Advance Access publication December 30, 2010

Political Analysis (2011) 19:66–86
doi:10.1093/pan/mpq037

Inferential Network Analysis with Exponential Random Graph Models

Skyler J. Cranmer
Department of Political Science, University of North Carolina at Chapel Hill, 311 Hamilton Hall, Chapel Hill, NC 27599
e-mail: skyler@unc.edu (corresponding author)

Bruce A. Desmarais
Department of Political Science, University of Massachusetts Amherst, Thompson Hall, 200 Hicks Way, Amherst, MA 01003
e-mail:desmarais@polsci.umass.edu

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 introduce and evaluate a general model for inference with network data, the Exponential Random Graph Model (ERGM) and several of its recent extensions. The ERGM simultaneously allows both inference on covariates and for arbitrarily complex network structures to be modeled. Our contributions are three-fold: beyond introducing the ERGM and discussing its limitations, we discuss extensions to the model that allow for the analysis of non-binary and longitudinally observed 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 political networks have not kept pace with the breadth and complexity of the substantive questions being asked of such data. In relational data, it is common that the relational tie between a given pair of actors depends on one or more of the other ties in the network. For example, if we consider the conflict network of World War II, it is clear that the United Kingdom's choice to declare war on Germany was strongly influenced by the conflict ties that Germany had already initiated with other states, Poland in particular. Relational dependencies need not even exert their influence on actors directly involved in the influential relation: Canada, the Netherlands, the United Kingdom, Australia, New Zealand, and others declared war on Japan in the immediate wake of the Pearl Harbor attack even though none of them were directly involved. The extent of the complex dependence can be further seen by fact that Italy and Germany then quickly declared war on the United States. This *dependence* among relationships in network data constitutes the core statistical challenge, and also a great opportunity, associated with the empirical study of networks.

The two common approaches to the statistical analysis of relational data in political science are (1) to assume the covariates in a dyadic regression model are sufficient to account for dependence among the observations and (2) to condition out interdependence through the innovative use of random effects. The first approach is necessarily biased in the many instances where relationships between the observations cannot be represented as a covariate, violating the regression assumption of conditional independence of observations. Complex random-effect specifications are often sufficient to condition out the dependence among the

Authors' note: Many thanks to Tom Carsey, James Fowler, Justin Gross, and Peter Mucha for their valuable comments on a previous version of this article. Supplementary materials for this article are available on the *Political Analysis* Web site.

observations, but such approaches are not generally capable of characterizing the precise forms of interdependence that characterize the data—properties that may be theoretically important to political scientists.

Our aim is to introduce a model for inference on network data called the Exponential Random Graph Model (ERGM) as well as several of its recent extensions and demonstrate the application of these methods on political science data. The ERGM is a statistical model that can be used to estimate the effects of covariates on the ties in a network while simultaneously estimating parameters that provide a precise and parsimonious description of the forms of dependence that can exist in relational data. We introduce the ERGM, its estimation, and its limitations; the most restrictive of which are the fact that ERGMs are only defined for single networks with binary ties. We then, by means of example, discuss the common applied problems of valued edges and longitudinally observed networks—demonstrating some extensions of the ERGM that allow the analyst to overcome these problems. Our applications show that using ERGMs can not only avoid faulty inference on covariates but can also provide new insight to substantive problems via empirical characterization of the interdependence among political relationships.

Standard regression models are designed to estimate the effect of covariates on the outcome, not to estimate the influence of the outcomes on each other. Therefore, the regression framework is not only unable to model network dependencies but the validity of regression results breaks down when such dependencies go unmodeled. Consider, for example, a model of international conflict where the outcome variable is war and the single independent variable is joint democracy; both variables are measured on the dyad. If it is the case that the enemy of my enemy is likely to be my friend, the outcome observations (conflict relationships) are nonindependent because of an endogenous effect in the network generating process. A triadic effect such as this cannot be captured by covariates and so, in the context of a traditional regression, must be omitted from the model specification. The effect of this omission will be bias in the coefficients and inconsistency in the SEs because of model misspecification. The bias here is a result of unobserved heterogeneity. Proper specification is essential to produce a valid result for any type of parameter (covariate or structural). As such, even if the researcher is only interested in evaluating a particular hypothesis about a covariate, bias due to the omission of relevant structural effects can compromise the analysis. Furthermore, clustering the SEs could not correct the problem because the nonindependence of nodes in complex networks prevents the node-wise partitioning necessary to cluster the SEs.[1]

The ERGM permits construction of a comprehensive model for the state of the network through a mechanism that departs markedly from regression. Because the observed network is treated as a single realization from a multivariate distribution, no assumptions about the independence of actors or ties within the network are necessary. Furthermore, the ERGM can model both exogenous effects (covariates) and effects that are endogenous to the network (structural effects in the network). Dependence terms in an ERGM specification, which are endogenous measures on the network, capture theories postulating relationships among the ties in the network (e.g., the friend of my friend is my friend). In other words, researchers can proceed with ERGM analysis based on hypotheses similar to those that would produce regression specifications (i.e., covariate x is expected to affect the outcome y), and as much network structure (dependence) as they see fit. Moreover, the ERGM is poised for widespread application in political science and is implemented in a number of software packages. The ERGM is, as we will show, widely applicable to network analysis in political science and is remarkably flexible in its ability to model relational interdependence.

We are not the first to have noticed that treating relational data as independent, an assumption required by the regression framework, is a problematic practice (Hoff and Ward 2004; Ward, Siverson, and Cao 2007). Qualitative theory in various subfields has discussed the interdependence of relationships for some time (i.e., Hoag 1960, Rosecrance and Stein 1973, Rosecrance et al. 1977, Keohane and Nye 1989, and Kroll 1993) and descriptive techniques for network data are now reasonably well known (i.e., Fowler 2006 and Maoz et al. 2006; Maoz 2006, 2009). Although 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; Schneider et al. 2003; Lazer 2005; Fowler 2006; Maoz et al. 2006; Franzese and Hays 2006; Scholz and Wang 2006; Crescenzi 2007; Ward and Hoff 2007; Ahlquist and Ward 2009; Hays, Kachi, and Franzese 2010).

[1]Additionally, unlike the linear regression model, robust-clustered SEs have been shown to be inconsistent for grouped discrete-choice data (i.e., clustered logit) Greene (2008, p. 517).

One effort to extend the regression framework to the network context involves the inclusion of network statistics as covariates in regression analyses. 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 endogenous network structures to network outcomes. However, because it applies standard regression to account for this interdependence, it does not do away with the independence assumption and, as such, does not solve the bias problem discussed above. Furthermore, this approach treats structural effects, which are necessarily endogenous to the outcome network, as exogenous to the network. The ERGM framework we will introduce easily handles both endogenous and exogenous effects and, more importantly, departs from the regression framework and its independence assumptions.

A number of promising approaches have been developed outside the ERGM framework (Hoff, Raftery, and Handcock 2002; Hoff and Ward 2004; Franzese and Hays 2006; Hays, Kachi, and Franzese 2010). Perhaps the most visible approach 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 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 networks; the best approach will vary by application based on the substance of the problem and the quest for a universally best method is quixotic. Instead, we seek to introduce the ERGM and some of its extensions as an approach that may be useful in many political science analyses, particularly those where actors are thought to condition their tie-formation behaviors on the ties formed by others in the network and/or exogenous covariates.

Though the general framework for ERGMs is well developed, there remain a number of applied problems that currently limit the utility of ERGMs for political scientists. Chief among these problems are the inability of ERGMs to model longitudinally observed networks or handle non-binary edge types. We present extensions to the basic ERGM that can overcome the problem in the case of longitudinally observed networks and can work around the valued-edges problem by transforming a valued-edge network into a binary network. We demonstrate the efficacy of these extensions with two illustrative examples—one on cosponsorship in the U.S. House of Representatives and another on international conflict.

We begin by deriving the ERGM and discussing the ways in which it requires us to think differently about network data, how it accommodates dependence among the edges, how it can be estimated, and some limitations of the model. We demonstrate the ERGM, along with extensions to valued (non-binary) edges and longitudinally observed networks using two examples chosen for both salience and exposition: cosponsorship networks in the U.S. Congress and conflict networks in the international system. We feel that these are good expository applications because they pose challenges to ERGM analysis that are likely to be encountered in many other political network analyses: the cosponsorship network has non-binary edges and the conflict network is observed over a long period of time. 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 The Exponential Random Graph Model

The ERGM has evolved out of a series of efforts by scholars in many disciplines to address the same limitations of classical statistical models that 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.[2] 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 specified ERGM was first derived.

To make inference about the determinants of outcomes that are dependent upon each other (i.e., the ties within a network), even if one is only interested in the effects of exogenous covariates on the individual outcomes, those individual outcomes must be combined within the context of a multivariate distribution (i.e., joint distribution) that accurately represents the dependence among them (Hyvrinen, Hurri and Hoyer 2009). The failure of a model to recognize dependences among outcomes is as threatening to the validity of results as omitting an important covariate. There are many established methods in political methodology that take this approach—accounting for correlation among observed or latent outcomes that is not captured by the covariates—including, seemingly unrelated regression (Jackson 2002), multinomial probit (Alvarez and Nagler 1998), spatial regression (Franzese and Hays 2007; Hays, Kachi, and Franzese 2010), and copula modeling (Boehmke 2006). Aside from being tailored to network data, the characteristic that differentiates the ERGM from these other multivariate methods is that the researcher must fully specify the forms of interdependence that are captured by the model. With that in mind, an appropriate model for statistical inference on network data must meet two criteria. First, the model must be sufficiently general so as to allow ties to depend upon covariates as well as other ties in the network. Second, to allow for this dependence, the model must be derived as a proper joint distribution so as to avoid, altogether, the restrictive assumption that the ties are conditionally independent. The ERGM meets both of these criteria.[3]

In order to construct a model without assuming relational independence, we need to think in a somewhat different fashion about the data generating process. We can construct a likelihood function for any network 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 conditional 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. 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. It is important to note that this generalization does not require us to abandon the ability to characterize the conditional probability of a tie given covariates, as in a regression. Once a proper joint distribution is defined for the ties in the network, the conditional distribution of any tie given covariates is implied. We provide the formula for the conditional probability of a tie in Section 2.2 on interpretation below (equation 5 specifically). 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 Γ_i, $i = 1,.., k$ that can be computed on the network (or graph) Y—an $N \times N$ adjacency matrix—which the analyst believes affect the likelihood of observing Y. These statistics can include measures of interconnectedness, reciprocity in directed networks, and node (e.g., political party of a legislator) or dyad-level (e.g., trade between two states) covariates. For example, node and dyad-level covariates, X_n and X_d, respectively, are included in an ERGM via

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

[3]To revisit a topic from above, including network statistics as predictors in a regression meets the first criteria in that ties are rendered dependent upon other ties, but does not meet the second.

$$\Gamma_{X_n}(Y, X_n) = \sum_{i \neq j} X_i X_j Y_{ij} \text{ and } \Gamma_{X_d}(Y, X_d) = \sum_{i \neq j} X_{ij} Y_{ij}, \tag{1}$$

and reciprocity, a dependence term, is accommodated as

$$\Gamma_R(Y) = \sum_{i<j} Y_{ij} Y_{ji}. \tag{2}$$

Like with the selection of independent variables, the selection of network statistics should be motivated by theory. For example, in our international relations application below, we have a theoretical reason for believing that the enemy of my enemy should not be my enemy. In other words, we suspect that we will find a negative coefficient for the count of closed triangles in the conflict network. Although a detailed review of possible network statistics to include is beyond the scope of this discussion, interested readers are referred to Handcock et al. (2010) 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 Y, there are M possible sets of edges on N. This means that the observed network Y 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 Y arises stochastically from the support $Y_{M.}$

The general ERGM is derived using a moment-matching exercise. We start by making one of only two assumptions that are invoked to derive the ERGM: we assume that each of the network statistics calculated on graph Y are the expected values of those statistics across all possible graphs: $E[\Gamma_i] = \Gamma_i$, 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 Y. Although 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 Y_M:

$$E[\Gamma_m] = \sum_{m=1}^{M} P(Y_m)\Gamma_m. \tag{3}$$

Yet the distribution over Y_M is not fully identified at this point. We must also assume that only the statistics included in Γ influence the probability that graph m is observed, $P(Y_{\mathrm{m}})$.[4] This is accomplished by maximizing the Gibbs entropy, $S = -\sum_{m=1}^{M} P(Y_m)\ln P(Y_m)$, on the discrete distribution of graphs in Y_M subject to the condition in equation (3). In effect, the network statistics pull the distribution of networks away from a uniform distribution over all possible networks to the extent that such a weighting increases the likelihood of observing Y_m.

As a result of this optimization, we can recover an elegant formula reflecting the relationship between the probability of a network m and the network statistics in Γ,

$$P(Y_m) = \frac{\exp\left(-\sum_{j=1}^{k} \Gamma_{mj}\theta_j\right)}{\sum_{m=1}^{M} \exp\left(-\sum_{j=1}^{k} \Gamma_{mj}\theta_j\right)}, \tag{4}$$

where θ is the vector of k parameters that describe the dependence of $P(Y_m)$ on the network statistics in Γ. Equation (4) is the full joint distribution of the edges in Y. We describe below, a number of ways that this distribution can be used for inference on θ.

[4]This boils down to an assumption that the model is correctly specified—an assumption that underlies many methods used for statistical inference in political science, including the generalized linear model.

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 (i.e., the model is correctly specified). 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.

The ERGM is a natural way to model a number of network generating processes, both structural (i.e., network level) and individual, common to political science. One approach to thinking about network generating processes is to hypothesize that certain substructures are likely to emerge in the network. For example, theory may imply that dense clusters (e.g., polarized political parties) should exist in the network; a theory that may not be well articulated at the individual choice level. The ERGM can also be used to study relationships among individuals who are thought to condition their individual choices on the behaviors of others in the network. An alternative way to think about the ERGM and the network generating process is that the ERGM can be viewed as the limiting joint distribution of ties formed conditional on all other ties in the network. In other words, when individuals—say legislators—condition their tie formation behavior on the tie formation behavior of others in the network, a complex structure easily modeled with the ERGM emerges. The applications in sections 4.1 and 4.2 make use of the individual and structural perspectives on the network generating process, respectively.

2.1 *A Small-Scale Illustration*

As noted above, the largest conceptual difference between the conventional regression framework and the ERGM is that the model used to describe the data is specified at the network level and cannot be constructed from the individual links or nodes *independently*. Given a set of network statistics (Γ), the parameters (θ) are estimated to maximize the likelihood of observing the network on hand. To fit the ERGM well, the analyst should choose Γ such that the vector of statistics collectively differentiates the observed network from the other possible networks that could have been observed.

Consider a small-scale hypothetical example of a directed network defined on three nodes $\{A, B, C\}$, with an exogenous edge attribute (i.e., directed, dyadic covariate) X. Suppose X takes on one of three discrete values: low, moderate, and high. The distribution of possible graphs under a naive uniform assumption is shown in Fig. 1; there are 64 possible realizations of this three-node directed network. The probability of each, given in the center of the graph, is 1/64.

Suppose the starred network in Fig. 2 is observed. This network contains four out of six possible edges, one reciprocal edge, two edges corresponding to low covariate values, one edge corresponding to a moderate covariate value, and one edge corresponding to a high covariate value. We now specify an ERGM on this network including three statistics: the number of edges (E), the number of reciprocal edges (R), and the sum of the covariate values corresponding to the edges (S_X). The maximum likelihood estimates of the effects of these statistics, estimated from the starred network, are $\theta_E = 16.96$, $\theta_R = -14.61$, and $\theta_{S_x} = -1.07$. A negative (positive) parameter estimate means that the probability of observing a network with a higher value of the corresponding statistic relative to some hypothetical baseline network is lower (higher) than the probability of observing the baseline. Figure 2 gives the distribution over three-node directed networks implied by these ERGM parameters and statistics. The probability of having observed the starred networks is 0.06, almost four times as high as the uniform probability of 0.016.

This small-scale example shows a microcosm of the process of statistical inference on network data using ERGM. The null model (i.e., when $\theta = 0$) is a uniform distribution across the possible networks. A set of network characteristics are hypothesized to account for the configuration of the observed network—identifying the factors that collectively differentiate the observed network from the others that could have been observed. Parameters are then estimated that give the effect of network characteristics on the probability of observing that network. The distribution of networks can then be reweighted to account for the systematic structure discovered in the modeling process.

2.2 *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.

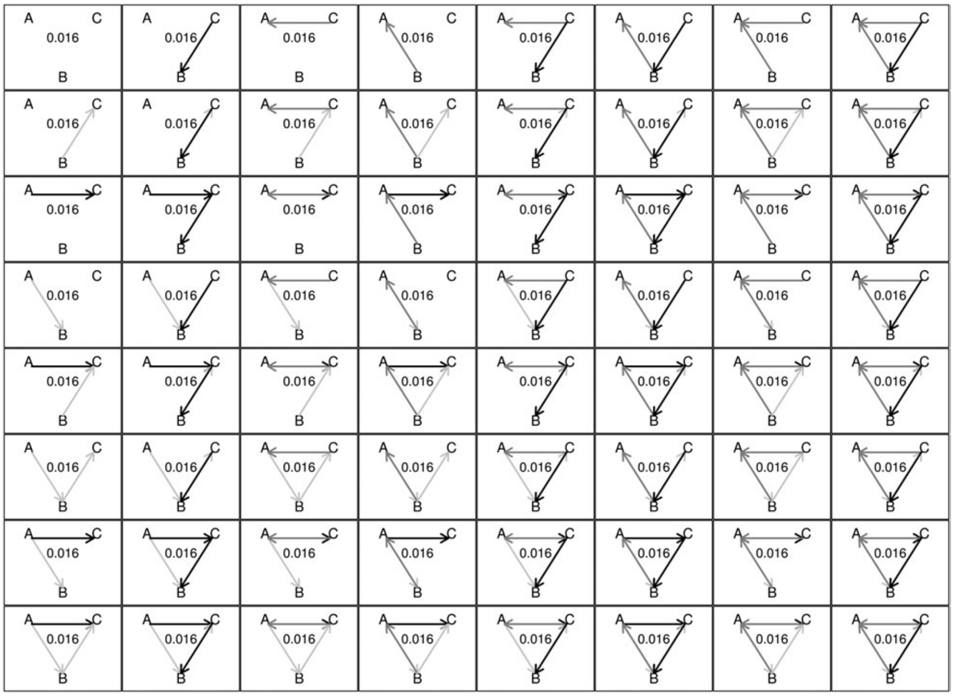

Fig. 1 The uniform distribution over a three-node directed network. The darker the edge, the higher the value of a hypothetical edge-wise covariate with three possible values. The number in each cell gives the probability of observing that configuration.

At the network level, the network statistics directly condition the probability of the entire network—allowing us to estimate the effect of within-network configurations (e.g., alignment of edges with covariates and/or reciprocal ties) on the predicted probability of observing a particular instance of the network. Specifically, the probability of observing (Y) is proportional to the exponent of the sum of network statistics, weighted by their respective parameters. If θ_i is positive (negative), then the effect of $\Gamma_i(Y)$ on the predicted probability of Y is positive (negative). It can be illustrative to compare the predicted probabilities of two possible realizations of the network, $Y^{(1)}$and $Y^{(2)}$. Suppose $Y^{(1)}$ and $Y^{(2)}$ are equal on every network statistic besides $\Gamma_i(Y)$. Let $\Gamma_i(Y^{(1)}) - \Gamma_i(Y^{(2)}) = \delta$, then $\frac{P\left(Y^{(1)}\right)}{P\left(Y^{(2)}\right)} = \exp\left\{\theta_i\delta\right\}$ (Desmarais 2010, 49). In other words, we are $\exp\{\theta_i\ \delta\}$ times as likely to observe $Y^{(1)}$ as we are $Y^{(2)}$.

Another popular network-level approach to interpretation of the ERGM is mechanically very similar to the method advocated by King et al. (2001). The large number of networks to be summed over in the denominator in equation (4) make it difficult to arrive at a closed-form expression for predicted quantities such as the average number of ties in the network or the distribution of ties about nodes (the degree distribution). 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. 2010). 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 form of a logistic regression. At the dyad level, the quantity of primary interest in most networks is the probability of the existence of a particular edge (e.g., how likely is it that there is war between the United States and Canada?). The probability of an edge between nodes i and j is written,

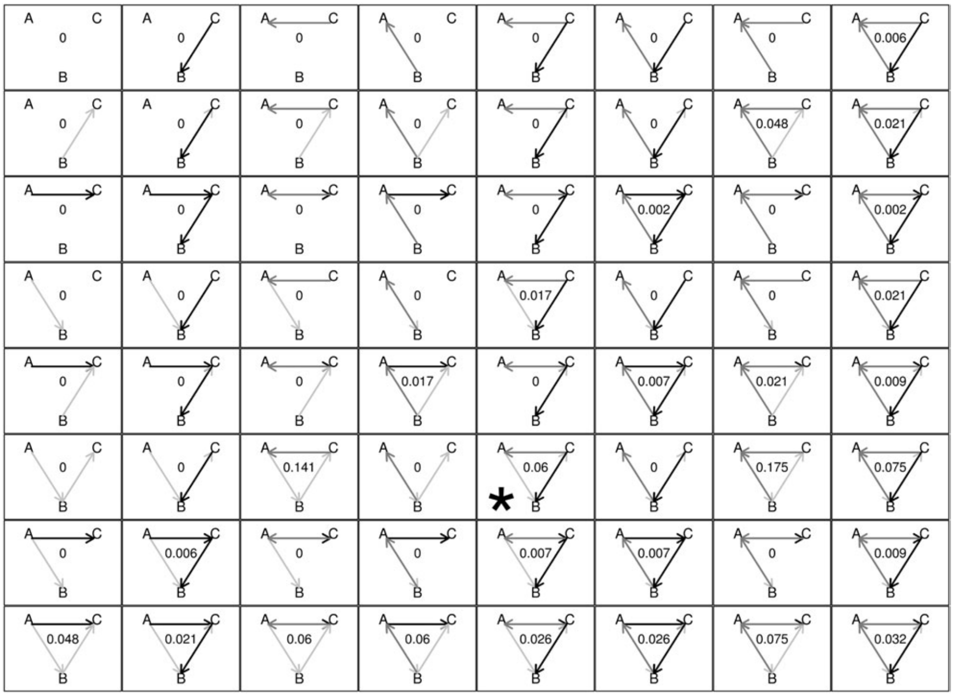

Fig. 2 The ERG distribution over a three-node directed network. The parameters were estimated by MLE on the starred network. The darker the edge, the higher the value of a hypothetical edge-wise covariate with three possible values. The number in each cell gives the probability of observing that configuration.

$$P(Y_{ij}=1|n,Y_{ij}^c)=\text{logit}^{-1}\left(\sum_{k=1}^{K}\theta_k\delta(\Gamma_{yk})\right), \tag{5}$$

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 (5), 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, $P(Y_{ij})$ is dependent on the dyad-wise outcomes of *every* other dyad.

It is important to note that a special case of equation (5) is equal to logistic regression. Consider a simple example where there are three network statistics in the model: the count of the number of edges in the network ($\Gamma_E\,(Y)$), a dyadic covariate ($\Gamma_{X_d}(Y)$), and reciprocity ($\Gamma_R\,(Y)$). In this model, the change statistics for the *ijth* dyad on these statistics are

$$\delta\big(\Gamma_E(Y)\big)=1,\delta\big(\Gamma_{X_d}(Y)\big)=X_{ij},\text{and}\,\delta\big(\Gamma_R(Y)\big)=Y_{ji},\ \text{such that,}$$

$$P(Y_{ij}=1|Y_{\text{ij}}^c)=\text{logit}^{-1}\Big(\theta_E+\theta X_dX_{ij}+\theta_RY_{ji}\Big).$$

If $\theta_R = 0$, then $P(Y_{ij} = 1|\ Y^c{}_{ij}) = \text{logit}^{-1}\,(\theta_E + \theta X_d\, X_{ij})$. Note that this is equal to the predicted probability in a logistic regression with θ_E equal to the intercept and θ_{Xd} the regression coefficient for X_d. Indeed, if the coefficients on the dependence terms—terms for which the change statistics involve other edges in Y—are all zero, the ERGM reduces to a logistic regression on the binary dyadic edge values, with the intercept equal to the edges coefficient and the regression coefficients equal to the exogenous covariate parameters (Strauss and Ikeda 1990; Wasserman and Pattison 1996).

2.3 *Limitations of ERGMs*

Though the ERGM is a powerful and flexible model, it suffers from a number of limitations that may hamper its utility to researchers. The major limitations of the ERGM are degeneracy problems, a sensitivity to missing data, an inability to model networks over time or networks with nonbinary 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, where we examine some extensions to the ERGM that can be useful for addressing these problems.

Degeneracy is an estimation problem associated with models that fit the data poorly. Essentially, degeneracy results from the specification of a model that 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 chain will move to an extreme graph of all or no edges, where it will stay, and the model is said to be degenerate.

Although degeneracy can be frustrating, it is not necessarily a major hindrance to applied research. Because models that 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 that is degenerate is sure not to fit well. Since including parameters or variables that 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" that 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.

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 (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 that 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 that recent work on inference with sampled networks is attempting to address Handcock and Gile (2010) but remains problematic from an applied perspective.

3 Estimation of the ERGM

Parameterized in the form of equation (4), 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, 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 (4), 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.[5] 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 Maximum Likelihood Estimation (MCMC-MLE) algorithm in Geyer and Thompson (1992), which is the algorithm used in most software packages for ERGM estimation, is given in Fig. 1 of our online supplement.

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. For a given edge ij, the edge-wise probability of a tie, given the rest of the network, (p_{ij}) is given in equation (5). The maximum pseudolikelihood is computed by using a hill-climbing algorithm to find the vector of parameters that maximize

$$\log\left(\prod_{i=2}^{n}\prod_{j=1}^{i} p_{ij}^{Y_{ij}}\left(1-p_{ij}\right)^{1-Y_{ij}}\right).$$

This is 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.

As two methods of approximation, there are pros and cons associated with each approach. The advantage of the 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 finite-sample 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 and Saul and Filkov 2007). 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.[6] More alarming for the task of hypothesis testing is the finding that the confidence intervals derived from the inverse 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.

[5]Most ERGM software comes complete with a suite of convergence diagnostics for MCMC estimation. See Handcock et al. (2010) for a detailed discussion.

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

4 Applications

Opportunities for analysis with the ERGM abound in political science. Indeed, data for which the unit of analysis is the relationship between actors can occur in any subfield of the discipline. The degree to which an application conforms to what we might call a ''basic'' ERGM setup, one in which a network with naturally binary relationships is observed in a single cross section, does much to inform how straightforward the ERGM analysis will be.

In some instances, the data will conform well to the basic ERGM setup. For example, in an application that lies at the intersection of comparative politics and interest group studies, Lusher and Ackland (2010) examine the structure of hyperlinks between the Web sites of 144 Australian groups that advocate for Asylum seekers. They use the ERGM to study the structure of this network, in which a directed edge exists from group A to group B if there is a link to group B on group A's Web site, at two cross sections.[7] Lusher and Ackland (2010) begin by developing a typology of advocacy groups, classifying them as either lobbying, research, or service-provision groups. They then specify an ERGM with exogenous terms to capture intratype homophily (the tendency to link to groups of the same type), type-sender effects, and type-receiver effects as well as terms to account for nine possible endogenous effects including, for instance, reciprocity, popularity, and cyclicality. Like our applications below, Lusher and Ackland (2010) estimate a dyadic logit model and an ERGM. They find that eight of the nine endogenous effects are statistically significant at the 0.05 level (two tailed) and that the substantive inferences regarding the exogenous sender and receiver effects differ between logit and the ERGM. Specifically, in the logit model, groups that lobby are found to receive more links than groups that do not lobby (but do not differ in their link-sending behavior), and in the ERGM, this finding is reversed, with groups that lobby found to send more links than those that do not (but do not differ on the tendency to receive links). This shows that (1) the ERGM can shed light on interesting dependencies within networks and (2) remove bias from the estimates of covariates that results from the inappropriate assumption that there is no endogeneity in the network.

Unfortunately, many applications in political science do not conform directly to the basicERGM setup. We now develop and consider two such applications. 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, the different subfields from which the applications are drawn demonstrate the broad applicability of ERGMs.

For each of the applications, we walk through the specification process and the considerations that 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 Valued Cross-Sectional Networks:*

4.1.1 Cosponsorship in the U.S. Congress

Our first application is drawn from American politics and extends ideas developed by Fowler (2006). In the U.S. Congress, legislators commonly cosponsor legislation sponsored by other legislators.[8] 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). Although 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;

[7]They impose the restriction that the ERGM parameters are the same in both periods. These parameters are estimated with the software LPNet: http://www.sna.unimelb.edu.au/pnet/pnet.html.
[8]Congressional rules dictate that there may be only one sponsor to a bill, and there are no joint sponsorships.

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).[9] This data set records the sponsorship and cosponsorship 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 nonnegative integer.

When undertaking the analysis, we are quickly presented with a technical problem. Within Congresses, the time interval 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, the 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.[10] 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. We can set thresholds of cosponsorship above which a tie is said to exist and, thus, coerce a count-valued network into a binary network that can be modeled with an ERGM. Both Fowler (2006) and Faust and Skvoretz (2002) address this problem by thresholding and treating any positive cosponsorship count greater than zero as a cosponsorship.[11] The second problem relates to the structure of the cosponsorship network specifically. As can be seen in Fig. 3, treating any cosponsorship act as an instance of a cosponsorship relationship would produce an extremely dense

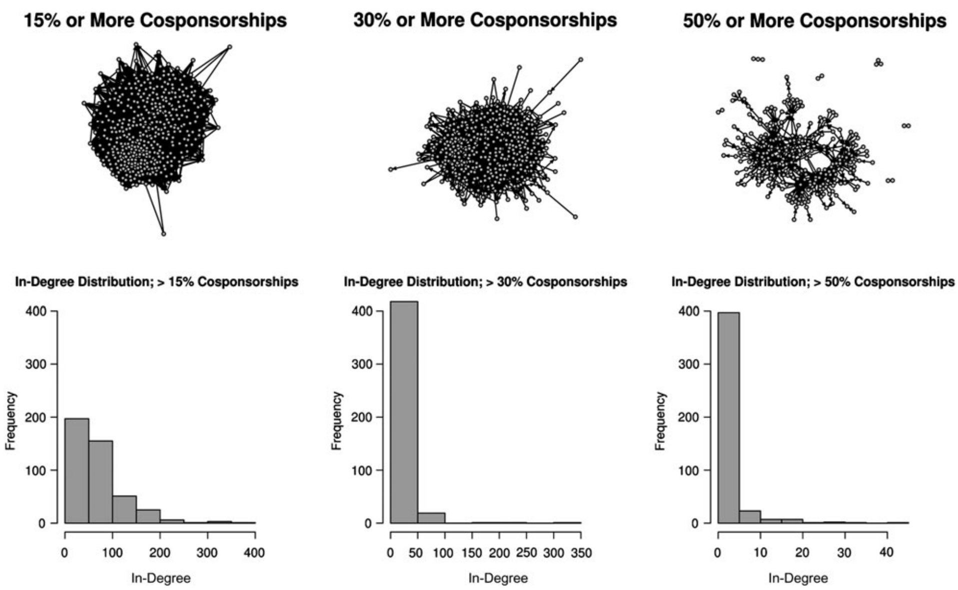

Fig. 3 This figure has six cells each showing the 108th House cosponsorship network at the same time t. The network is thinned by drawing a link from a legislator that cosponsors at least X% of the legislation sponsored by the receiver. The degree of thinning increases as increases from left to right (15%, 30%, and 50%, respectively). The top row depicts the network and the bottom row shows histograms of the in-degree of legislators (i.e., the in-degree distribution).

[9]Fowler's cosponsorship data are freely available from his Web site at http://jhfowler.ucsd.edu/cosponsorship.htm.

[10]The ERGM requires a finite and discrete number of possible networks. As can be seen in equation (4), the denominator sums over all possible networks. Because there is no upper bound for Poisson variables, a "Poisson ERGM" would require summation over an infinite space and is thus not possible.

[11]An alternative solution to this problem would be to treat the network as bipartite—meaning that two types of nodes (called modes in this context) exist—members of Congress are one type of mode and bills are another type of mode. This would eliminate the count-valued-edges problem because the unit of analysis would be changed from the legislator to the bill and, accordingly, there would be only one edge receiver and several edge-senders with no possible cosponsorship structure *between* them. This approach would lead to an analytic setup suitable for regression modeling. However, the majority of the cosponsorship literature treats the network as unipartite (as we do) because the structure of the unipartite network is substantively interesting (Fowler 2006; Zhang et al. 2007; Faust and Skvoretz 2002).

network. If we attempt to estimate an ERGM on this network, the parameters will produce a sample of networks where every possible tie exists and the MCMC sample will be degenerate as discussed above.

In order to transform the count-based ties into binary edges as well as produce a network that permits nondegenerate 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 that low threshold edges represent a fundamentally different type of edge than high threshold edges (in this case, that casual cosponsorship is fundamentally different from very frequent cosponsorship) and that fundamentally different processes are at work.

Our first step 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. We restrict our focus 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 substantively 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 $\{i \rightarrow k, i \rightarrow j, j \rightarrow k\}$.[12]

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 Bayesian Information Criteria (BIC) by more than 25%. The signs of all 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.0074 in the restricted model that does not account for network structure but rises to .64 when network parameters are included. This is a clear and consequential example of the faulty weight that 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

[12]We used the R package ERGM (Handcock et al. 2010) to fit the model.

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

	Full model			*Covariates-only model*		
Term	*Estimate*	*SE*	*p Value*	*Estimate*	*SE*	*p Value*
Edges	–2.8894	0.0155	0.0000	–0.6831	0.0301	0.0000
Same State	2.0864	0.0290	0.0000	1.8000	0.0249	0.0000
Nominate Distance	–1.0456	0.0269	0.0000	–1.4015	0.0327	0.0000
Same Party	0.0029	0.0062	0.6371	0.0688	0.0257	0.0074
Shared Committees	0.4822	0.0119	0.0000	0.3543	0.0111	0.0000
Asymmetric Ties	–0.1896	0.0075	0.0000	–	–	–
Cyclic Triples	–0.0300	0.0047	0.0000	–	–	–
Transitive Triples	0.0297	0.0016	0.0000	–	–	–
BIC	139,506			196,043		

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, in 23 of the 30 ties, the full model provides more accurate predictions.

4.2 *A Network over Time*

4.2.1 Conflict in the international system

Our second application is drawn from international relations; we seek to replicate and reanalyze 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 of the international context and implications of conflict decisions, dyadic independence is a problematic assumption. Few studies in international relations have specified inferential models of conflict as a network phenomenon; the article by Maoz et al. (2006) 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). Before going into greater detail regarding the specification of the model, we present an innovation in the estimation of ERGMs for longitudinally observed networks that we use to analyze the conflict network.

4.2.1.1 *ERGM for longitudinal networks.* 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 international system from 1870–1996, a total of 450,306 dyadic observations. The dependent variable is an indicator for the occurrence of at least one 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; ERGMs were developed to handle single networks and cannot be used to compute a single set of estimates that covers multiple pooled networks. Because the network of interest is observed longitudinally, we apply a longitudinal extension to the ERGM called the temporal Exponential Random Graph Model (TERGM), developed by Hanneke and Xing (2007) and extended by Desmarais and Cranmer (2010). One helpful quality of the TERGM for the purpose of comparison with dyadic logit is that the TERGM parameters (like those of ERGM discussed above) reduce to those of logistic regression if the coefficients on the network statistics are zero (i.e., dependence terms are either not included or have no effect). The parameters in this model are interpreted in the same way as ERGM parameters, with the multivariate observation (Y) being the network-year. The likelihood function for the TERGM, with T networks is the product over equation (4) evaluated on each of the T networks in the

Table 2 A selected House sociomatrix with conditional probabilities

	Pelosi	*Kucinich*	*Gephardt*	*Delay*	*Hastert*	*Pryce*
Pelosi		(0, 0.037, 0.308)[a]	(0, 0.16, 0.351)[a]	(0, 0.036, 0.135)[a]	(1, 0.056, 0.145)	(0, 0.264, 0.121)
Kucinich	(1, 0.95, 0.308)[a]		(0, 0.161, 0.308)[a]	(0, 0.023, 0.083)[a]	(0, 0.028, 0.089)[a]	(1, 0.736, 0.405)[a]
Gephardt	(1, 0.312, 0.351)	(0, 0.072, 0.308)[a]		(0, 0.019, 0.099)[a]	(0, 0.024, 0.107)[a]	(0, 0.043, 0.121)[a]
Delay	(0, 0.034, 0.135)[a]	(0, 0.011, 0.083)[a]	(0, 0.016, 0.099)[a]		(1, 0.15, 0.415)	(0, 0.087, 0.302)[a]
Hastert	(0, 0.052, 0.145)[a]	(0, 0.009, 0.089)[a]	(0, 0.014, 0.107)[a]	(0, 0.096, 0.415)[a]		(0, 0.071, 0.32)[a]
Pryce	(0, 0.2, 0.121)	(0, 0.106, 0.405)[a]	(0, 0.03, 0.121)[a]	(1, 0.076, 0.302)	(1, 0.132, 0.32)	

Note. Each entry gives (1) a binary indicator of a cosponsorship relationship, (2) a conditional probability of cosponsorship from the full model, and (3) a conditional probability of cosponsorship from the covariates-only model. The rows correspond to the cosponsor and the columns to the sponsor.

[a]Indicates the full model makes the better prediction.

series. Another extension is that $\Gamma(\cdot)$ can be a function of past networks (e.g., a lagged network as a dyadic covariate or the delayed reciprocation of edges).[13]

The choice we would typically have to make between maximum pseudolikelihood and MCMC to estimate the TERGM is nullified in this case: the computational burden associated with the use of MCMC-MLE is insurmountable given current technology and the fact that we have to estimate the model on 126 networks. We use maximum pseudolikelihood, which is shown by Strauss and Ikeda (1990), to produce consistent estimates of the parameters and implement a bootstrap resampling method (Efron 1981) for computing confidence intervals. The method we use is the nonparametric bootstrap, and the resampled unit is the network-year. For each iteration in each model, we resample with replacement 126 networks, find the maximum pseudolikelihood estimate for that sample, and store the TERGM. Desmarais and Cranmer (2010) show that this bootstrap scheme, when applied to the TERGM estimated by pseudolikelihood, provides bootstrapped percentile confidence intervals with coverage probabilities that are asymptotically equivalent to their nominal levels. The algorithm for bootstrapped pseudolikelihood estimation in the TERGM is given in Fig. 2 of our online supplement to this article. 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.

4.2.1.2 *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.[14] 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.[15] They find that *SEq* has a negative effect on the probability of conflict between any two states. We use the replication data by Maoz et al. and specify the same theoretical model they did: militarized interstate disputes 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 model by Maoz et al. (2006) 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 counterproductive 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.[16]

The results from the logistic regression and TERGM analysis of the MID network 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

[13]For a comparison of TERGM with other methods for longitudinal social network analysis, particularly that proposed by Snijders, van de Bunt, and Steglich (2010), see Desmarais and Cranmer (2010).

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

[15]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.

[16]We estimate each model with a lagged MID network as a covariate.

Table 3 Replication and extension of model 1 from Maoz et al. (2006)

	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]

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 to 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 or beyond. LDV indicates that a lagged dependent variable was included in the model specification.

covariates that motivate collections of states to 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 triadic effects, though highly intuitive, can be difficult to accommodate outside the ERGM framework, whereas ERGMs make the inclusion of such effects simple both theoretically and computationally. 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 TERGM 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 the conflict literature. This illustrates the inferential power and broad applicability of the TERGM. Yet one question we have yet to answer is whether the TERGM provides a better fit to the data than logistic regression. Given that we have used pseudolikelihood, and not maximum likelihood, we cannot assess model fit using 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 TERGM 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 TERGM 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 nonparametrically assess the performance of competing methods for classifying binary outcomes (Pete, Pattipati, and Kleinman 1993; Pepe 2000). We will use it to compare logit to ERGM in terms of 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 the R^2, the greater the area under the

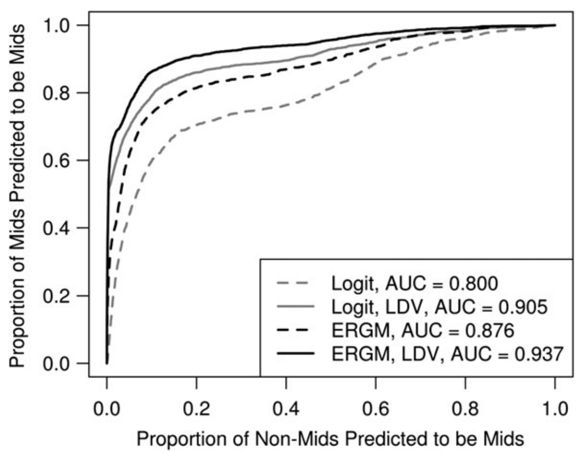

Fig. 4 ROC curves demonstrating the in-sample predictive performance of the models for MIDs. The area under the ROC curve (AUC) listed in the legend is a common nonparametric measure of predictive fit. The closer the AUC is to one, the closer the model is to perfect classification. LDV indicates that a lagged dependent variable was included in the model specification.

ROC curve, the better the predictive value of a model. If classification is perfect, the area under the ROC curve is 1.0 and if the classifier is never right, it 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 Fig. 4. The TERGM with a lagged network performs the best, the TERGM without a lagged network performs better than the original logistic regression model, and the logit with a lagged network performs better than the TERGM without a lagged network. As such, we have identified two important, and previously unverified, factors related to the prediction of conflict. First, there is memory in the conflict network, and it is thus important to account for the past behavior of a dyad. Second, network structure is nontrivially important. Whether a lagged network is or is not included, moving from the logistic to the ERGM modeling framework with network structure improves the fit of the model. This reinforces confidence in our 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 is not statistically significant.

5 Concluding Thoughts

Relational data, as they commonly occur in political science, can often be conceptualized as networks rather than sets of conditionally independent dyads. The ERGM is a means of conducting inference on network data. By accounting for the interdependence common in social networks, the ERGM eliminates the threats to valid inference induced by the conditional independence assumption associated with traditional regression models. Moreover, the ERGM permits a precise account of the forms of interdependence present in the network—allowing researchers to formulate and test hypotheses about the dependencies among ties in the network. When inappropriately applied to relational data, as we demonstrate in the empirical applications, i.i.d. regression models can cause explanatory power generated by systematic interdependence between relationships to be falsely attributed to the covariates.

We reviewed the basic concept of the ERGM, its strengths, limitations, and estimation. We then presented solutions to two of the ERGM's limitations: difficulties with non-binary edges 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 homophily characterizes the cosponsorship network. Then, in our international conflict application, we demonstrated that there is a popularity effect in conflict networks—meaning that there is positive correlation among states in their decisions to go to war with other states—and that there is little triangle formation in the conflict network, supporting the claim that "the enemy of your enemy is your friend." Also, once we account for the network structure underlying international conflict, there is less support for the operation of the democratic peace. We feel that our discussions and applications highlight the promise of network analysis in political science generally and the utility of the ERGM approach to inference specifically.

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, Chicago, IL.

Alvarez, R. Michael, and Jonathan Nagler. 1998. When politics and models collide: estimating models of multiparty elections. *American Journal of Political Science* 42:55–96.

Baybeck, Brady, and Robert Huckfeldt. 2002. Spatially dispersed ties among interdependent citizens: connecting individuals and aggregates. *Political Analysis* 10:261–75.

Besag, Julian. 1975. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician)* 24:179–95.

Boehmke, Frederick J. 2006. The influence of unobserved factors on position timing and content in the NAFTA vote. *Political Analysis* 14:421–38.

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* 51:301–27.

Crescenzi, Mark J. C. 2007. Reputation and interstate conflict. *American Journal of Political Science* 51:382–96.

Desmarais, Bruce A. 2010. Modeling interdependence in collective political decision making. PhD thesis University of North Carolina at Chapel Hill.

Desmarais, Bruce A., and Skyler J. Cranmer. 2010. *Analyzing longitudinal networks: the temporal exponential random graph model* In Midwest Political Science Association Annual Meeting, April 22, Chicago, IL.

Efron, Bradley. 1981. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* 68:589–99.

Faust, Katherine, and John Skvoretz. 2002. Comparing networks across space and time, size and species. *Sociological Methodology* 32:267–99.

Fienberg, Stephen E., and Stanley S. Wasserman. 1981. Categorical data analysis of single sociometric relations. *Sociological Methodology* 12:156–92.

Fowler, James H. 2006. Connecting the Congress: A study of cosponsorship networks. *Political Analysis* 14:456–87.

Frank, Ove, and David Strauss. 1986. Markov graphs. *Journal of the American Statistical Association* 81:832–42.

Franzese, Robert J. Jr., and Jude 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* 7:167–89.

———. 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis* 15:140–64.

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)* 54:657–99.

Goodreau, Steven M., James A. Kitts, and Martina Morris. 2009. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* 46:103–25.

Greene, William H. 2008. Econometric analysis. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Handcock, Mark S. 2003. Assessing degeneracy in statistical models of social networks. Center for Statistics and the Social Sciences, University of Washington, Working Paper No. 39.

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. Version 2.2–2*. Seattle, WA.Project home page at url http://statnetproject.org. http://CRAN.R-project.org/package=ergm

Handcock, Mark Stephen, and Krista Jennifer Gile. 2010. Modeling networks from sampled data. *Annals of Applied Statistics* 4:5–25.

Hanneke, Steve, and Eric P. Xing. 2007. Discrete temporal models of social networks. In *Statistical network analysis: Models, issues, and new directions.* Vol. 4503 of Lecture Notes in Computer Science, 115–125. Berlin, Germany: Springer.

Hays, Jude C., Aya Kachi, and Robert J. Franzese, Jr. 2010. A spatial model incorporating dynamic, endogenous network interdependence: A political science application. *Statistical Methodology* 7:406–28.

Hoag, Malcolm W. 1960. What interdependence for NATO? *World Politics* 12:369–90.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97:1090–8.

Hoff, Peter D., and Michael D. Ward. 2004. Modeling dependencies in international relations networks. *Political Analysis* 12:160–75.

Holland, Paul W., and Samuel Leinhardt. 1981. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* 76:33–50.

Hyvrinen, Aapo, Jarmo Hurri, and Patrik O. Hoyer. 2009. *Natural image statistics*. London: Springer.

Jackson, John E. 2002. A seemingly unrelated regression model for analyzing multiparty elections. *Political Analysis* 10:49–65.

Keohane, Robert O., and Joseph S. Nye. 1989. *Power and interdependence*. 2nd ed. Glenview, IL: Scott Foresman.

Kessler, Daniel, and Keith Krehbiel. 1996. Dynamics of cosponsorship. *The American Political Science Review* 90:555–66.

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* 95:49–69.

Kroll, John A. 1993. The complexity of interdependence. *International Studies Quarterly* 37:321–47.

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.

Lusher, Dean, and Robert Ackland. 2010. A relational hyperlink analysis of an online social movement. *Journal of Social Structure* 11. http://www.cmu.edu/joss/content/articles/volume11/Lusher/.

Maoz, Zeev. 2006. Network polarization, network interdependence, and international conflict, 1816–2002. *Journal of Peace Research* 43:391–411.

———. 2009. The effects of strategic and economic interdependence on international conflict across levels of analysis. *American Journal of Political Science* 53:223–40.

Maoz, Zeev, and Bruce Russett. 1993. Normative and structural causes of democratic peace, 1946–1986. *The American Political Science Review* 87:624–38.

Maoz, Zeev, Ranan D. Kuperman, Lesley Terris, and Ilan Talmud. 2006. Structural equivalence and international conflict: A social networks analysis. *The Journal of Conflict Resolution* 50:664–89.

Morris, Martina, Mark S. Handcock, and David R. Hunter. 2008. Specification of exponential-family random graph models: Terms and computational aspects. *Journal of Statistical Software* 24:1–24.

Park, Juyong, and M. E. J. Newman. 2004. Statistical mechanics of networks. *Physical Review E* 70:66117–30.

Pepe, Margaret Sullivan. 2000. An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* 56:352–9.

Pete, Andras, Krishna R. Pattipati, and David L. Kleinman. 1993. Optimal team and individual decision rules in uncertain dichotomous situations. *Public Choice* 75:205–30.

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* 29:192–215.

Rosecrance, R., A. Alexandroff, W. Koehler, J. Kroll, S. Laqueur, and J. Stocker. 1977. Whither interdependence? *International Organization* 31:425–71.

Rosecrance, Richard, and Arthur Stein. 1973. Interdependence: myth or reality. *World Politics* 26:1–27.

Saul, Zachary M., and Vladimir Filkov. 2007. Exploring biological network structure using exponential random graph models. *Bioinformatics* 23:2604–11.

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* 47:143–58.

Scholz, John T., and Cheng-Lung Wang. 2006. Cooptation or transformation? Local policy networks and federal regulatory enforcement. *American Journal of Political Science* 50:81–97.

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

Snijders, Tom A. B., Gerhard G. van de Bunt, and Christian E. G. Steglich. 2010. Introduction to stochastic actor-based models for network dynamics. *Social Networks* 32:44–60.

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* 85:204–12.

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* 31:52–62.

Ward, Michael D., and Peter D. Hoff. 2007. Persistent patterns of international commerce. *Journal of Peace Research* 44:157–75.

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* 51:583–601.

Wasserman, Stanley, and Philippa Pattison. 1996. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and *p**. *Psychometrika* 61:401–25.

Zhang, Yan, A. J. Friend, Amanda L. Traud, Mason A. Porter, James H. Fowler, and Peter J. Mucha. 2007. Community structure in Congressional cosponsorship networks. *Physica A: Statistical Mechanics and Its Applications* 387:1705–12.