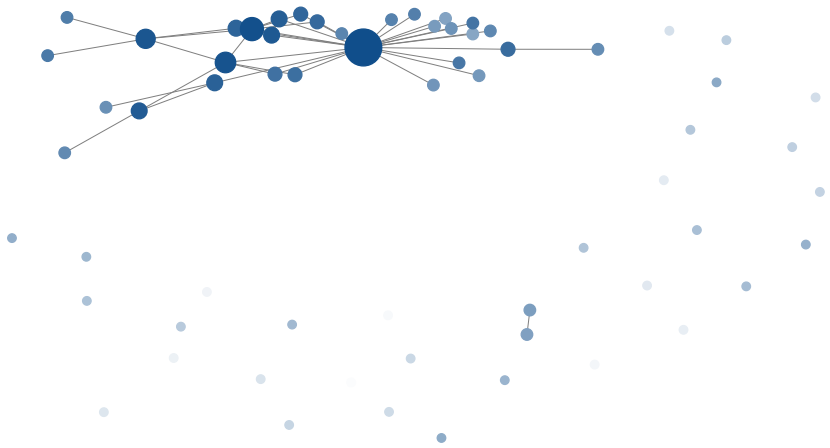

Figure 7: Militarized Interstate Dispute Network, 1939. Nodes are sized and colored according to increasing degree.

other factors collinear with regime type (Gartzke, 2007; Dafoe et al., 2013), with a recent increasing emphasis on network explanations (Cranmer and Desmarais, 2011; Campbell et al., 2018). With this turn towards network explanations, and the widely known heterogeneity that plagues militarized interstate disputes data (Box-Steffensmeier et al., 2003b,a), the study of MIDs seems a perfect application for the FERGM. A cross-section of this network is visualized in Figure 7 for 1939.

Unobserved heterogeneity in the causes of conflict may take a variety of forms, including leader or populace-based factors. For example, it would be difficult to explicitly model the factors that lead Hitler to be more aggressive and initiate World War 2. To demonstrate how this heterogeneity may influence the inferences drawn, we apply the FERGM and ERGM to a cross-sectional model approximating the Dafoe et al. (2013) model specification. In particular, we focus on the network of militarized disputes for 1939 that is of interest to many.

The best fitting ERGM found and its FERGM equivalent were fit with the following terms: Edges, Alternating K-Stars ($\alpha = 0.75$), GWESP ($\alpha = 0.3$), Capability Ratio, Highest Composite Index of National Capabilities (CINC) Score, Number of States in the System, Alliance Pact, and Weak-Link Regime Score. The democratic peace would be uncovered by a negative and statistically significant effect for the Weak-Link Regime Score dyad covariate which is measured as the lowest Polity score within a dyad (Marshall et al., 2002). As the value for this variable increases, a dyad becomes more democratic.

Overall, the FERGM was found to fit better than its ERGM counterpart. There is approximately a 2% improvement in tie prediction accuracy, as illustrated in Figure 13. There are also tremendous changes when interpreting model results. Figure 8 presents the changes in effect estimates when accounting for unobserved heterogeneity. Interestingly, the substantive inferences drawn from the models differ significantly. Many of the estimates that are significant in the ERGM are insignificant in the FERGM, including Capability Ratio,

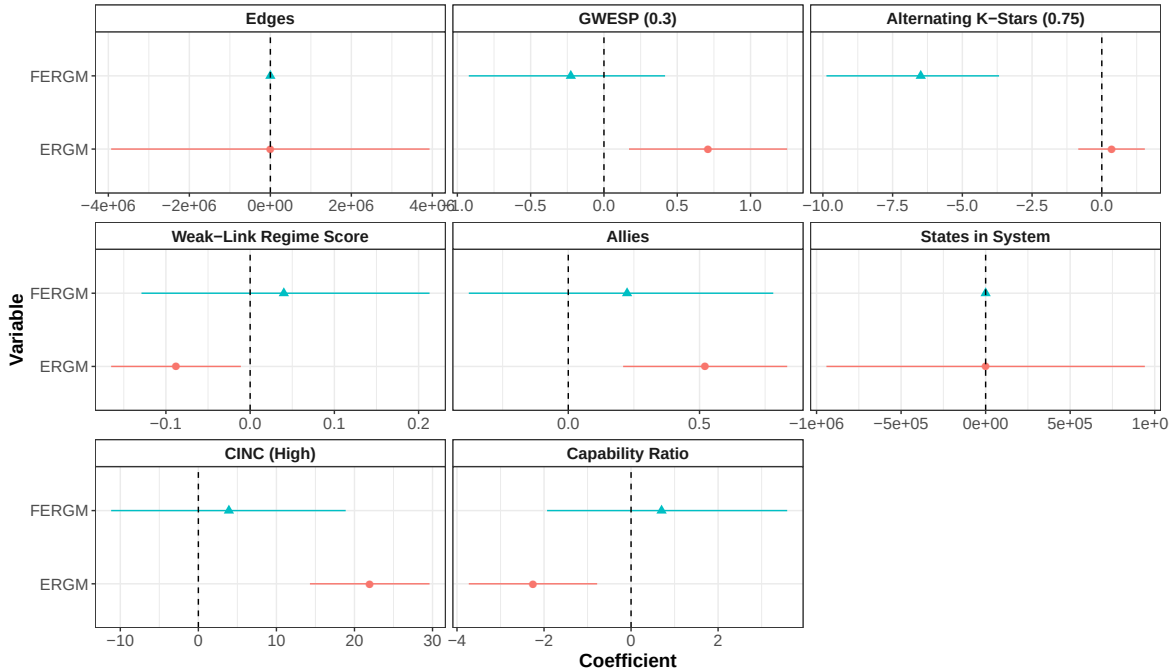

Figure 8: Coefficient Plot, Militarized Interstate Dispute Network. 95% confidence intervals presented.

CINC, Alliance Pact, GWESP, and Weak-Link Regime Score. Only the Alternating K-Stars went from insignificant to significant. Most importantly, perhaps, the democratic peace, represented by a negative effect for Weak-Link Regime Score, uncovered when using the ERGM is not detected when using the FERGM. In the ERGM a negative and statistically robust effect at $p < 0.05$ is detected for the Weak-Link Regime Score variable. While this effect is close to zero, once when a frailty term is included the effect becomes positive and statistically insignificant. In other words, when failing to account for heterogeneity in sender-receiver effects using the ERGM, the democratic peace is uncovered. This potentially underscores the importance of accounting for heterogeneity, that when using FERGM, the democratic peace is problematized.

This finding may shed light on one of the most important literatures in international relations. While some effects remain unchanged, such as the edges term or the number of states in the system, the majority of variables' effects change in interesting ways that may highlight the importance of sender or receiver effects, such as leader-based attributes like psychology, or populace-based attributes like shared identity.

3.1.5 Replication 5: Regional Planning Networks

Observers of politics, professional and casual alike, are often interested in the degree to which actors with similar political views interact or collaborate. Homophily is well established in a variety of networks (McPherson et al., 2001; Fowler et al., 2011; Sinclair, 2012; Victor et al., 2016), and in networks of political actors, those with similar political views are thought to

affiliate at higher levels (Huckfeldt and Sprague, 1995). Political homophily is now taken as a given in many social networks, and in fact is steadily becoming an empirical regularity (Fowler et al., 2011). This phenomena is well documented, and the desire of individuals to select their alters based upon political similarity has caught national attention – political homophily is being used to explain the spread of misinformation and fake news, political polarization, and ultimately Brexit and the election of Donald Trump (Margetts, 2017).

One prominent example from this literature is Gerber et al. (2013) (GHL). GHL examine political homophily in the context of land-use planning, which can influence a community's development, social structure, tax base, and quality of life (Levy, 2009). Land-use planning refers to the collaboration and competition among local governments to create land-use plans and zoning ordinances to specify allowed development levels. Regional land-use planning, the interest of GHL, refers to attempts to mitigate the negative spill-over effects associated with local land-use planning by encouraging collaboration among regional actors to address common goals. GHL are less interested in the actors and more interested in homophily within political networks, and in particular, within institutional collective action (ICA) contexts where actors weigh the benefits and transaction costs from interaction. Theoretically, GHL argue that political homophily reduces transaction costs as similar groups may have more common policy objectives, and as such, face fewer costs from their selectorates. Nevertheless, unobserved heterogeneity may confound our ability to truly understand the effect of political homophily on the generative process for land-use collaboration networks. This unobserved hetereogeneity may take the form of latent confounders, including the industries that may exist in two counties.

To examine the effect of these unobserved confounders on the inferences drawn from the GHL ERGM fit on their land-use planning network, we apply the FERGM to an approximate replication of their model. The land-use planning network in presented in Figure 9. To approximate GHL's Model 2, we include the following terms: Edges, Geometrically Weighted Edgewise Shared Partners (GWESP), Geometrically Weighted Dyadwise Shared Partners (GWDSP), Political Homophily, the distance in cities in terms of the percentage of the population that is Latino (Percent Latino), and Median Household Income. This closely mirrors GHL's Model 2 specification, save Alternating K-Stars, which are excluded as they produce an ill-fitting and degenerate model.[8]

Important and dramatic changes occur when using the FERGM, as evidenced by the effects presented in Figure 10. First, and foremost, the positive and statistically significant effect for the *Party Registration* variable that GHL rely upon becomes negative and insignificant when accounting for unobserved heterogeneity. In other words, it appears that political homophily may not influence collaboration in the land-use planning network. In addition, we find that the GWDSP has no effect in the ERGM, but it is negative and statistically significant in the FERGM, while income is significant in the ERGM but not in the FERGM. However, the GWESP remains positive across the models, indicating a strong tendency towards triadic closure within this network. Similar to the previous case, this demonstrates that unobserved heterogeneity may confound the substantive effect of key variables of in-

[8]This particular model is chosen given the availability of replication archives suitable for our purposes.

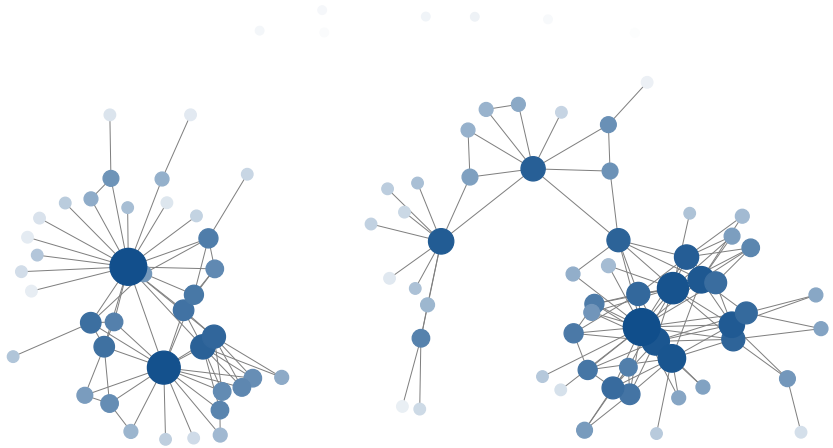

Figure 9: Land-use Planning Network. Nodes are sized and colored according to increasing degree.

terest. In addition, there is evidence that the FERGM fits better than the GHL-inspired ERGM as there is a 1.7% improvement in the ability of the FERGM to accurately predict tie formation, as presented in Figure 13.

The previous application has demonstrated that unobserved heterogeneity may confound the ability to cleanly study effects of interest, including foundational findings like homophily in political networks. When accounting for unobserved confounders, analysts can be more confident that they have gotten a clean read on effects of interest, and given the improved predictive performance, inferences may be more credible.

3.1.6 Replication 6: Brain Networks

In another interesting application, Simpson et al. (2012) argue that ERGMs are useful for modeling brain networks, providing a means of understanding complex brain function and how it may change. This ERGM based approach may be preferable to using correlation networks as it explicitly accounts for patterns in brain topology. However, it is possible that due to individual-level variation, there may be heterogeneous differences in tie formation across and within brain networks. Illustrated in Figure 11, examining this network offers an opportunity to demonstrate that heterogeneity in tie formation occurs in both biological and social life.

To explore the role of heterogeneity in a brain's topological organization, we fit a FERGM on the brain network of Simpson et al.'s (2012) Subject 12. By accounting for this heterogeneity, the tie prediction accuracy of the model increases by 3.1%, as presented in Figure 13. For these fits, we include the same terms as Simpson et al. (2012), Edges, GWNSP ($\alpha = 0.75$), and GWESP ($\alpha = 0.75$). When fitting the ERGM, the inferences drawn in the original work do not differ from those indicated in Figure 12. However, when using the FERGM the inferences drawn from these results change for two of three variables. Both the Edges and GWESP terms are found to be significant in the ERGM, only to have null effects in the FERGM. Only GWNSP has a consistent effect that does not differ across models. The

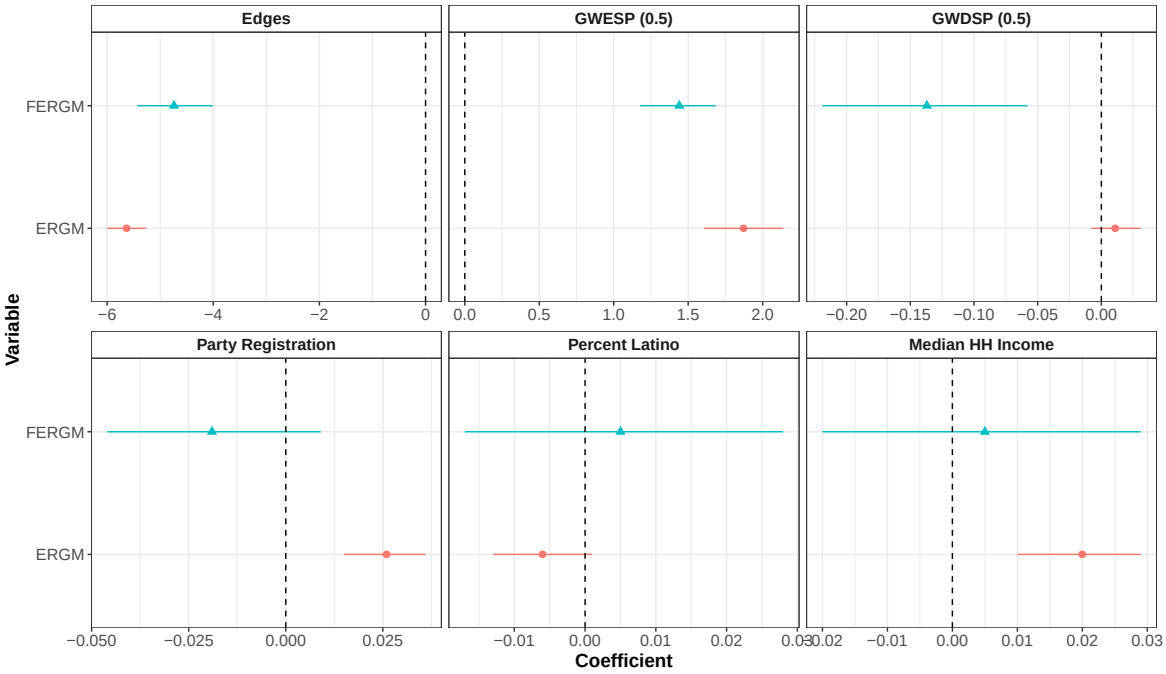

Figure 10: Coefficient Plot, Land-use Planning Network. 95% confidence intervals presented.

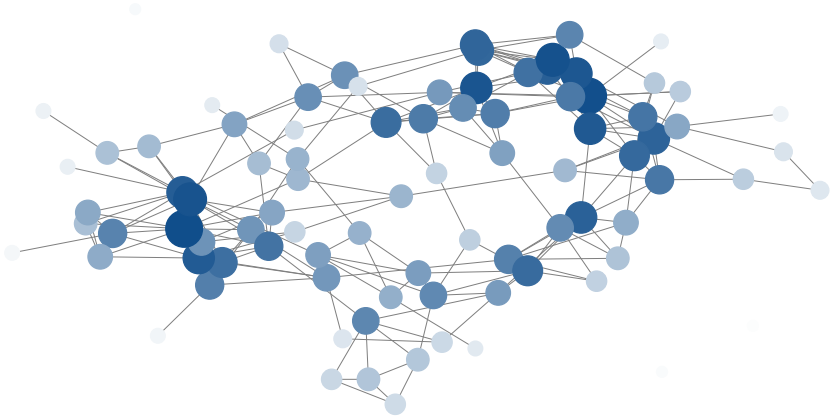

Figure 11: Subject 12 Brain Network. Nodes are sized and colored according to increasing degree.

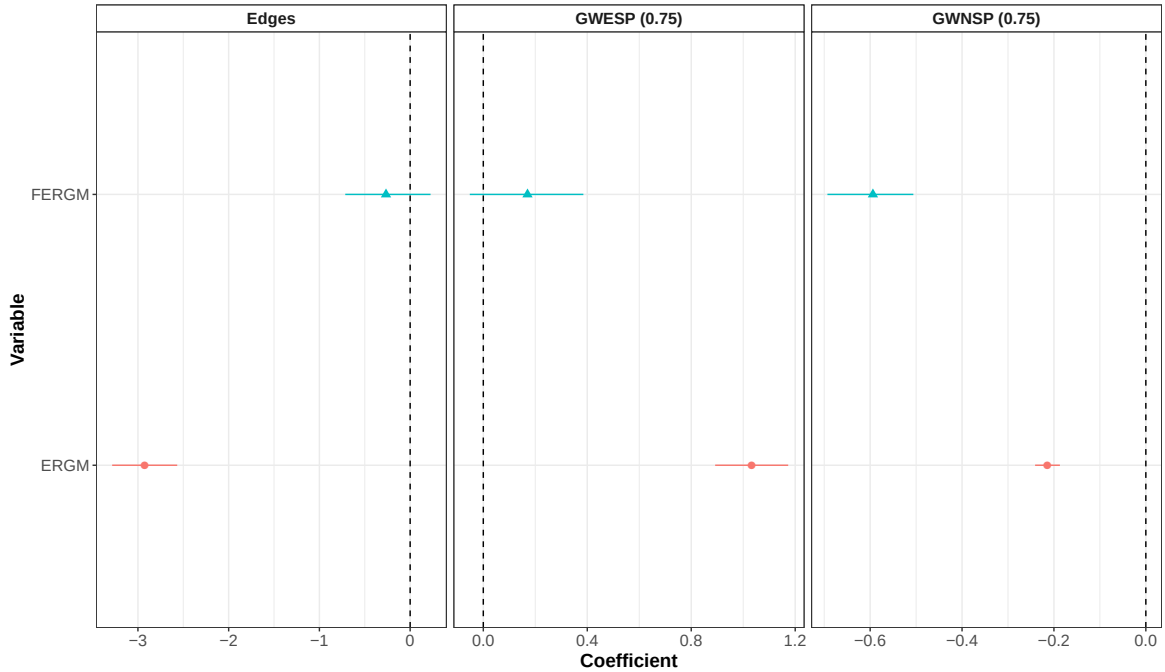

Figure 12: Coefficient Plot, Subject 12 Brain Network. 95% confidence intervals presented.

results presented here illustrate how nodal heterogeneity may influence the inferences drawn even ewhen analyzing biological networks, in this case brain networks. While it is unclear what this heterogeneity means in substantive terms, it may have important implications for predicting brain topology.

3.2 General Findings

While the prior sections have discussed the inferential differences between the ERGM and FERGM within particular cases, here we emphasize the general patterns uncovered. In aggregate, we find that inferences drawn from FERGMs differ from those drawn from ERGMs in important and theoretically intuitive ways. The FERGM also predicts, on average, much better than the ERGM out-of-sample. We will discuss each of these aggregate findings in the following pages.

3.2.1 Theoretical and Empirical Implications for Substantive Networks

Using the fits from the previously discussed applications, we examine the change in the substantive conclusions drawn from the ERGM and the FERGM for each of six networks used. To do so, we look for the covariates that have the same substantive conclusion (interpreted as common sign and significance for both the ERGM and FERGM). Not only does the FERGM demonstrate superior model fit, but, overall, the substantive implications one may draw from these applications changes dramatically as well. Table 1 demonstrates that by accounting for unit frailty there are often *dramatic* changes in the substantive implications

	Same Effect	Different Effect
Brain Networks	33.3%	66.6%
Militarized Interstate Disputes (MIDs)	25%	75%
Mesa High	83.3%	16.6%
Regional Planning	50%	50%
Florentine Marriage	57.1%	42.9%
Military Alliances	57.1%	42.9%

Table 1: Change in Substantive Interpretation of Covariate Effects. Percentages calculated by taking the percentage of total covariates in the model that either had the same sign and significance between the ERGM and FERGM (same effect) or changed in sign and significance between them (different effect).

drawn from the model. This variation is not just among exogenous nodal or edge covariates, but also occurs among endogenous terms.

3.2.2 Predictive Performance for Substantive Networks

To better understand the improvements made through accounting for unit-level heterogeneity, we move to an out-of-sample predictive exercise mirroring that introduced by Box-Steffensmeier et al. (2017). For the estimated ERGM, 500 networks were simulated and then the percentage of correctly predicted ties based upon the estimated parameters were calculated. Then, for the FERGM, we calculated the predicted probability of a tie for each of the 500 posterior draws from the parameter estimates. These probabilities were then used to simulate ties according to a Bernoulli distribution for each dyad in the network. We then compared the FERGM predicted ties to the observed ties and calculated the percent correctly predicted to get a quantity comparable for the ERGM simulations. Across all 500 simulations we are left with two quantities: the mean percentage of ties correctly predicted by the ERGM and the mean percentage of ties correctly predicted by the FERGM. These can be compared to assess which has superior model fit.

The results of this exercise are presented in Figure 13. The percent improvement is calculated as the difference in the mean accuracy of tie prediction for the FERGM relative to the ERGM across 500 simulations. In each application, except the Mesa High Network, there is a fairly large increase in the percent of ties accurately predicted by using the FERGM. The ERGM outperforms the FERGM by only .06% in the Mesa High Network, which, as we discussed above, makes sense as the Mesa High Network is itself simulated from an ERGM fit on an AddHealth network (Hunter et al., 2008). Across the other networks, the degree of improvement in predictive accuracy ranges from minor in the application of the FERGM to the Military Alliance Network (1.4%) to fairly significant in the Florentine Marriage Network ($> 10\%$).

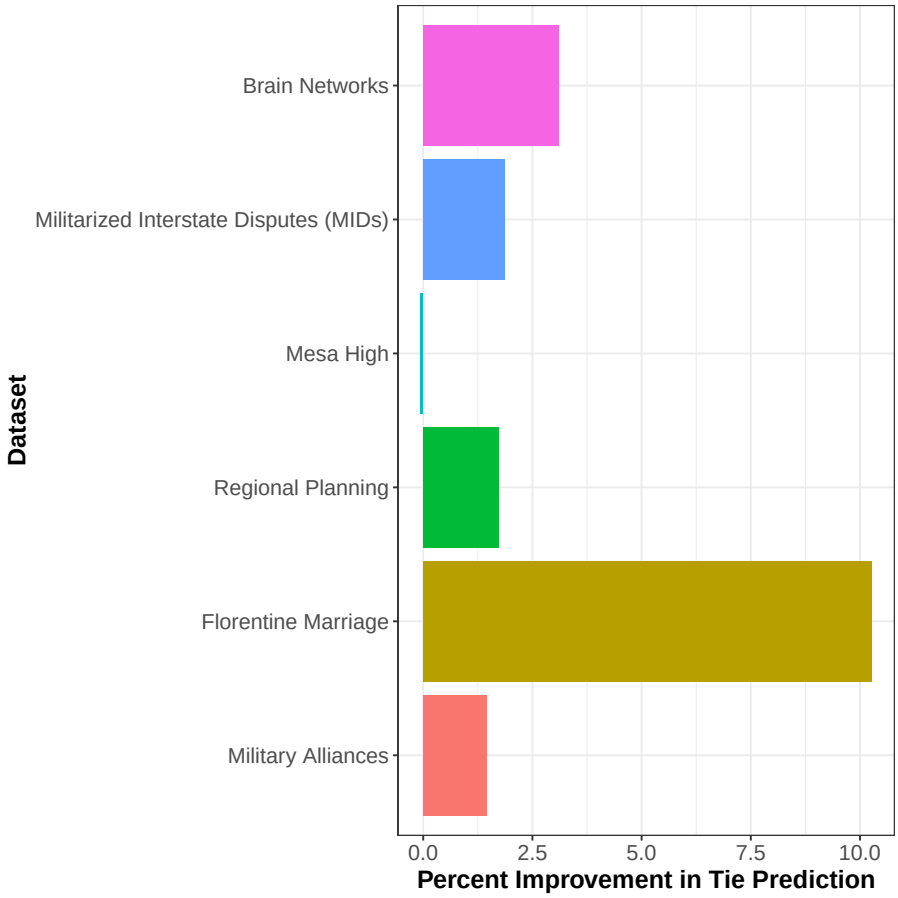

Figure 13: FERGM v. ERGM Performance. Improvement in model performance in terms of the difference in mean percentage of ties correctly predicted in 500 simulated networks

4 Concluding Thoughts

With the increasing prevalence of ERGMs, scholars must be increasingly aware that unobserved unit or group-level heterogeneity may pose a set of inferential challenges in applying these models. Often times, this heterogeneity and the omission of important variables may lead to a degenerate model specification. In other cases, models may estimate but the effects identified may be biased and confounded by heterogeneity. With great computational power comes great analytic responsibility—scholars must be aware that even if ERGMs can be estimated, the effects identified may be biased by the presence of unobserved confounding.

Take the case of military alliance formation, diplomatic historians have long acknowledged that actors have various incentives for signing alliance treaties (Taylor, 1954; Morrow, 1991; Schroeder, 1996). Measuring these incentives directly has long eluded scholars; while capturing heterogeneity in alliance formation through frailty-based linear models may be possible, it comes at the cost of ignoring the network dependencies that are known to influence alliance formation (Cranmer et al., 2012a,b). The FERGM offers an opportunity to capture both of these factors simultaneously. Doing so reveals evidence that the long-dominant paradigm for considering alliances, that states form them to counter external threats, may be impacted by omitted variable bias and unobserved hterogeneity. Similar patterns hold when examining Regional Planning wherein partisan homophily may not matter as much as previously thought, or in the application to militarized disputes where the democratic peace, an empirical regularity approaching law-like status, is not uncovered. The applications presented in this piece—while ranging markedly in academic discipline—all show significant evidence for how unobserved heterogeneity manifests itself and has consequences for our inferences.

As our six applications have illustrated, this bias may create non-trivial inferential errors and undermine the accuracy of model forecasts. Fortunately, a solution exists that works fairly well in these cases and should do so beyond them where unobserved heterogeneity is likely. The FERGM offers an opportunity to overcome the problems of model degeneracy, inferential errors, and poor model fit that stem from sender or receiver unobserved heterogeneity.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., Xing, E. P., 2008. Mixed membership stochastic blockmodels. Journal of Machine Learning Research 9 (Sep), 1981–2014.

Anderson, C. J., Wasserman, S., Crouch, B., 1999. A p* primer: Logit models for social networks. Social networks 21 (1), 37–66.

Box-Steffensmeier, J., Christenson, D., Morgan, J., 2017. Modeling unobserved heterogeneity in social networks with the frailty exponential random graph model. Political Analysis 26 (1), 3–19.

Box-Steffensmeier, J. M., Campbell, B. W., Christenson, D. P., Navabi, Z., 2018. Role analysis using the ego-ergm: A look at environmental interest group coalitions. Social Networks 52, 213–227.

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

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

Box-Steffensmeier, J. M., De Boef, S., 2006. Repeated events survival models: the conditional frailty model. Statistics in medicine 25 (20), 3518–3533.

Box-Steffensmeier, J. M., De Boef, S., Joyce, K. A., 2007. Event dependence and heterogeneity in duration models: The conditional frailty model. Political Analysis 15 (3), 237–256.

Box-Steffensmeier, J. M., Reiter, D., Zorn, C., 2003a. Nonproportional hazards and event history analysis in international relations. Journal of Conflict Resolution 47 (1), 33–53.

Box-Steffensmeier, J. M., Reiter, D., Zorn, C. J., 2003b. Temporal dynamics and heterogeneity in the quantitative study of international conflict. Economic interdependence and international conflict: New perspectives on an enduring debate, 273–289.

Bridge, R., Bullen, R., 2014. The Great Powers and the European States System 1814-1914. Routledge.

Campbell, B., Cranmer, S., Desmarais, B., 2018. Triangulating war: Network structure and the democratic peace. arXiv preprint arXiv:1809.04141.

Campbell, B. W., 2018a. Detecting heterogeneity and inferring latent roles in longitudinal networks. Political Analysis 26 (3), 292–311.

Campbell, B. W., 2018b. Measuring and assessing latent variation in alliance design and objectives. In: annual meeting of the SPSA, New Orleans, LA.

Carr, E. H., 1946. The twenty years' crisis, 1919-1939: an introduction to the study of international relations. MacMillan and Co. Ltd.

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

Cranmer, S. J., Desmarais, B. A., Kirkland, J. H., 2012a. Toward a network theory of alliance formation. International Interactions 38 (3), 295–324.

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

Cranmer, S. J., Leifeld, P., McClurg, S. D., Rolfe, M., 2016. Navigating the range of statistical tools for inferential network analysis. American Journal of Political Science.

Dafoe, A., Oneal, J. R., Russett, B., 2013. The democratic peace: Weighing the evidence and cautious inference. International Studies Quarterly 57 (1), 201–214.

Desmarais, B. A., Cranmer, S. J., 2012. Statistical inference for valued-edge networks: the generalized exponential random graph model. PloS one 7 (1), e30136.

Fellows, I., Handcock, M. S., 2012. Exponential-family random network models. arXiv preprint arXiv:1208.0121.

Fordham, B. O., 2010. Trade and asymmetric alliances. Journal of Peace Research 47 (6), 685–696.

Fosdick, B. K., Hoff, P. D., 2015. Testing and modeling dependencies between a network and nodal attributes. Journal of the American Statistical Association 110 (511), 1047–1056.

Fowler, J. H., Heaney, M. T., Nickerson, D. W., Padgett, J. F., Sinclair, B., 2011. Causality in political networks. American Politics Research 39 (2), 437–480.

Frank, O., Strauss, D., 1986. Markov graphs. Journal of the american Statistical association 81 (395), 832–842.

Gartzke, E., 2007. The capitalist peace. American journal of political science 51 (1), 166–191.

Gerber, E. R., Henry, A. D., Lubell, M., 2013. Political homophily and collaboration in regional planning networks. American Journal of Political Science 57 (3), 598–610.

Gormley, I. C., Murphy, T. B., 2010. A mixture of experts latent position cluster model for social network data. Statistical methodology 7 (3), 385–405.

Handcock, M. S., Robins, G., Snijders, T. A., Moody, J., Besag, J., 2003. Assessing degeneracy in statistical models of social networks. Tech. rep., Citeseer.

Hegre, H., 2014. Democracy and armed conflict. Journal of Peace Research 51 (2), 159–172.

Hoff, P., 2008. Modeling homophily and stochastic equivalence in symmetric relational data. In: Advances in neural information processing systems. pp. 657–664.

Hoff, P. D., 2005. Bilinear mixed-effects models for dyadic data. Journal of the american Statistical association 100 (469), 286–295.

Hoff, P. D., 2009. Multiplicative latent factor models for description and prediction of social networks. Computational and mathematical organization theory 15 (4), 261.

Holland, P. W., Laskey, K. B., Leinhardt, S., 1983. Stochastic blockmodels: First steps. Social networks 5 (2), 109–137.

Holland, P. W., Leinhardt, S., 1981. An exponential family of probability distributions for directed graphs. Journal of the american Statistical association 76 (373), 33–50.

Huckfeldt, R. R., Sprague, J., 1995. Citizens, politics and social communication: Information and influence in an election campaign. Cambridge University Press.

Hunter, D. R., Goodreau, S. M., Handcock, M. S., 2008. Goodness of fit of social network models. Journal of the American Statistical Association 103 (481), 248–258.

Krivitsky, P. N., Handcock, M. S., Raftery, A. E., Hoff, P. D., 2009. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. Social networks 31 (3), 204–213.

Lake, D. A., 2009. Hierarchy in international relations. Cornell University Press.

Levy, J. M., 2009. Contemporary urban planning. Routledge.

Levy, J. S., 1998. The causes of war and the conditions of peace. Annual Review of Political Science 1 (1), 139–165.

Long, A. G., 2003. Defense pacts and international trade. Journal of Peace Research 40 (5), 537–552.

Maoz, Z., Russett, B., 1993. Normative and structural causes of democratic peace, 1946–1986. American Political Science Review 87 (3), 624–638.

Margetts, H., 2017. Political behaviour and the acoustics of social media. Nature Human Behaviour 1, 0086.

Marshall, M. G., Jaggers, K., Gurr, T. R., 2002. Polity IV project. Center for International Development and Conflict Management at the University of Maryland College Park.

McPherson, M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. Annual review of sociology 27 (1), 415–444.

Minhas, S., Hoff, P. D., Ward, M. D., 2016. Inferential approaches for network analyses: Amen for latent factor models. arXiv preprint arXiv:1611.00460.

Morgan, J. W., Campbell, B. W., Christenson, D. P., Box-Steffensmeier, J. M., 2018. fergm: Estimation and Fit Assessment of Frailty Exponential Random Graph Models. R package version 1.1.1.
URL `CRAN.R-project.org/package=fergm`

Morgenthau, H., 1948. Politics Among Nations: The struggle for power and peace. Nova York, Alfred Kopf.

Morrow, J. D., 1991. Alliances and asymmetry: An alternative to the capability aggregation model of alliances. American Journal of Political Science, 904–933.

Padgett, J. F., Ansell, C. K., 1993. Robust action and the rise of the medici, 1400-1434. American journal of sociology 98 (6), 1259–1319.

Pevehouse, J. C., 2002. Democracy from the outside-in? international organizations and democratization. International organization 56 (03), 515–549.

Powers, K., 2004. Regional trade agreements as military alliances. International Interactions 30 (4), 373–395.

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., et al., 1997. Protecting adolescents from harm: findings from the national longitudinal study on adolescent health. Jama 278 (10), 823–832.

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

Robins, G., Snijders, T., Wang, P., Handcock, M., Pattison, P., 2007b. Recent developments in exponential random graph (p*) models for social networks. Social networks 29 (2), 192–215.

Russett, B., 1994. Grasping the democratic peace: Principles for a post-Cold War world. Princeton university press.

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

Schroeder, P. W., 1996. The transformation of European politics, 1763-1848. Oxford University Press.

Simpson, S. L., Moussa, M. N., Laurienti, P. J., 2012. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. Neuroimage 60 (2), 1117–1126.

Sinclair, B., 2012. The social citizen: Peer networks and political behavior. University of Chicago Press.

Snijders, T. A., 2002. Markov chain monte carlo estimation of exponential random graph models. Journal of Social Structure 3 (2), 1–40.

Snijders, T. A., Nowicki, K., 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. Journal of classification 14 (1), 75–100.

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

Snyder, G. H., 1997. Alliance politics. Cornell University Press.

Sprecher, C., Krause, V., Long, A. G., Leeds, B. A., 2006. Trading for security: Military alliances and economic agreements. Journal of Peace Research 43 (4), 433–451.

Stillman, P. E., Wilson, J. D., Denny, M. J., Desmarais, B. A., Bhamidi, S., Cranmer, S. J., Lu, Z.-L., 2017. Statistical modeling of the default mode brain network reveals a segregated highway structure. Scientific reports 7 (1), 11694.

Taylor, A. J. P., 1954. The struggle for mastery in Europe, 1848-1918. Oxford University Press.

Thiemichen, S., Friel, N., Caimo, A., Kauermann, G., 2016. Bayesian exponential random graph models with nodal random effects. Social Networks 46, 11–28.

Van Duijn, M. A., Gile, K. J., Handcock, M. S., 2009. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Social Networks 31 (1), 52–62.

Victor, J. N., Montgomery, A. H., Lubell, M., 2016. The Oxford Handbook of Political Networks. Oxford University Press.

Walt, S. M., 1990. The origins of alliance. Cornell University Press.

Waltz, K., 1979. Theory of international relations. Reading, Mass.: Addison-Webley, 111–114.

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

White, A., Murphy, T. B., 2016. Mixed-membership of experts stochastic blockmodel. Network Science 4 (1), 48–80.

Wilson, J. D., Denny, M. J., Bhamidi, S., Cranmer, S. J., Desmarais, B. A., 2017. Stochastic weighted graphs: Flexible model specification and simulation. Social Networks 49, 37–47.

Yan, T., Jiang, B., Fienberg, S. E., Leng, C., 2018. Statistical inference in a directed network model with covariates. Journal of the American Statistical Association (just-accepted), 1–33.