Inferential Network Analysis with Exponential Random Graph Models

Skyler J. Cranmer   Bruce A. Desmarais

April 15, 2010

Abstract

Methods for descriptive network analysis have reached statistical maturity and general acceptance across the social sciences in recent years. However, methods for statistical inference with network data remain fledgling by comparison. We evaluate and extend a new model for inference with network data, the Exponential Random Graph Model (ERGM), that simultaneously allows both inference on covariates and for arbitrarily complex network structures to be modeled. Our contributions are four fold: beyond introducing the ERGM and discussing its limitations, we extend the model to allow for the analysis of longitudinally observed networks, develop guidelines for analyzing highly dense and highly sparse networks, and show through applications that network-based inference can improve our understanding of political phenomena.

1   Introduction

Over the past several decades, political scientists have increasingly looked to relational data generally and network data specifically to address important substantive questions. However, statistical tools for the analysis of network data have not kept pace with the breadth and complexity of the substantive questions being asked of such data. Because methods which can accommodate complex interdependencies have not been readily available, it has become commonplace in many subfields of political science to use traditional statistical tools, such as generalized linear models, on network data. Such an approach necessarily produces biased results because the structure of the data generating process cannot be modeled with tools that assume conditional independence of observations. Worse still, the unmodeled effects of network structure will be falsely attributed to the covariates; thus perilously raising the probability of faulty inference. While there are several useful approaches for analyzing relational data, we address an area of particular concern for political scientists: how to make unbiased inferences about covariate effects, and even account for arbitrarily complex interdependencies, when analyzing political networks. Our aim here is to introduce and improve upon a model for inference on network data called the Exponential Random Graph Model (ERGM) in such a way as to make it widely useful to political scientists. Our four contributions are that we introduce the ERGM and discuss several applied problems that currently limit its utility to substantive researchers, develop guidelines for analyzing highly dense and sparse networks with the ERGM, offer a solution to the current inability of ERGMs to model longitudinally observed networks, and show by means of example that network inference using ERGMs can not only avoid faulty inference on covariates, but can provide new insight to substantive problems.

The exponential random graph model (ERMG), has evolved out of a series of efforts by statisticians and sociologists to address the same limitations of classical statistical models

which have frustrated political scientists using network data. The theoretical foundations for the ERGM were originally laid by Besag (1975) who proved that there exists a class of probability distributions that are consistent with the Markovian property that the value of any one location is dependent only on the values of its contiguous neighbors.[1] Building on Besag's (1975) work, Holland and Leinhardt (1981) derived the exponential family of distributions for networks. That same year, Fienberg and Wasserman (1981) put the exponential family of models into the context of log-linear models and examined issues with the likelihood estimation of such models. This class of models and the accompanying estimation techniques were developed further by Frank and Strauss (1986) and Strauss and Ikeda (1990) but it was not until Wasserman and Pattison (1996) that the fully generalized ERGM was first derived.

In its simplest form, the ERGM provides the same general sort of inference to which political science is accustomed: an outcome variable is predicted by several independent variables, only no assumptions of independence are necessary. Imagine the outcome variable of interest is alliance ties between two countries. Rather than assume that each state decides its alliance portfolio independent of all other decisions made by all other states and that decisions within a state's alliance portfolio are independent, the ERGM considers the network of alliance relationships, models the probability of observing that network over the other networks we could have observed, and allows the analyst to draw conditional inference on any dyadic relationship within the network. In other words, the ERGM allows researchers to perform an analysis loosely akin to logistic regression on network data, but in addition to modeling the effects of covariates, they can take as much network dependence into account as they see theoretically or empirically fit. This powerful and flexible tool is straightforward to understand – though it requires us to think in a fundamentally different way about the

---

[1]Specifically, Besag (1975) proved the Hammersley-Clifford Theorem which shows this in the context of spatial data, but that theorem is naturally generalized to networks.

structure of the data – and since it is already well supported by software, it is also quite easy to use. The ERGM is, as we will show, widely applicable to relational analysis in political science and is remarkably flexible in its ability to model complex networks.

Though the general framework for ERGMs has already been developed, there remain a number of applied problems which currently limit the utility of ERGMs for political scientists, chief amongst which are their inability to model longitudinally observed networks and coding challenges which can alter the substantive meanings of ERGM results. We propose solutions to these problems respectively and demonstrate their efficacy with political science data from American politics and international relations.

We are not the first to have noticed that treating relational data as independent is a problematic practice. Qualitative theorists in various subfields have discussed interdependence of relationships for some time and descriptive techniques for network data are now reasonably well known. While the discipline as a whole has yet to move away from regression on network data, many recent works have striven to adapt regression modeling to the network framework (Baybeck and Huckfeldt 2002; Crescenzi 2007; Fowler 2006; Franzese and Hays 2006; Lazer 2005; Maoz et al. 2006; Maoz 2009; Schneider et al. 2003; Scholz and Wang 2006; Ward and Hoff 2007; Ahlquist and Ward 2009).

Most efforts to extend the regression framework to the network context have involved the inclusion of network statistics as covariates in regression analyses. As one example, Maoz (2009) models the dyad-level incidence of conflict, as well as the annual aggregate count of militarized interstate disputes, as functions of characteristics of the international trade and alliance networks. Also, Maoz et al. (2006) compute structural equivalence scores for each dyad in the network and include those estimates as covariates in a logistic regression of international conflict. This approach to modeling networks, as applied to both international relations and American politics, is pivotal in its recognition of the importance of network

characteristics to network outcomes, but because it applies standard regression to account for this interdependence, it does not do away with the independence assumption.

Some promising approaches have also been developed out of spatial statistical methods (Franzese and Hays 2006; Hoff, Raftery and Handcock 2002; Hoff and Ward 2004). Perhaps the most visible approach based on spatial statistics is the latent space network model originally proposed by Hoff, Raftery and Handcock (2002). The latent space model takes account of network dependencies by mapping each node to a position in k-dimensional Euclidean space. The positions of the nodes in this latent space are then linked back to the network through the assumption that the closer any two nodes are to one another in the latent space, the more likely they are to share an edge. A logistic regression is then estimated where, in addition to covariates, the Euclidean distance between each node and each other node is controlled for. Since its development, the latent space approach has since been applied to international conflict (Ward, Siverson and Cao 2007) and rather extensively to trade, job protection and international political economy generally (Cao, Prakash and Ward 2007; Ward and Hoff 2007; Ahlquist and Ward 2009).

It is not our intention here to pit the ERGM against these alternative approaches and discover the "best" method for modeling complex interdependencies; the "best" approach will vary by application based on substance and the quest for a universally best method is quixotic. Instead, we seek to introduce and extend the ERGM as an approach which may be useful in many contexts.

We begin by deriving the Exponential Random Graph Model and discussing the ways in which it requires us to think differently about network data, how it solves the dyadic dependence problem, and some limitations of the model. We then propose a novel solution to a problem often faced by political scientists, but not addressed systematically in other literatures: what to do when the network of interest is observed over time. We demonstrate

the ERGM and our proposal for analyzing longitudinal networks using two examples chosen for both salience and exposition: cosponsorship networks in the US Congress and conflict networks in the international system. We feel that these are good expository applications because they bring up problems with coding rules for the network ties which can fundamentally change the meaning of one's analyses; the cosponsorship network is quite dense (highly interconnected), requiring the introduction of what we call network "thinning," and the conflict network is rather sparse, suggesting the possibility of "thickening" the network. Through these examples, we will also show that using ERGMs to model these networks not only rectifies bias problems often encountered with the erroneous application of classical methods, but can also provide deeper substantive insight and suggest new theoretical approaches.

2 Exponential Random Graph Models (ERGMs)

A useful statistical model for networks must meet two major criteria. First, it must avoid the problematic assumption that relational observations – observations on the dyad – are conditionally independent from one another. Second, it must in some way allow for network structures to be modeled. Ideally, this latter characteristic would be highly flexible, transparent and intuitive. We feel that the major strength of the ERGM approach to inferential statistics in networks is that it satisfies both of these criteria.

In order to construct a model without assuming dyadic independence, we need to think in a somewhat different fashion about the data generating process. We can construct a likelihood function for any vector of interest Y without assuming independence by considering Y to be a single observation from a multivariate probability distribution. So instead of thinking about Y as a series of values drawn from a univariate distribution (as is the case for standard regression models), we think of it as a single draw from a multivariate distribution

where many other draws (many other realizations of the network) are possible. In other words, there is a multivariate distribution from which the realized network is drawn. If Y is a single realization from a multivariate distribution, we no longer have to assume independence among the values of Y in any way. We will still be able to draw inferences about particular dyads, so the unit of analysis need not change, but the multivariate treatment of Y is the conceptual leap necessary to avoid the problems associated with independence violations. The modeling challenge is then shifted from constructing a model of a univariate outcome, the parameters of which are estimated using many observations, to constructing a model of the multivariate distribution from which a single network is drawn and whose parameters are estimated with a single observation. ERGMs provide a general approach to constructing these probability distributions.

To formalize the intuition given above, we follow and build upon the derivation of the general class of ERGMs presented by Park and Newman (2004). Suppose there are k statistics $\Gamma_i, i = 1, \ldots, k$ that can be computed on the network (or graph) G, which the analyst believes affect the likelihood of observing G. These statistics can include measures of interconnectedness, reciprocity in directed networks and node or dyad-level covariate values. The Γ parameter is how covariates and structural elements of the network are modeled at the same time. While a detailed review of possible network parameters to include is beyond the scope of this discussion, interested readers are referred to Handcock et al. (2008) and Snijders et al. (2006) for such a treatment. The only condition we require is that no subset of Γ be linearly dependent on another; in other words, there can be no perfect collinearity. Given N, the (fixed) number of nodes in G, there are M possible sets of edges on N. This means that the observed network G is one of M possible networks with the same number of nodes that could have been observed. This illustrates the assumption that the observed network G arises stochastically from the support G_M.

The general ERGM is derived using a moment-matching method. We start by making

one of only two assumptions necessary for the ERGM: we assume that each of the network statistics calculated on graph G are the expected values of those statistics across other possible graphs:

$$E\left[\Gamma_i\right] = \Gamma_i, \tag{1}$$

where Γ_i is any network statistic. In other words, we assume that the expected value of the network statistics are their values as computed from G. While this may seem like a strong assumption, one must keep in mind the fact that, in many cases, we will only observe a single realization of the network (i.e. there is only one realized Supreme Court citation network) and so the observed value of the statistic Γ_i is actually our best indication of its expected value $E[\Gamma_i]$. This assumption is necessary because it establishes an identifying condition on the probabilities of the networks in G_M:

$$E\left[\Gamma_m\right] = \sum_{m=1}^{M} P(G_m)\Gamma_m. \tag{2}$$

Yet the distribution over G_M is not fully identified at this point. In addition to the restriction made in equation 1, we must also assume that only the statistics included in Γ influence the probability that graph m is observed, $P(G_m)$. This is accomplished by maximizing the Gibbs entropy, $S = -\sum_{m=1}^{M} P(G_m) \ln P(G_m)$, on the discrete distribution of graphs in G_M subject to the condition in equation 2. This identifies the discrete probability mass function over the networks in G_M to be as close as possible to a discrete uniform distribution while satisfying the constraint in equation 2. The maximization is performed using the Lagrange multiplier method.

As as result of this optimization, we can recover an elegant formula reflecting the rela-

tionship between the probability of a graph i and the network statistics in Γ,

$$P(G_i) = \frac{\exp(-\sum_{j=1}^{k} \Gamma_{ij}\theta_j)}{\sum_{m=1}^{M} \exp(-\sum_{j=1}^{k} \Gamma_{mj}\theta_j)} \tag{3}$$

where θ is the vector of k parameters that describe the dependence of $P(G_i)$ on the network statistics in Γ (see the Appendix for a full derivation). We can then choose between maximum likelihood or Bayesian methods to estimate the θ's.

The power of the ERGM approach lies in the fact that we have derived the parametric form of the distribution of interest from very general conditions. The only assumptions we have made are (1) that we observe the expected values of Γ and (2) that we have identified the factors that influence the probability of observing any given graph. ERGMs thus provide a method capable of estimating standard covariate effects as well as the effects of other network properties without having to make any assumptions about independence.

2.1 Interpreting ERGMs

As is the case with standard regression models, there are several ways in which one can interpret the estimated parameters of an ERGM. We focus on two levels of interpretation: the network-level and the dyad-level.

At the network-level, it will often be of interest to characterize the expected distribution of a network statistic. For instance, the distribution of the number of edges in a conflict network is important if one is interested in the expected frequency of interstate war. The challenge is that the parametric form of the distribution of a function of the graph is not directly recoverable. This is because parameters of the ERGM describe the relationship between the properties of a network of interest G and the probability of observing G. The parameters are thus not directly linked to a moment of the distribution of the outcome

variable (the network) as they are in a likelihood model. Since ERGMs produce unique and complex distributions of networks that involve every possible realization of edge combinations with the same number of nodes, closed form manipulation of these distributions is usually impossible due to the fact that M in the denominator of equation 3 is often a large number.

In the absence of closed form solutions, simulation techniques provide a straightforward means of approximating the distribution of a given network statistic. Morris, Handcock and Hunter (2008) develop an algorithm to simulate networks from the distributions implied by ERGM parameters. Thus if one is interested in the distribution of some statistic $\Gamma(\cdot)$ defined on the network, an approximation to this distribution is easily obtainable by simulating a large number of networks based on the ERGM parameters and computing $\Gamma_i(\cdot)$ for each network (Morris, Handcock and Hunter 2008; Handcock et al. 2008). The larger the number of simulations, the more accurate the approximation. Once the distribution is obtained, approximations to the characteristics of the distribution of the network statistic can be obtained with summary statistics (e.g. what is the variance in the number of wars?).

The dyadic level of interpretation will be the most familiar to political scientists as it closely resembles the results from a logistic regression. At the dyad-level, the quantity of primary interest in most networks is the probability of the existence of particular edges (e.g. how likely is it that there is war between the U.S. and Canada?). The probability of an edge between nodes i and j is written,

$$Pr(Y_{ij} = 1 | n, Y_{ij}^c) = \text{logit}^{-1} \left(\sum_{k=1}^{K} \theta_k \delta(\Gamma_{yk}) \right), \tag{4}$$

where n is the number of nodes, Y_{ij}^c denotes all dyads other than Y_{ij}, K is the number of network statistics in the ERGM, and $\delta(\Gamma_{yk})$ is the amount by which Γ_{yk} changes when Y_{ij} is toggled from 0 to 1 (Goodreau, Kitts and Morris 2009). Using equation 4 we can produce predicted probabilities for the edges in the network. Notice that the predicted probabilities

cannot be produced without including Y_{ij}^c; this is a major departure from the computation of predicted probabilities in a logit or probit model. The inclusion of Y_{ij}^c is necessary because in ERGMs, $Pr(Y_{ij})$ is deterministically dependent on the dyad-wise outcomes of *every* other dyad. Its differences aside, equation 4 gets as close as possible to the sort of interpretation which is commonplace with logit models and does not represent a major departure from the sort of interpretations the field is accustomed to.

When covariates are included in a standard likelihood model, variance in the outcome or dependent variable is regressed on variance in the independent variables. But in the case of an ERGM, there is no variance in the dependent variable (the network, recall, is a single observation). In order to estimate the influence of covariates, ERGMs exploit the hypothetical variation over networks contained in the denominator of equation 3. For each network computed in equation 3, the sum of the covariate values for each edge are added up (noting that the existing edge set changes for each network). The parameter corresponding to a specific covariate then tells us if the probability of observing a graph increases or decreases as the the sum of the covariate values for all connected dyads increases. A positive parameter value indicates that graphs where dyads that are connected have high covariate values are highly likely to be observed relative to graphs where connected dyads have low covariate values. The interpretation of the covariate parameters is similar to the regression case. A positive valued parameter means it is likely that high covariate values correspond to connected dyads and a negative valued parameter means it is likely that low covariate values correspond to connected dyads.

2.2   Limitations of ERGMs

Though the ERGM is a powerful and flexible model, it suffers from a number of limitations which may hamper its utility to researchers. The major limitations of the ERGM are:

degeneracy problems, an inability to model networks over time, a sensitivity to missing data, a limited ability to model networks with non-binary edges, and (related) the fact that ideal coding rules for edges are often not obvious. We discuss most of these limitations here; in some cases, we argue that the drawbacks are not as limiting as they might seem. We reserve a discussion of the binary edges/coding rule problems and the longitudinal analysis problem for sections 4.1 and 4.2 respectively as we propose specific solutions to those problems.

Degeneracy is an estimation problem associated with models that fit the data poorly. Essentially, degeneracy results from the specification of a model which is so unlikely to have generated the network, that the ERGM cannot be computed. Degeneracy occurs when the model lumps all or most of its probability mass on just one or a few possible graphs. In most cases of degeneracy, disproportionate probability mass is placed either on the complete (fully connected) or empty (entirely unconnected) networks (Handcock 2003). Once the ERGM has been specified, it is estimated using Markov chain Monte Carlo, but degenerate or near degenerate models cause problems in the estimation process (see Snijders (2002) for a detailed description). If the model is not well specified, the estimated probabilities will walk the chain to an extreme graph of all or no edges, where it will stay, and the model is said to be degenerate.

While degeneracy can be frustrating, it is not necessarily a major hinderance to applied research. Because models which are highly unlikely to have generated the network will be degenerate, degeneracy is, in some sense, a statement about model fit. This certainly should not be taken to mean that a given model fits well simply because it is not degenerate, but any model which is degenerate is sure not to fit well. Since including parameters or variables which contribute no meaningful information to the prediction of the outcome network will usually cause degeneracy, one must think carefully about the components of the models to be estimated and sometimes specification searches may be necessary. Quite unlike standard regression models, ERGMs with a set of "standard controls" which do nothing to predict

the outcome will cause degeneracy problems and thus should not be specified. The major "weakness" of ERGMs then is the fact that one's model must actually fit the data reasonably well; not much of a limitation in our opinion.

It is natural to weigh the occasional frustrations researchers may experience with model degeneracy against the convenience of the traditional regression framework. The fundamental issue here is whether one should prefer a violation of the independence assumption with a computable MLE to a model which produces an error when it is grossly misspecified. We feel that degeneracy actually helps the ERGM analyst in this case in that it will return a degeneracy error before it returns results for uninformative models. In other words, logistic regression on network data makes it quick and easy to find an answer – including the wrong answer – whereas researchers using ERGMs may occasionally find that they have to alter the specification of their model in order to specify a model which fits the data well enough to be estimable.

Another potential shortcoming of the ERGM approach to modeling networks is the assumption that the appropriate network statistics to include in the model are known. While this assumption is not different in nature than the assumption of "correct specification" made in any regression style model, some may claim that the difficulty of surmounting this hurdle is magnified by the fact that political scientists rarely have theoretical expectations about network structures such as closed triangles, reciprocity, transitivity, or any other of the myriad of network statistics that could be included in an ERGM specification. We see this critique as being decreasingly valid. It is certainly that case that few political science theories explicitly speak to network statistics, but a history of not theorizing about network structure is not grounds to cast network structure aside as not theoretically useful. Indeed, a number of theories – such as balance of power and all manner of strategic games – imply roles for network structure. Furthermore, some recent studies have formulated theories which generate specific expectations for certain network structures. For example, Cranmer,

Desmarais and Kirkland (2010) develop a theory of synergistic gains in utility from triadic closure in alliance networks and then find triadic closure to be an important empirical driver of alliance formation. In short, we recognize that network theories of international relations are not currently ubiquitous, but posit that such theories are certainly possible and indeed the field may be moving in that direction.

Missing data can also be especially problematic in a network context. Specifically, inferences on the importance of covariate or structural parameters can be altered by missing edge values; in certain cases, small amounts of missing data can have pronounced effects. In some ways, this is not different from missing data problems encountered in the traditional regression framework: small amounts of missingness (particularly missing observations for rare events) can have pronounced effects on inference (Rubin 1976; King et al. 2001). Where the missing data challenge is magnified for ERGMs is that effective multiple imputation is more difficult to achieve than it is for models which maintain the traditional rectangular structure of the data. One option for solving this problem is to use multiple imputation on the edge-list (dyadic) representation of the network. One can even compute descriptive network statistics on the node level – such as centrality, betweenness and so on – and include those in the imputation model. This is an imperfect solution however because dependencies from certain network structures could not be taken into account. This is an open problem which recent work on inference with sampled networks is attempting to address (Handcock and Gile forthcoming), but remains problematic from an applied perspective.

3   Estimation of the ERGM

Parameterized in the form of equation 3, the ERGM has an exponential family form log-likelihood function, and as a result, the log-likelihood surface is globally concave in the parameters (van Duijn, Gile and Handcock 2009). This is an ideal setup for Newton-Raphson

type maximization of the likelihood function to find the parameters, but a challenge arises in the fact that the exact computation of the likelihood function is too computationally demanding for most networks and network statistics (Γs) of practical interest. As can be seen in the denominator of equation 3, the computation of the likelihood function requires the summation over all possible configurations of the network. For an undirected network with n nodes, this constitutes $2^{\binom{n}{2}}$ networks. A network with just 9 nodes can assume 68,719,476,736 configurations, a number that increases by a *multiplicative* factor of 1,073,741,824 if the number of nodes increases to 12. Needless to say, in order to estimate the ERGM on a network of 100-200 nodes, the likelihood function must be approximated. Two methods of approximation have found regular application in the literature – maximum pseudolikelihood (Frank and Strauss 1986) and Markov Chain Monte Carlo (MCMC) maximum likelihood (Geyer and Thompson 1992). The MCMC approach is currently the default for most software packages and is a form of maximum simulated likelihood. The method is iterative and the algorithm works as follows: in a given optimization iteration, the sum in the denominator of the likelihood function is approximated using a series of networks sampled from the distribution parameterized with those parameters that maximized the likelihood using the previous sample of networks. This iterative optimization proceeds until there is little change in the approximate likelihood function value. The covariance matrix of the parameters is then computed as the inverse of the negative Hessian of the log-likelihood function. Pseudocode for the MCMC-MLE algorithm in Geyer and Thompson (1992), which is the algorithm used in most software packages for ERGM estimation, is given in figure 1.

[Figure 1 about here.]

The pseudolikelihood technique is an analytic approximation method. The joint likelihood of the ties in the network is replaced with the product over the conditional probability of each tie given the other ties in the network. That is, for a given edge ij, the edge-wise

probability of a tie (p_{ij}) is given in equation 4. The maximum pseudolikelihood is computed by using a hill-climbing algorithm to find the vector of parameters that maximize $\log(\prod_{i=2}^{n} \prod_{j=1}^{i} p_{ij})$. This is very convenient in that it can be computed using standard logistic regression software, with the dyadic covariates and network change statistics composing the design matrix. As with the MCMC method, an approximate covariance matrix for the maximum pseudolikelihood is formed by inverting the observed information matrix (i.e. the asymptotic sampling variance of the logistic regression estimator used to compute the pseudolikelihood).

As two methods of approximation, there are pros and cons associated with each approach. The advantage of MCMC approach is that in the limit (with an infinite number of draws from the distribution of networks), it produces estimates equivalent to the MLE. The disadvantages of the technique are that the number of draws must be finite, computation can prove quite difficult for large networks and – because simulation is used – two analysts can produce different results when conducting the same study. For the pseudolikelihood approach, computation is typically fast and convergence certain, but there are few concrete results that characterize the degree of bias or loss of efficiency induced by replacing the joint likelihood with a product over conditionals. MCMC estimation has been favored in recent analyses, though not without exception (see e.g. Faust and Skvoretz (2002), Saul and Filkov (2007) and Danny, Choudhury and Bilmes (2009)). van Duijn, Gile and Handcock (2009) perform a Monte Carlo experiment to compare pseudolikelihood and MCMC methods; their findings support the recent dominance of the MCMC-MLE. For network statistics, they find that the maximum pseudolikelihood is $60 - 80\%$ as efficient as the MCMC-MLE, and for exogenous covariate effects it is 80-95% as efficient as the MLE.[2] More alarming for the task of hypothesis testing is the finding that the confidence intervals derived from the inverse

---

[2]Efficiency is measured as the inverse of the mean squared error of the parameter estimate in the Monte Carlo experiment.

of the observed fisher information are biased using pseudolikelihood, but not with MCMC. Specifically, they find that the coverage probability for the nominal 95% confidence interval is only 74.6% for a statistic that captures clustering in the network. In the first application below, we use MCMC, which is implemented in the R package `ergm` (Handcock et al. 2010), and in the second application we adapt the pseudolikelihood approach for the estimation of a single vector of ERGM parameters that covers multiple realizations of the same network.

4   Applications

We now consider two applications of ERGMs to political science data. We have chosen these applications because they allow us not only to further illustrate the methods we have described above, but also because each forces us to grapple with a different (and common) applied problem we have not addressed specifically above: what to do when the network is both dense and valued (i.e. the edges can assume more values than 0 and 1) and how to work with multiple realizations of a particular network. Lastly, we hope that the different subfields from which the applications are drawn will demonstrate the broad applicability of ERGMs in political science.

For each of the applications, we walk through the specification process and the considerations which go into it from beginning to end more as a researcher would in their office than readers are accustomed to seeing in print. This is done strictly for exposition.

4.1   Analyzing Cross-Sectional Networks:
Cosponsorship in the US Congress

Our first application is drawn from American politics and extends ideas developed by Fowler (2006). In the US Congress, legislators commonly cosponsor legislation sponsored by other

legislators.[3] The cosponsorship network is realized when one considers a legislator's cosponsorship of another legislator's legislation to be an edge between the two legislators. It is worth noting that there is some disagreement about the precise meaning of a cosponsorship tie: some have argued that it represents a signal from the cosponsor of support for the subject matter of the legislation (Kessler and Krehbiel 1996). Others have argued that it represents a show of support from one legislator to another (Fowler 2006). While it is beyond the scope of this application to sort out the precise meaning of a cosponsorship tie, we believe that we can remove the definitional problem by viewing the cosponsorship tie as a form of closeness between any two legislators; whether they are in agreement with regard to policy or politics, the cosponsorship tie is indicative of active engagement between the legislators.

We examine the cosponsorship network using data from Fowler (2006).[4] This dataset records the cosponsorship and sponsorship of every bill in the U.S. House and Senate for the 93rd through the 109th congress. The edge from legislator i to legislator j in the network is the number of times legislator i has cosponsored legislator j. Note that the network is directed and that the edge value is a non-negative integer.

When undertaking the analysis, we are quickly presented with a technical problem. Within a session of Congress, the time interval that we use to define an instance of the cosponsorship network, it is possible that any representative cosponsors another numerous times. This is problematic for two reasons. First, ERGM has only been developed to handle binary ties, so we cannot use ERGM technology to model the number of times each representative cosponsors others within a session. Many tools for network analysis are only appropriate for binary ties, and we are not the first to confront this problem – with the Congressional Cosponsorship network specifically. Both Fowler (2006) and Faust and Skvoretz

---

[3]Congressional rules dictate that there may be only one sponsor to a bill, there are no joint sponsorships.

[4]Fowler's cosponsorship data is freely available from his website at `http://jhfowler.ucsd.edu/cosponsorship.htm`.

(2002) address this problem by thresholding, and treating any positive cosponsorship count greater than zero as a cosponsorship. The second problem relates to the structure of the cosponsorship network specifically. As can be seen from the upper-left cell of Figure 2, treating any cosponsorship act as an instance of a cosponsorship relationship produces an extremely dense network. If we attempt to estimate an ERGM on this network, the parameters will produce a network where every possible tie exists and the model will be degenerate as discussed above.

[Figure 2 about here.]

In order to transform the count-based ties into binary edges as well as produce a network that permits non-degenerate ERGM estimation, we elect to create a binary network that is less dense than that in previous treatments by requiring a larger number of cosponsorships than one to indicate a cosponsorship relationship. The average number of cosponsorship ties between legislators in the 108th House is 13.76. To "thin" the network, we must pick a threshold (number of cosponsorship ties between any two legislators) above which we will code an edge and below which we will not. Obviously, when thinning a network, careful attention must be paid to the substantive meaning of the thinning as well as the sensitivity of the model ultimately produced to the thinning rule. The selection of such a threshold is necessarily somewhat arbitrary and so it is important to try several threshold values as a sensitivity analysis. When testing the sensitivity of the model, one should be looking for both significant changes in the estimates as well as problems with model degeneracy.

Further, we should point out that sensitivity of the model to the thinning thresholds chosen is not necessarily a strictly bad thing. If the results are noticeably different between, say, low and high threshold values, that tells us something quite interesting about the data. It could even suggest low threshold edges represent a fundamentally different type of edge than high threshold edges (in this case, that casual cosponsorship is fundamentally different

18

from very frequent cosponsorship) and that fundamentally different processes are at work.

Our first step in model fitting is to customize our coding (thinning) rule to the substantive application. A constant threshold may be inappropriate due to the fact that there is variance in the number of bills each legislator sponsors. Therefore censoring at a threshold is in part a census on legislator i's propensity to cosponsor legislator j *and* legislator j's frequency of sponsorship. Because we are only interested in the former, we apply a new coding rule by considering legislator i a cosponsor of legislator j if legislator i has cosponsored π percent of legislator j's sponsored legislation. Our focus is restricted here to the 108th House for the sake of brevity.

After experimenting with several thinning threshold values we found that thresholds between 1% and 10% produce reasonably dense networks capturing between 20% and 50% of all possible ties. This medium level density produces invariant model estimates and avoids degeneracy. The covariates that we include in the model are an indicator of whether legislators share the same party, the absolute difference in the first dimension of DW-Nominate, a count of committee service overlap, and a shared-state indicator. We also include network parameters for reciprocity (asymmetric tie count), cyclicality (count of cyclic triples), and transitivity (count of transitive triples). Recall that an asymmetric tie is one of the form $\{i \rightarrow j\}$ and a cyclic triple takes the form $\{i \rightarrow j, j \rightarrow k, k \rightarrow i\}$, whereas a transitive triple is of the form $\{i \rightarrow k, i \rightarrow j, j \rightarrow k\}$. We used the `R` package `ERGM` (Handcock et al. 2010) to fit the model.[5]

[Table 1 about here.]

The results of the ERGM with and without network statistics are presented in Table 1. A first look at model fit tells us that systemic properties of the Cosponsorship network are extremely important; the inclusion of the three network parameters reduces the BIC by more

---

[5]Complete R code for this application is in the Appendix.

than 25%. The signs of all of the covariates are in the expected direction. Representatives from the same state are more likely to be cosponsors and so are members from the same party, those who share committee memberships and those that are ideologically similar.

One striking finding is that when the network characteristics are accounted for, shared party is no longer statistically significant at any traditional threshold. The p-value for shared party is approximately .002 in the restricted model which does not account for network structure, but rises to 0.65 when network parameters are included. This is a clear and consequential example of the faulty weight which can be attributed to covariate effects when network structure is not properly accounted for. As can be seen in Table 1, artificially inflated covariate effects can easily result in faulty inference. The House network exhibits a high degree of reciprocity, a low degree of cyclicality and a high degree of transitivity; failing to account for these pronounced effects is akin to model misspecification and the restricted model in Table 1 reflects the resulting bias.

We can also use the ERGMs we have specified to produce better descriptive understandings of the relationships between legislators. Table 2 presents an alternative look at how taking account of network factors can change (and improve) inferences about dyadic outcomes in the cosponsorship network. Each off-diagonal entry is a three element set that includes (1) a binary indicator of a cosponsorship relationship, (2) a conditional probability of cosponsorship from the full model, (3) a conditional probability of cosponsorship from the covariates-only model. The rows correspond to the cosponsor and the columns the sponsor. We see that the full and covariates-only models produce noticeably different probabilities, and that, on average, the full model provides more accurate predictions.

[Table 2 about here.]

20

4.2   A Network Over Time:

Conflict in the International System

Our second application is drawn from international relations; we seek to replicate and re-analyze the model of militarized interstate disputes (MIDs) originally developed by Maoz et al. (2006) using ERGM technology. Our reasons for selecting this application are two fold: dyad-year data are extremely common in conflict scholarship and, because it is not uncommon for states to be involved in multiple disputes in a given year, independence is a problematic assumption. Few studies in international relations have specified inferential models of conflict as a network phenomenon; Maoz et al.'s (2006) article is appealing because the authors explicitly argue that international conflict should be treated as a network (other pioneering examples include Maoz (2006) and Hoff and Ward (2004)). We will go into greater detail regarding the specification of the model we replicate shortly, but first we present our innovation in the estimation of ERGMs for longitudinally observed networks which we use to analyze the conflict network.

4.2.1   The Pooled Exponential Random Graph Model

In this second application we construct a model that is as close as possible to one estimated in Maoz et al. (2006) in order to observe the substantive and technical advantages of working within the ERGM framework. The original application considers all dyads in the state system from 1870 – 1996, a total of 450,306 dyadic observations. The dependent variable is an indicator for the occurrence of at least one militarized interstate dispute (MID) between the two states in a given dyad (MID =1, No MID = 0). The method used in the original application is logistic regression, and, critically, one set of parameters is estimated that covers the entire time period – pooling all years into the same analysis. This last point poses a challenge for our replication; existing ERGM specifications and software have only been

developed to handle single networks and cannot be used to compute a single set of estimates that covers multiple pooled networks. We develop and implement a pooled ERGM for this application. One helpful quality of the pooled ERGM for the purpose of comparison with dyadic logit, is that the ERGM parameters reduce to those of logistic regression if the coefficients on the network statistics are zero (i.e. network structure is either not included or has exactly no effect).

The choice we would typically have to make between maximum pseudolikelihood and MCMC to estimate the pooled ERGM is nullified in this case: the computational burden associated with estimating the pooled ERGM by MCMC-MLE is insurmountable given current technology and the fact that we have to estimate the model on 126 networks. This computational problem will not necessarily affect every pooled ERGM analysis, but it is not unique to our application so it is worth considering. The problem presents at the Newton-Raphson optimization step in the iterative estimation of the ERGM parameters. Given the last set of parameters found by maximizing the approximate likelihood with the previous sample of networks, the current hill-climbing optimization requires that the entire sample of networks be held in memory. Because the size of the MID network varies by year, a unique sample of networks must be stored for each year. Each series of the 127 series of networks is of length 10,000 – using 10,000 networks for the likelihood approximation. Estimating such a model quickly overwhelms the capabilities of (currently) very powerful servers.[6] Instead of MCMC, we use maximum pseudolikelihood and implement a bootstrap resampling method (Efron 1981) for computing confidence intervals.

As it happens, the computational challenge of a large series of networks is something of a blessing in disguise. Given that, in most ERGM specifications, all of the dyads are dependent on each other, it is impossible to partition the sample of dyads into independent units. If such a partition were possible, a nonparametric resampling procedure such as the bootstrap

---

[6]An attempt at implementing this algorithm overloaded the 96 gigabytes of ram on a server.

or jackknife could be used to compute confidence intervals, and the analyst would not need to rely on the poor approximation provided by the inverse Fisher information. Since the current application consists of multiple MID networks over time, it is possible to implement a resampling method. The assumption underlying our approach is that, conditional upon the configuration of the network at time $t-1$ and exogenous covariates, dyads in t are independent of those in other years. To lend robustness to this assumption by accounting for memory in the MID network, we estimate each model with a lagged MID network as a dyadic covariate. The method we use is the non-parametric bootstrap, and the resampled unit is the network-year. This is a recently innovated bootstrapping approach for hierarchically organized data, where observations (e.g. dyads) are independent across units (e.g. years), but not within units (Field and Welsh 2007; Harden Forthcoming). For each iteration in each model, we resample with replacement 127 networks, find the maximum pseudolikelihood estimate for that sample and store the pooled ERGM estimates for that sample. We use 1,000 bootstrap iterations for each model. The 95% confidence intervals are constructed with the 2.5th and 97.5th percentiles of the sample of 1,000 bootstrap estimates.

4.2.2 Specification and Results

Maoz et al. (2006) argue that a network statistic called structural equivalence, which measures the similarity of ties held by nodes in a number of important international networks, is a measure of international affinity.[7] They compute structural equivalence statistics for each node with respect to alliance relationships, trade relationships, and IGO membership – ultimately combining these measures into an integrated structural equivalence (SEq) score.[8]

---

[7]See Maoz et al. (2006) for details.

[8]Maoz et al. (2006) estimate models with just the integrated measure and with the three components as independent variables. Since the BIC indicates that breaking the integrated score into its components does not improve model fit, we elect to replicate the model with the integrated score.

They find that *SEq* has a negative effect on the probability of conflict between any two states. We use Maoz et al.'s replication data and specify the same theoretical model they did: militarized interstate dispute are predicted by the dyad's weak-link regime score (Maoz and Russett 1993) (*Minimum Regime Score*), military-industrial capabilities ratio (*Capability Ratio*), geographical distance between capitols (*Distance*), and its integrated structural equivalence score (*SEq*).

We expand the Maoz et al. (2006) model to include two likely structural characteristics of the conflict network. First, it is likely that we will see an effect akin to the "popularity" of a state in the conflict network; by which we mean that the covariates may not capture everything that renders certain countries likely targets of conflict. In an age where international emergencies, including conflicts, receive highly organized and internationally coordinated responses, it is likely we will see association among states in their decisions to enter into conflict with other states. In the context of ERGMs, the "popularity" property is captured with the use of a two-star statistic (Robins et al. 2007), which is the number of times in the network where two states are at war with the same state.

Second, we contend that another local configuration, the triangle, is especially unlikely to be present in the conflict network. More specifically, if state i is at war with j and state k is at war with j (i.e. i and k have a common enemy), it is very unlikely that i and k are also at war. At a minimum, a war between i and k would be counter-productive to their efforts against j, weakening the enemy of their enemy. With regard to network effects, the presence of more two-stars and less triangles should increase the likelihood of observing a particular realization of the conflict network. Specifically, we expect that the two-star parameter will be positive and the triangles parameter will be negative.

[Table 3 about here.]

The results from the logistic regression and pooled ERGM analysis of the MID net-

work are presented in table 3. The results confirm our expectations regarding the network structure that characterizes the MID network. The positive two-star coefficient, which is statistically significant at the 0.05 level based on the 95% bootstrapped confidence interval, indicates that there are forces above and beyond the covariates that motivate collections of states become involved in conflicts with the same country. The negative and statistically significant effect of triangles confirms that states are unlikely to enter into conflicts with the enemy of their enemy. We should note that this last finding, though highly intuitive, is impossible to accommodate outside of the ERGM framework. Overall, we find that the network structure represents an essential component of the process underlying the MID network.

Equally important considerations in the analysis are the effects that accounting for network structure have on inferences relating to the dyadic covariates. A particularly important inference changes once we move from dyadic logit to ERGM: the democratic peace effect. The variable *Minimum Regime Score* measures the minimum democratic regime score among the two states in the dyad. Within the logistic regression framework of the original analysis, this variable has a negative effect on the likelihood of conflict that is statistically significant at the 0.05 level. This is consistent with the democratic peace – the theory suggesting that democratic states are unlikely to fight one another. In the pooled ERGM results, the effect of *Minimum Regime Score* is not distinguishable from zero at the 0.05 level of significance. In other words, we cannot reject the null hypothesis that the effect of *Minimum Regime Score*, the operationalization of the democratic peace, is equal to zero.

In the previous two paragraphs we have claimed findings of interesting network structure in the MID network, and called into question one of the central findings of international conflict literature. This, we hope, illustrates the inferential power and broad applicability of the pooled ERGM. Yet one question we have yet to answer is whether the pooled ERGM provides a better fit to the data than logistic regression. Given that we have used maximum pseudolikelihood, and not maximum likelihood, assessing model fit requires more work be-

cause we cannot rely on information-theoretic parametric measures of fit such as the BIC or AIC. Though we cannot claim to have discovered the set of parameters that maximize the likelihood function, in our use of the bootstrapped pseudolikelihood, we have provided an alternative algorithm for finding pooled ERGM parameters. These parameters constitute an ERGM distribution for each network in the series. We can simulate hypothetical conflict networks from the distributions implied by the estimated parameters, and the simulated networks can be compared to the observed networks in order to assess model fit. Both the logistic regression and pooled ERGM estimates can thus be used to predict the binary event that there is or is not a MID between two states in any given year.

The receiver operating characteristic (ROC) curve is a tool that can be used to non-parametrically assess the performance of competing methods for classifying binary outcomes (Pete, Pattipati and Kleinman 1993; Pepe 2000). In this case, we will use it to compare logit to ERGM for classifying MIDs. The ROC curve plots the degree of true positive predictions against the degree of false positive predictions. The area under the ROC curve gives the integrated difference between the predictive success and predictive error of a classifier, and represents a scalar-valued measure of model fit. Similar to an R^2 in linear regression, the greater the area under the ROC curve, the better the predictive value of a model. If the objects were classified perfectly, the area under the ROC curve is 1.0 and if the classifier is never right, the area under the curve is 0. The ROC curve has been used previously to assess the ability of an ERGM to predict edge values (Saul and Filkov 2007). We use the ROC curve to compare the performance of the specifications in table 3. The ROC curves as well as the areas under them are presented in figure 3. We see that the ERGM with a lagged network performs the best, the ERGM without a lagged network performs better than the original logistic regression model, and the logit with a lagged network performs better than the ERGM without a lagged network. As such, we have identified two important, and previously unverified, factors related to the prediction of conflict. First, there is

26

memory in the conflict network, and it is thus important to account for the past behavior of a dyad. Second, network structure is non-trivially important. Whether a lagged network is or is not included, moving from the logistic to the ERGM modeling framework with network structure improves the predictive performance of the model. This reinforces confidence in findings that (1) there is a "popularity" effect in the MID network, (2) states are not likely to fight the enemies of their enemies, and (3) once we have controlled for network structure, the effect of the democratic peace drops out of significance.

[Figure 3 about here.]

5 Concluding Thoughts

We have argued that relational data, as it commonly occurs in political science, can be conceptualized as network data rather than as a set of conditionally independent dyads. We have also introduced and extended the Exponential Random Graph Model (ERGM) as a means of conducting regression-style inference on network data. The problem with using traditional regression models on network data is that those models assume that the observations (dyadic relationships) are conditionally independent of one another. When violated, this assumption causes explanatory power attributable to the interdependence between nodes to be falsely attributed to the covariates. Many before us have noted that this is a substantively problematic assumption, but because network analysis technology such as the ERGM has only recently become available, researchers have had little choice but to use standard regression models on network data or rely on descriptive statistics to capture network effects.

We reviewed the basic concept of the ERGM, is strengths, limitations, and estimation. We then proposed our own solutions to two of the ERGM's limitations: difficulties with overly dense and edge-valued networks and its inability to cope with longitudinally observed

networks. Perhaps our most notable findings come from our applications of the ERGM to political science data. First, we found that several previously unexplored network parameters are strong predictors of the U.S. House of Representatives' legislative cosponsorship network. Specifically, we found that the House cosponsorship network exhibits high reciprocity, low cyclicality and high transitivity, *and* that the failure to account for network structure in the model would lead one to falsely conclude that party-based assortative mixing characterizes the cosponsorship network. Then, in our international conflict application, we made the rather striking discovery that some strong and significant findings generated through standard logistic regression drop out of significance when analyzed with an ERGM; this suggest the possibility that logistic regression was indeed falsely attributing effect strength and significance to the covariates. We hope that our arguments and findings will serve as both a caution against the use of traditional regression methods on network data and as a set of guidelines for researchers to draw sound inferences from relational data.

Appendix 1: Full Derivation of the ERGM

This derivation generally follows Park and Newman (2004) with only minor modifications to their notation. We seek to identify the probability distribution that will maximize Gibbs entropy

$$S = - \sum_{G_i \in G_M} P(G) \ln P(G) \tag{5}$$

subject to the assumed constraint

$$\sum_{i=1}^{M} P(G_i) \Gamma_i = E[\Gamma_i] \tag{6}$$

and the normalizing condition

$$\sum_{i=1}^{M} P(G_i) = 1, \tag{7}$$

where Γ_i is a statistic computed on the graph G. We can then identify the maximum entropy distribution by introducing the Lagrange multipliers α and θ_i. The distribution satisfying

$$\frac{\partial}{\partial P(G_i)} \left[S + \alpha \left(1 - \sum_{i=1}^{M} P(G_i) \right) + \sum_{j} \theta_j \left(E[\Gamma_j] - \sum_{i=1}^{M} P(G_i) \Gamma_{ij} \right) \right] = 0 \tag{8}$$

for all G_M has maximum entropy. Taking the derivative gives,

$$\ln P(G_i) + 1 + \alpha + \sum_{j} \theta_j \Gamma_j = 0 \tag{9}$$

which, solving for $P(G_i)$, gives

$$P(G_i) = \frac{\exp(-H(G_i))}{Z} \tag{10}$$

where $H(G_i)$ is the Hamiltonian

$$H(G_i) = \sum_j \theta_j \Gamma_{ij} \tag{11}$$

and Z is the partition function

$$Z = \exp(\alpha + 1) = \sum_{v=1}^{M} \exp(-H(G_i)) = \sum_{v=1}^{M} \exp\left(-\sum_{j=1}^{k} \theta_j \Gamma_{vj}\right). \tag{12}$$

Substituting 11 and 12 into 10 gives

$$P(G_i) = \frac{\exp(-\sum_{j=1}^{k} \Gamma_{ij}\theta_j)}{\sum_{v=1}^{M} \exp(-\sum_{j=1}^{k} \Gamma_{vj}\theta_j)} \tag{13}$$

which is presented as equation 3 in the main text.

Appendix 2: R Code for Conducting ERGM Analysis

Below is our replication R code. We have commented it as thoroughly as possible so as to demonstrate the use of ERGM software.

```
install.packages("ergm")
library(ergm)


#Assume the network is in a simple matrix of zeros and ones.
ergm.fit <- ergm(network(net,directed=F) ~ edges + edgecov(minregl)
    + edgecov(capratl) + edgecov(distancel) + edgecov(atopsl)
    + edgecov(tradesl) + edgecov(igosecl), maxit=50)


#The network() command transforms the matrix into a proper network object
#The edgecov() command tells ergm that the matrix in () is to be treated
#as a dyadic covariate
#The parameter estimates and standard errors can be obtained by:


summary(ergm.fit)
#Many other useful functions are defined in the ergm documentation
help("ergm")
```

References

Ahlquist, John Stephen and Michael D. Ward. 2009. "Factor Endowments, Trade, and Democracy." *Paper presented at the Annual Meeting of the Midwest Political Science Association* .

Baybeck, B. and R. Huckfeldt. 2002. "Spatially Dispersed Ties Among Interdependent Citizens: Connecting Individuals and Aggregates." *Political Analysis* 10(3):261–275.

Besag, Julian. 1975. "Statistical Analysis of Non-Lattice Data." *Journal of the Royal Statistical Society. Series D (The Statistician)* 24(3):179–195.

Cao, Xun, Aseem Prakash and Michael D. Ward. 2007. "Protecting Jobs in the Age of Globalization: Examining the Relative Salience of Social Welfare and Industrial Subsidies in OECD Countries." *International Studies Quarterly* 51(2).

Cranmer, Skyler J., Bruce A. Desmarais and Justin H. Kirkland. 2010. "Towards A Network Theory of Alliance Formation." *Working Paper* .

Crescenzi, Mark J. C. 2007. "Reputation and Interstate Conflict." *American Journal of Political Science* 51(2):382–396.

Danny, Wyatt, Tanzeem Choudhury and Jeff Bilmes. 2009. "Dynamic Multi-Valued Network Models for Predicting Face-to-Face Conversations." *Proceedings of the NIPS-09 workshop on Analyzing Networks and Learning with Graphs.* .

Efron, Bradley. 1981. "Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods." *Biometrika* 68(3):589–599.

Faust, Katherine and John Skvoretz. 2002. "Comparing Networks across Space and Time, Size and Species." *Sociological Methodology* 32:267–299.

Field, Christopher A. and A. H. Welsh. 2007. "Bootstrapping clustered data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69(3):369 – 390.

Fienberg, S. E. and S.S. Wasserman. 1981. "Categorical Data Analysis of Single Sociometric Relations." *Sociological Methodology* 12:156–192.

Fowler, James .H. 2006. "Connecting the Congress: A study of cosponsorship networks." *Political Analysis* 14(4):456.

Frank, Ove and David Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81(395):832–842.

Franzese, R.J. Jr. and J.C. Hays. 2006. "Strategic interaction among EU governments in active labor market policy-making: Subsidiarity and policy coordination under the European employment strategy." *European Union Politics* 7(2):167.

Geyer, Charles J. and Elizabeth A. Thompson. 1992. "Constrained Monte Carlo Maximum Likelihood for Dependent Data." *Journal of the Royal Statistical Society. Series B (Methodological)* 54(3):657–699.

Goodreau, Steven .M., J. 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):103–25.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, Martina Morris and Pavel Krivitsky. 2010. *ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks.* Seattle, WA: . Version 2.2-2. Project home page at url-http://statnetproject.org.
URL: *http://CRAN.R-project.org/package=ergm*

Handcock, Mark Stephen and Krista Jennifer Gile. forthcoming. "Modeling Networks from Sampled Data." *Annals of Applied Statistics* .

Handcock, M.S. 2003. Assessing degeneracy in statistical models of social networks. In *Workshop on Dynamic Social Network Analysis, Washington, DC, November.*

Handcock, M.S., D.R. Hunter, C.T. Butts, S.M. Goodreau and M. Morris. 2008. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24(3):1–29.

Harden, Jeffrey J. Forthcoming. "A Bootstrap Method for Conducting Statistical Inference with Clustered Data." *State Politics and Policy Quarterly* .

Hoff, P.D. and M.D. Ward. 2004. "Modeling dependencies in international relations networks." *Political Analysis* 12(2):160–175.

Hoff, Peter D., Adrian E. Raftery and Mark S. Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97(460):1090–1098.

Holland, Paul W. and Samuel Leinhardt. 1981. "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 76(373):33–50.

Kessler, Daniel and Keith Krehbiel. 1996. "Dynamics of Cosponsorship." *The American Political Science Review* 90(3):555–566.

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing incomplete political science data: An alternative algorithm for multiple imputation." *American Political Science Review* 95(1):49–69.

Lazer, David. 2005. "Regulatory Capitalism as a Networked Order: The International System as an Informational Network." *Annals of the American Academy of Political and Social Science* 598:52–66.

Maoz, Z., RD Kuperman, L. Terris and I. Talmud. 2006. "Structural Equivalence and International Conflict, 1816–2001. A Network Analysis of Affinities and Conflict." *Journal of Conflict Resolution* 50(5):1–26.

Maoz, Zeev. 2006. "Network Polarization, Network Interdependence, and International Conflict, 1816-2002." *Journal of Peace Research* 43(4):391–411.

Maoz, Zeev. 2009. "The Effects of Strategic and Economic Interdependence on International Conflict Across Levels of Analysis." *American Journal of Political Science* 53(1):223–240.

Maoz, Zeev and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *The American Political Science Review* 87(3):624–638.

Morris, M., M.S. Handcock and D.R. Hunter. 2008. "Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects." *Journal of statistical software* 24(4):1548.

Park, Juyong. and M.E.J. Newman. 2004. "Statistical mechanics of networks." *Physical Review E* 70(8):66117–66130.

Pepe, Margaret Sullivan. 2000. "An Interpretation for the ROC Curve and Inference Using GLM Procedures." *Biometrics* 56(2):352–359.

Pete, Andras, Krishna R. Pattipati and David L. Kleinman. 1993. "Optimal Team and Individual Decision Rules in Uncertain Dichotomous Situations." *Public Choice* 75(3):205–230.

Robins, Garry, Tom Snijders, Peng Wang, Mark Handcock and Philippa Pattison. 2007. "Recent developments in exponential random graph (p*) models for social networks." *Social Networks* 29(2):192 – 215.

Rubin, Donald B. 1976. "Inference and missing data." *Biometrika* 1976:581–592.

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

Schneider, Mark, John Scholz, Mark Lubell, Denisa Mindruta and Matthew Edwardsen. 2003. "Building Consensual Institutions: Networks and the National Estuary Program." *American Journal of Political Science* 47(1):143–158.

Scholz, John T. and Cheng-Lung Wang. 2006. "Cooptation or Transformation? Local Policy Networks and Federal Regulatory Enforcement." *American Journal of Political Science* 50(1):81–97.

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

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins and Mark S. Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36:99–153.

Strauss, David and Michael Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85(409):204–212.

van Duijn, Marijtje A.J., Krista J. Gile and Mark 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(1):52 – 62.

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

Ward, Michael D., Randolph M. Siverson and Xun Cao. 2007. "Disputes, Democracies, and Dependencies: A Re-examination of the Kantian Peace." *American Journal of Political Science* 51(3).

Wasserman, S. and P. Pattison. 1996. "Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^*." *Psychometrika* 61(3):401–425.

t = number of simulated networks used to approximate likelihood
α = threshold for stopping iterative optimization
$\boldsymbol{\theta}$ = parameter vector
Γ_{hm} = the mth (out of k) statistics computed on the hth network.
Δ_{ll} = change in log-likelihood
o = indicator for the observed network
LL = log-likelihood

Initialize Δ_{ll} to ∞
Initialize LL to $-\infty$
Initialize $\boldsymbol{\theta}$ to starting values

while($\Delta_{ll} > \alpha$){
1. Draw t networks from the distribution parameterized with $\boldsymbol{\theta}$
2. Using a hill-climbing algorithm, find $\boldsymbol{\theta}^*$ to maximize $LL^* = \log\left(\frac{\exp(-\sum_{j=1}^{k}\Gamma_{oj}\theta_j^*)}{\sum_{i=1}^{t}\exp(-\sum_{j=1}^{k}\Gamma_{ij}\theta_j^*)}\right)$
3. Store $\boldsymbol{\theta} = \boldsymbol{\theta}^*$
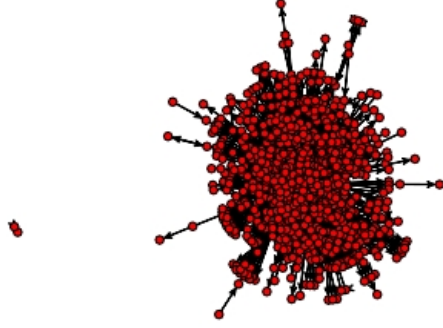4. Store $\Delta_{ll} = LL^* - LL$
5. Store $LL = LL^*$
}

$\boldsymbol{\theta}$ is now the MCMC-MLE

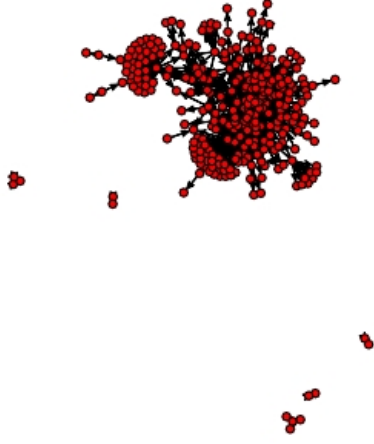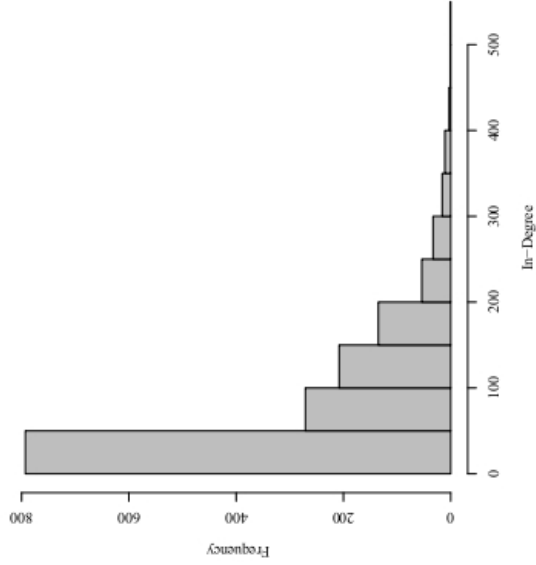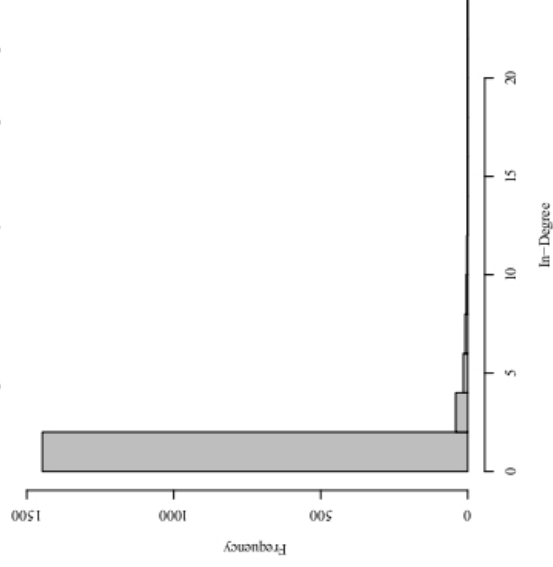Figure 1: This figure gives the MCMC-MLE algorithm used to estimate ERGMs.
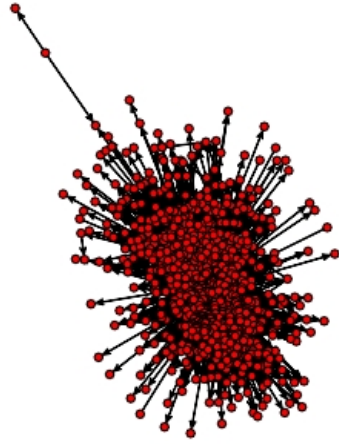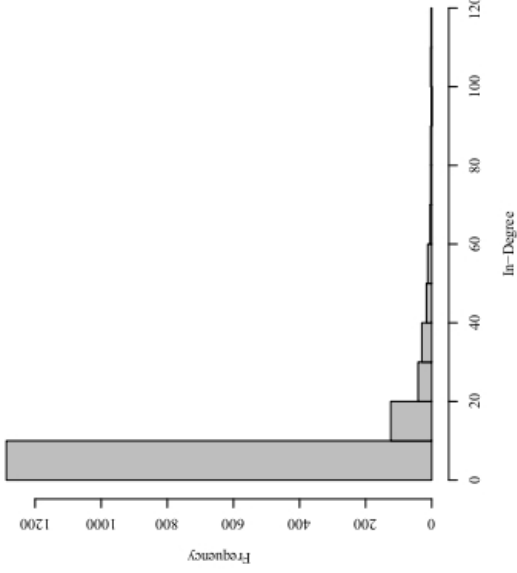
Figure 2: This figure has six cells each showing the 108th House network at the same (interesting) time t: the [1,1] cell shows the completely un-thinned network (an edge exists if there was <u>any</u> cosponsorship), the [1, 2] cell shows the network with a low thinning threshold, and the [1, 3] cell shows the network with a high thinning threshold (most sparse). The [2, 1] cell shows a histogram of the number of cosponsorship ties between legislators. The [2, 2] and [2, 3] cells show the same thing after thinning.
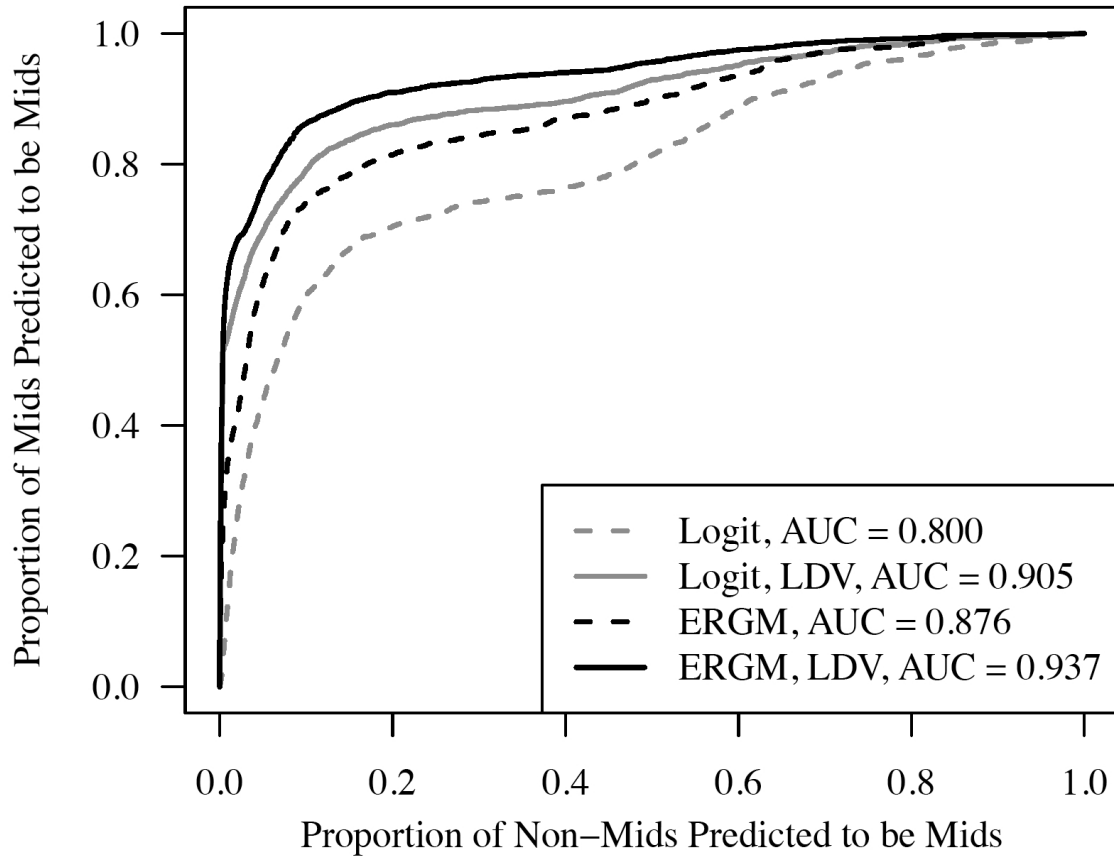
Figure 3: ROC curves demonstrating the in-sample predictive performance of the models for MIDs. The area under the curve ROC curve (AUC) listed in the legend is a common nonparametric measure of predictive fit. The closer the AUC is to one, the closer the model is to perfect classificaiton.

	Full Model			Covariates-Only Model		
Term	Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
Edges	-2.897	0.0957	0.0000	-0.6778	0.0300	0.0000
Same State	2.0262	0.2722	0.0000	1.80230	0.02492	0.0000
Nominate Distance	-0.9951	0.2483	0.0044	-1.3932	0.03260	0.000
Same Party	0.0106	0.0211	0.6147	0.07752	0.02562	.0025
Shared Committees	0.4670	0.0371	0.0000	0.3541	0.0110	0.0000
Asymmetric Ties	-0.1989	0.0562	0.0000	–	–	–
Cyclic Triples	-0.0340	0.0001	0.0000	–	–	–
Transitive Triples	0.0269	0.0000	0.0000	–	–	–
BIC	140,523			197,019		

Table 1: ERGM fit for the 108th U.S. House Cosponsorship Network

	Pelosi	Kucinich	Gephardt	Delay	Hastert	Pryce
Pelosi	0	(0, 0.023, 0.311)*	(0, 0.144, 0.354)*	(0, 0.034, 0.137)*	(1, 0.054, 0.147)	(0, 0.23, 0.122)
Kucinich	(1, 0.932, 0.311)*	0	(0, 0.151, 0.311)*	(0, 0.023, 0.084)*	(0, 0.028, 0.09)*	(1, 0.698, 0.409)*
Gephardt	(1, 0.29, 0.354)	(0, 0.067, 0.311)*	0	(0, 0.02, 0.1)*	(0, 0.024, 0.108)*	(0, 0.042, 0.122)*
Delay	(0, 0.034, 0.137)*	(0, 0.011, 0.084)*	(0, 0.016, 0.1)*	0	(1, 0.15, 0.418)	(0, 0.091, 0.305)*
Hastert	(0, 0.05, 0.147)*	(0, 0.009, 0.09)*	(0, 0.014, 0.108)*	(0, 0.095, 0.418)*	0	(0, 0.067, 0.323)*
Pryce	(0, 0.172, 0.122)	(0, 0.081, 0.409)*	(0, 0.029, 0.122)*	(1, 0.075, 0.305)	(1, 0.127, 0.323)	0

Table 2: Selected House Sociomatrix with Conditional Probabilities. Each entry gives ((1) binary indicator of a cosponsorship relationship, (2) conditional probability of cosponsorship from the full model, (3) conditional probability of cosponsorship from the covariates-only model). The rows correspond to the cosponsor and the columns the sponsor.* Indicates the full model makes the better prediction.

	Logit	Logit, LDV	ERGM	ERGM, LDV
Edges	-3.14	-4.42	-3.61	-4.6
	[-3.43, -2.87]	[-4.68, -4.17]	[-3.79, -3.44]	[-4.79, -4.41]
Minimum Regime Score	-0.003	-0.002	-0.001	-0.001
	[-0.006, -0.00043]	[-0.0054, -0.00047]	[-0.0035, 0.00010]	[-0.0035, 0.00031]
Capability Ratio	0.00029	0.00027	0.00011	0.00021
	[0.0001, 0.0004]	[0.0001, 0.00039]	[-0.00010, 0.00028]	[0.00031, 0.00034]
Distance	-0.0005	-0.0003	-0.0005	-0.0003
	[-0.0006, -0.0004]	[-0.0004, -0.0002]	[-0.0006, -0.0005]	[-0.0004, -0.0003]
Integrated SEq	-0.867	-0.605	-0.511	-0.344
	[-1.09, -0.645]	[-0.822, -0.39]	[-0.682, -0.352]	[-0.544, -0.171]
Lagged MID	—	5.13	—	4.67
	—	[4.88, 5.35]	—	[4.48, 4.86]
Two Stars	—	—	0.335	0.272
	—	—	[0.302, 0.363]	[0.237, 0.308]
Triads	—	—	-0.583	-0.482
	—	—	[-0.743, -0.426]	[-0.707, -0.279]

Table 3: Replication and extension of model 1 from Maoz et al. (2006). Note, the baseline results presented in the first column of this table differ slightly from the original results in Maoz et al. (2006), because we omit the first year for the lagged network. Each model is estimated on 449,933 dyads representing the MID networks from 1870-1996. The estimates are given in the columns with 95% confidence intervals in brackets below. Confidence intervals not containing zero are significant at the common 0.05 p-value threshold.