

AMEN FOR LATENT SPACE MODELS

SHAHRYAR MINHAS, PETER D. HOFF, AND MICHAEL D. WARD

Interest in networks

Popular approach has been to employ latent space models
Variants.

Response to Cranmer et al. (2016).

Network analysis provides a way to represent and study “relational data”, that is data, with characteristics extending beyond those of the individual, or in the parlance of International Relations (IR), characteristics beyond the monadic. Data structures that extend beyond the monadic level are quite simply the norm when it comes to the study of events such as trade, interstate conflict, or the formation of international agreements. The dominant paradigm in international relations for dealing with such data structures, however, is not a network approach but rather a dyadic design, in which an interaction between a pair of countries is considered independent of interactions between any other pair in the system.¹

The implication of this assumption is that when, for example, Vietnam and the United States decide to form a trade agreement they make this decision independently of what they have done with other countries and what other countries in the international system have done among themselves.² An even stronger assumption is that Japan declaring war against the United States is independent of the decision of the United States going to war against Japan.³ A common refrain from those that favor the dyadic approach is that many events are not multilateral (Diehl and Wright, 2016), thus alleviating the need for an approach that incorporates interdependencies between observations. This is clearly wrong. The network perspective asserts that even the bilateral events and processes take place within a broader system. What takes place in one part of the system may be dependent upon another. At a minimum we don’t know whether independence of events and processes characterizes what we observe. At worst, we should examine this assertion.

¹To highlight the ubiquity of this approach the following represent just a sampling of the articles published from the 1980s to the present in the American Journal of Political Science (AJPS) and American Political Science Review (APSR) that assume dyadic independence: Dixon (1983); Mansfield, Milner and Rosendorff (2000); Lemke and Reed (2001); Mitchell (2002); Dafoe (2011); Fuhrmann and Sechser (2014); Carnegie (2014). See also XXXXX, iSQ special issue on networks.

²There has been plenty of work done on treaty formation that would challenge this claim, e.g., see Manger, Pickup and Snijders (2012); Kinne (2013).

³Maoz et al. (2006); Ward, Siverson and Cao (2007); Minhas, Hoff and Ward (2016) would each note the importance of taking into account network dynamics in the study of interstate conflict.

The potential for interdependence between observations poses a challenge to theoretical as well as statistical modeling since the assumption made by standard approaches used across the social sciences is that observations are, at least, conditionally independent (Snijders, 2011). The consequence of ignoring this assumption have been frequently noted within the political science literature already. For example, see Beck, Katz and Tucker (1998); Signorino (1999); Hoff and Ward (2004); Franzese and Hayes (2007); Cranmer and Desmarais (2011); Erikson, Pinto and Rader (2014). More relevant is the fact that a wealth of research from other disciplines suggests that carrying the independence assumption into a study with relational data is misguided and most often leads to biased inferences.⁴

Despite the hesitation among some in the discipline to adopt network analytic approaches, in recent years there has been a greater level of interest in understanding these approaches. For instance, in the past year special issues focused on the application of a variety of network approaches have come out in the *Journal of Peace Research* and *International Studies Quarterly*. Particularly notable is a recent overview and comparison of a handful of network based inferential models Cranmer et al. (2016).

Specifically, they focus on the exponential random graph model (ERGM), the multiple regression quadratic assignment procedure (MRQAP), and a latent distance approach developed by Hoff, Raftery and Handcock (2002). Their discussion around the differences among these approaches and their empirical comparison of them is valuable. At the same time, they overlook a decade worth of developments that latent variable model approaches have undergone. This is especially relevant in the context of providing an overview for the field by focusing on the results from one early attempt at a latent variable model, they end up overlooking a substantial amount of work using this

⁴From Computer Science see: Bonabeau (2002); Brandes and Erlebach (2005). From Economics see: Goyal (2012); Jackson (2014). From Psychology see: Pattison and Wasserman (1999); Kenny, Kashy and Cook (2006). From Statistics and Sociology see: Snijders (1996); Hoff, Raftery and Handcock (2002).

type of approach in political science.⁵ The principal latent variable approach used in political science has been the general bilinear mixed-effects (GBME) model developed by Hoff (2005). Examples of political science applications of the GBME include Hoff and Ward (2004); Ward, Siverson and Cao (2007); Cao (2009, 2010, 2012); Breunig, Cao and Luedtke (2012); Ward, Ahlquist and Rozenas (2012); Cao and Ward (2014); Metternich, Minhas and Ward (2015); Greenhill (2015), we are not aware of any political science applications using the latent distance approach.⁶ As Hoff (2008) shows both empirically and mathematically, the distinction between the latent distance and factor models, the next step of the GBME, is consequential when accounting for higher-order interdependencies, a point overlooked by Cranmer et al. (2016).

In this article, we introduce the additive and multiplicative effects network model (AMEN). To highlight the benefits of this approach we estimate this model using data from the application presented in Cranmer et al. (2016) and compare it to the other models presented in that article. By doing so we are able to show that AMEN provides a far superior goodness of fit to the data than alternative approaches.⁷ Further, through the AMEN approach we can estimate many different types of cross-sectional and longitudinal relational data structures (e.g., binomial, gaussian, and ordinal edges) in a straightforward way. The rest of this article proceeds as follows, we briefly motivate the need for network oriented approaches, introduce the AMEN modeling framework, compare it to previous implementations of latent variable approaches, and then end by showing how this approach fits the application presented in Cranmer et al. (2016).

We believe that this modeling framework can provide a flexible and easy to use scheme through which scholars can study relational data. It addresses the issue of

⁵Indeed, in so far as we can tell, no one in political science has actually employed the Euclidean approach they summarize.

⁶The code necessary to run the GBME has been available since 2004 at the following address: http://www.stat.washington.edu/people/pdhoff/Code/hoff_2005_jasa/.

⁷The AMEN approach has already been developed into a package named **amen** and is available on CRAN (Hoff et al., 2015).

interdependence while still allowing scholars to examine theories that may only be relevant in the monadic or dyadic level. Further at the network level it provides estimates of degree related effects, reciprocity, and provides a descriptive visualization of higher order dependencies such as homophily and stochastic equivalence.

1. ADDRESSING DEPENDENCIES IN DYADIC DATA

Relational, or dyadic, data structures provide measurements of how pairs of actors relate to one another. These structures encompass events of interest as diverse as the level of trade between i and j to the occurrence of an interstate conflict. The easiest way to organize such data is the directed dyadic design in which the unit of analysis is some set of n actors that have been paired together to form a dataset of z directed dyads. A tabular design such as this for a set of n actors, $\{i, j, k, l\}$ results in $n \times (n - 1)$ observations, as shown in Table 1.

Sender	Receiver	Event
i	j	y_{ij}
	k	y_{ik}
\vdots	l	y_{il}
j	i	y_{ji}
	k	y_{jk}
\vdots	l	y_{jl}
k	i	y_{ki}
	j	y_{kj}
\vdots	l	y_{kl}
l	i	y_{li}
	j	y_{lj}
\vdots	k	y_{lk}

Table 1. Structure of datasets used in canonical design.

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Table 2. Adjacency matrix representation of data in Table 1. Senders are represented in the rows and receivers the columns.

1.1. Limitations of the Standard Framework. When modeling these types of data structures, scholars typically employ a generalized linear model (GLM) estimated via

maximum-likelihood. This type of model can be expressed via a stochastic and systematic component (Pawitan, 2013). The stochastic component reflects our assumptions about the probability distribution from which the data is generated: $y_{ij} \stackrel{\text{iid}}{\sim} \mathcal{F}(\theta_{ij})$, where \mathcal{F} represents a probability distribution or mass function such as normal or binomial, and $\stackrel{\text{iid}}{\sim}$ represents the assumption that each dyad in our sample is independently drawn from that particular distribution. The systematic component characterizes the model for the parameters of that distribution and describes how θ_{ij} varies as a function of a set of nodal and dyadic covariates, \mathbf{X}_{ij} : $\theta_{ij} = \boldsymbol{\beta}^T \mathbf{X}_{ij}$. A fundamental assumption we make when applying this modeling technique is that given \mathbf{X}_{ij} and the parameters of our distribution, each of the dyadic observations is conditionally independent.

The import of this assumption becomes clearer in the process of estimating a GLM via maximum likelihood. After having chosen a set of covariates and specifying a distribution, we construct joint density function over all dyads.

$$(1) \quad \begin{aligned} Pr(y_{ij}, y_{ik}, \dots, y_{lk} | \theta_{ij}, \theta_{ik}, \dots, \theta_{lk}) &= \mathcal{F}(\theta_{ij}) \times \mathcal{F}(\theta_{ik}) \times \dots \times \mathcal{F}(\theta_{lk}) \\ Pr(\mathbf{Y} = (y_{ij}, y_{ik}, \dots, y_{lk}) | \boldsymbol{\theta} = (\theta_{ij}, \theta_{ik}, \dots, \theta_{lk})) &= \prod_{a=1}^z \mathcal{F}(\theta_a) \end{aligned}$$

We next convert the joint probability into a likelihood by assuming the observations are fixed but the distributional parameters, $\boldsymbol{\theta}$, are assumed to be random:

$$(2) \quad \begin{aligned} \mathcal{L}(\boldsymbol{\theta} | \mathbf{Y}) &= k(\mathbf{Y}) \times Pr(\mathbf{Y} | \boldsymbol{\theta}) \\ \mathcal{L}(\boldsymbol{\theta} | \mathbf{Y}) &= k(\mathbf{Y}) \times \prod_{a=1}^z \mathcal{F}(y_a | \theta_a) \\ \mathcal{L}(\boldsymbol{\theta} | \mathbf{Y}) &\propto \prod_{a=1}^z \mathcal{F}(y_a | \theta_a) \end{aligned}$$

This likelihood can be solved through maximization or numerical analysis. However, the important point to note is that the likelihood as defined above is only valid if we are able to make the assumption that, for example, y_{ij} is independent of y_{ji} and y_{ik} given the set of covariates we specified.⁸ Assuming that the dyad y_{ij} is independent of the dyad y_{ji} asserts that there is no level of reciprocity in a dataset, an assumption that in many cases would seem quite untenable.⁹ A harder problem to handle is the assumption that y_{ij} is independent of y_{ik} , the difficulty here follows from the possibility that i 's relationship with k is dependent on how i relates to j and how j relates to k , or more simply put the "enemy of my enemy [may be] my friend".

The presence of these types of interdependencies in relational data structures is exactly why we are interested in the, but complicates the a priori assumption of observational independence. Without this assumption the joint density function cannot be written in the way described above and a valid likelihood does not exist.¹⁰ Accordingly inferences drawn from misspecified models that ignore potential interdependencies between dyadic observations are likely to have a number of issues including biased estimates of the effect of independent variables, uncalibrated confidence intervals, and poor predictive performance. Just as important, however, is that by ignoring these interdependencies we ignore a potentially important part of the data generating process behind relational data structures, namely, network phenomena.

1.2. Social Relations Model: Additive Part of AMEN. The dependencies that tend to develop in relational data can be more easily understood when we move away from stacking dyads on top of one another and turn instead to adjacency matrices as shown

⁸The difficulties of applying the GLM framework to data structures that have structural interdependencies between observations is a problem that has long been recognized. Beck and Katz (1995), for example, detail the issues with pooling observations in time-series cross-section datasets. Ward and Gleditsch (2008) have done the same in the case of spatial dependence.

⁹For example, see Ward, Siverson and Cao (2007); Cranmer, Heinrich and Desmarais (2014); Dorff and Minhas (2016).

¹⁰This problem has been noted in works such as Lai (1995); Manger, Pickup and Snijders (2012); Kinne (2013).

in Table 2. Operationally, this type of data structure is represented as a $n \times n$ matrix, \mathbf{Y} , where the diagonals in the matrix are typically undefined.¹¹ The ij^{th} entry defines the relationship between i and j and can be continuous or discrete. If the matrix is undirected, the ji^{th} entry will equal the ij^{th} entry. For example, in undirected event data an event cannot be attributed to a specific sender or receiver rather it is just an indication of something that happened between a pair of countries. An example that commonly arises in the IR literature involves models of alliance relationships: two countries are allied. In directed matrices, the off-diagonal values are not symmetric and there is a clear sender and receiver as in the case of bi-lateral trade or multilateral aid.

A common type of structural interdependency that arises in relational data structures is “preferential attachment” (Barabási and Réka, 1999; Réka, Jeong and Barabási, 1999). This is typically categorized as a first-order, or nodal, dependency and represents the fact that we typically find significant heterogeneity in activity levels across nodes. The implication of this across-node heterogeneity is within-node heterogeneity of ties, meaning that values across a row, say $\{y_{ij}, y_{ik}, y_{iu}\}$, will be more similar to each other than other values in the adjacency matrix because each of these values has a common sender i . This type of dependency manifests in cases where sender i tends to be more active in the network than other senders. The emergence of this type of structure often occurs in relational datasets such as trade and conflict. In both these cases, there are a set of countries that tend to be more active than others. Similarly, while some actors may be more active in sending ties to others in the network, we might also observe that others are more popular targets, this would manifest in observations down a column, $\{y_{ji}, y_{ki}, y_{li}\}$, being more similar. Last, we might also find that actors who more likely to send ties in a network are also more likely to receive them, meaning that the row and column means of an adjacency matrix may be correlated. First-order dependencies

¹¹Most of the relational variables studied in political science do not involve events that countries can send to themselves.

are equally important to take into account in undirected relational structures, the only difference being that nodal heterogeneity will be equivalent across rows and columns. The presence of this type of heterogeneity in directed and undirected relational data structures leads to a violation of the conditional independence assumption underlying the models in our standard tool-kit.¹²

Another ubiquitous type of structural interdependency is reciprocity. This is a second-order, or dyadic, dependency relevant only to directed datasets, and asserts that values of y_{ij} and y_{ji} may be statistically dependent. In studies of social and economic behavior, direct reciprocity – the notion that actors learn to “respond in kind” to one another – is argued to be an essential component of behavior.¹³ More specifically, this concept actually has deep roots in political science theories of cooperation and the evolution of norms between states (Richardson, 1960; Choucri and North, 1972; Keohane, 1989; Rajmaira and Ward, 1990; Goldstein and Freeman, 1990, 1991; Ward and Rajmaira, 1992; Brandt, Colaresi and Freeman, 2008). The clearest example of the relevance of this dependency comes from the conflict literature, as we would expect that if, for instance, Iran behaved aggressively towards Saudi Arabia that this would induce Saudi Arabia to behave aggressively in return. The prevalence of these types of potential interactions within directed dyadic data structures also complicates the basic assumption of observational independence.

The relevance of modeling first- and second-order dependencies has long been recognized within some social sciences such as psychology. citetwarner:etal:1979 developed the social relational model (SRM), a type of ANOVA decomposition technique, that facilitates this undertaking.¹⁴ The SRM is of particular note as it provides the error

¹²It can lead to so-called power law dynamics, which has reinforced the popularity of the assumption of preferential attachment in network studies.

¹³For example, see Bolton, Brandts and Ockenfels (1998); Cox, Friedman and Gjerstad (2007).

¹⁴Dorff and Ward (2013) provide an introduction to this model and Dorff and Minhas (2016) apply this approach to studying reciprocal behavior in economic sanctions.

structure for the additive effects component of the AMEN framework that we introduce here. The goal of the SRM is to decompose the variance of observations in an adjacency matrix in terms of heterogeneity across row means (out-degree), heterogeneity along column means (in-degree), correlation between row and column means, and correlations within dyads. Wong (1982) and Li and Loken (2002) provide a random effects representation of the SRM:

$$\begin{aligned}
 y_{ij} &= \mu + e_{ij} \\
 e_{ij} &= a_i + b_j + \epsilon_{ij} \\
 (3) \quad \{(a_1, b_1), \dots, (a_n, b_n)\} &\stackrel{\text{iid}}{\sim} N(0, \Sigma_{ab}) \\
 \{(\epsilon_{ij}, \epsilon_{ji}) : i \neq j\} &\stackrel{\text{iid}}{\sim} N(0, \Sigma_\epsilon), \text{ where} \\
 \Sigma_{ab} &= \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \quad \Sigma_\epsilon = \sigma_\epsilon^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}
 \end{aligned}$$

The basic idea here is quite simple, μ provides a baseline measure of the density or sparsity of a network, and e_{ij} represents residual variation. We then decompose that residual variation into parts, namely, a row/sender effect (a_i), a column/receiver effect (b_j), and a within dyad effect (ϵ_{ij}). The row and column effects are modeled jointly to account for correlation in how active an actor is in sending and receiving ties. Heterogeneity in the row and column means is captured by σ_a^2 and σ_b^2 , respectively, and σ_{ab} describes the linear relationship between these two effects (i.e., whether actors who send [receive] a lot of ties also receive [send] a lot of ties). Beyond these first-order dependencies, variation across second-order dependencies is described by σ_ϵ^2 and a within dyad correlation, or reciprocity, parameter ρ .

Hoff (2005) shows that the SRM covariance structure described in Equation 3 can be incorporated into the systematic component of a GLM framework to produce a generalized (bi-)linear mixed effects model: $\beta^T \mathbf{X}_{ij} + a_i + b_j + \epsilon_{ij}$, where $\beta^T \mathbf{X}_{ij}$ accommodates the inclusion of dyadic, sender, and receiver covariates. This approach effectively incorporates row, column, and within-dyad dependence in way that is widely used and understood by applied researchers: a regression framework. Further this approach can be extended to handle a diversity of outcome distributions (e.g., binomial, ordinal, etc.). In the case of binary data this can be done by utilizing a latent variable representation of a probit regression model. Specifically, we model a latent variable, θ_{ij} , with a linear predictor and we model the error using the SRM from Equation 3: $\theta_{ij} = \beta^T \mathbf{X}_{ij} + e_{ij}$. Then we can simply utilize a threshold model linking θ_{ij} to our observed values of y_{ij} , in the case of a binomial outcome distribution the threshold model can be expressed as: $y_{ij} = I(\theta_{ij} > 0)$. It can also easily incorporate ordinal and rank ordered data.

1.3. Latent Factor Model: Multiplicative Part of AMEN. Missing from the framework provided by the SRM is an accounting of third order dependence patterns that can arise in relational data. The ubiquity of third order effects in relational datasets arises from the presence of some set of shared attributes between nodes that affects their probability of interacting with one another. For example, one finding from the gravity model of trade is that neighboring countries are more likely to trade with one another, in this case, the shared attribute is simply geographic proximity. A finding common in the political economy literature is that democracies are more likely to form trade agreements with one another, and the shared attribute here is a country's political system. Both geographic proximity and a country's political system are examples of homophily, which

captures the idea that the relationships between actors with similar characteristics in a network are likely to be stronger than nodes with varying characteristics.¹⁵

More generally, say that we have a binary network where actors tend to form ties to others based on some set of shared characteristics. This often leads to a network graph with a high number of “transitive triplets”, that is cases in which we have sets of actors $\{i, j, k\}$ each being linked to one another. The left-most plot in figure 1 provides a representation of a network that exhibits this type of pattern. Structures such as this can develop when the interactions between actors results from some set of shared attributes those actors may possess (e.g., they are neighbors of one another, part of an alliance agreement, share similar political systems). The relevant implication of this when it comes to conducting statistical inference is that—unless we are able to specify the list of exogenous variable that may explain homophily—the probability of j and k forming a tie is neither independent of the ties that already exist between those actors and i , nor higher than the probability that either of those actors might form a tie with another actor, l with whom they have no shared attributes.

Another third-order dependence pattern that cannot be accounted for in the additive effects framework discussed in the previous section is stochastic equivalence. A pair of actors i, j are stochastically equivalent if the probability of i relating to, and being related to, by every other actor is the same as the probability for j (Anderson, Wasserman and Faust, 1992). More simply put this refers to the idea that there will be groups of nodes in a network with similar relational patterns. The occurrence of a dependence pattern such as this is not uncommon in the social sciences. Manger, Pickup and Snijders (2012) posit and estimate a stochastic equivalence structure to explain the formation of preferential trade agreements (PTAs). Specifically, they suggest that PTA formation is related to differences in per capita income levels between countries. Countries falling

¹⁵Homophily can be used to explain the emergence of patterns such as transitivity (“a friend of a friend is a friend”) and balance (“an enemy of a friend is an enemy”). See Shalizi and Thomas (2011) for a more detailed discussion on the concept of homophily.

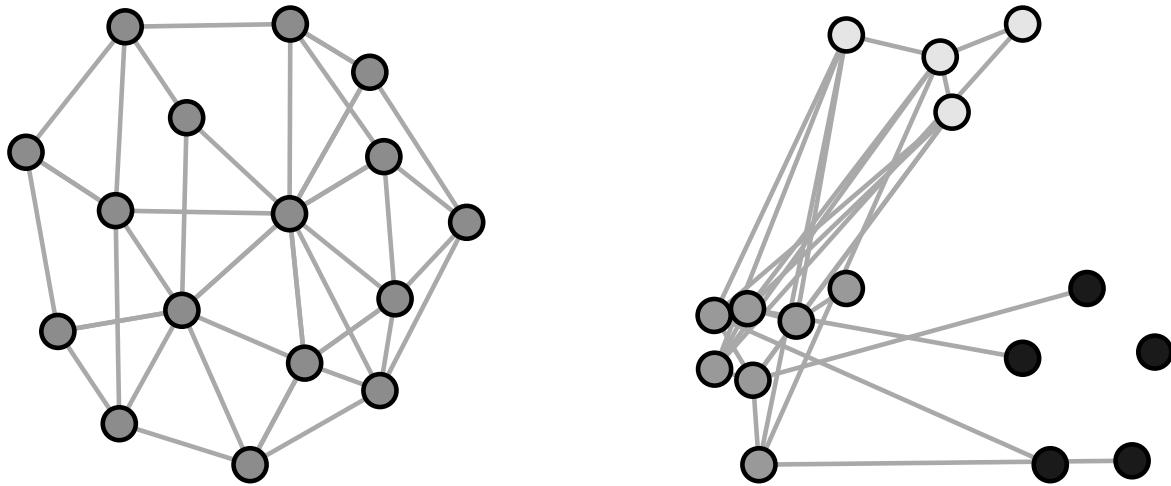


Figure 1. Graph on the left is a representation of an undirected network that exhibits a high degree of homophily, while on the left we show an undirected network that exhibits stochastic equivalence.

into high, middle, and low income per capita levels will have patterns of PTA formation that are determined by the groups into which they fall. They find that PTA formation occurs with greater probability in the following order high-middle, high-high, and middle-middle income groups, and that low income countries are rather unlikely to form PTAs with any partner. Such a structure is represented in the right-most panel of Figure 1, here the lightly shaded group of nodes at the top can represent high-income countries, nodes on the bottom-left middle-income, and the darkest shade of nodes low-income countries. The main point is that the behavior of actors in a network can at times be governed by group level dynamics, Failing to account for such patterns misleads in that it discounts or completely ignores an important part of the underlying data generating process that is the focus of inquiry.

If we are able to explicitly model the variety of shared attributes that might cause third-order dependence patterns to develop, then the additive effects described above justifies the conditional independence assumption that is central to the GLM framework we introduced earlier. The **amen** package even provides for the estimation of

that type of model using a Bayesian framework.¹⁶ There is also a set of utilities one can use to determine whether the inclusion of multiplicative effects is necessary, we will review these in the application section. In the context of most observational research, however, this assumption is untenable. The implausibility of this assumption is manifest.

1.3.1. ERGMs. Within political science the two most often used approaches to accounting for third order dependencies in relational data are ERGMs and latent space models. Exponential Random Graph Models were first developed by Erdős and Rényi (1959), but became more widely understood as they were applied to particular problems. Frank (1971) undertook an early examination and Julian Besag developed interesting applications and methods promoting their examination (Besag, 1977). But computing was complicated and expensive and it wasn't until Frank and Strauss (1986) and Wasserman and Pattison (1996) did these methods find widespread application. These so-called ERGM approaches are particularly useful when researchers are interested in the role that a specific list of network statistics have in giving rise to a certain network. These network statistics could include the number of transitive triads in a network, balanced triads, reciprocal pairs and so on.¹⁷ Having developed a set of network statistics, $S(\mathbf{Y})$, from a given network, \mathbf{Y} , the distribution of that network can be parameterized as:

$$(4) \quad Pr(Y = y) = \frac{\exp(\boldsymbol{\beta}^T S(y))}{\sum_{z \in \mathcal{Y}} \exp(\boldsymbol{\beta}^T S(z))}, y \in \mathcal{Y}$$

$\boldsymbol{\beta}$ represents a vector of model coefficients for the specified network statistics, \mathcal{Y} denotes the set of all obtainable networks, and the denominator is used as a normalizing factor (Hunter et al., 2008). This approach provides a way to state that the probability

¹⁶The main function in the **amen** package is titled “ame” and by default it runs a model assuming that no multiplicative effects are necessary.

¹⁷Morris, Handcock and Hunter (2008) and Snijders et al. (2006) provide a detailed list of network statistics that can be included in an ERGM model specification.

of observing a given network depends on the patterns that it exhibits, which are operationalized in the list of network statistics specified by the researcher. Within this approach one can test the role that a variety of network statistics play in giving rise to a particular network, additionally, researchers can easily accommodate nodal and dyadic covariates. Further because of the Hammersley-Clifford theorem any probability distribution over networks can be represented by the form shown in Equation 4 Hammersley and Clifford (1971).

One issue that arises when conducting statistical inference with this model is in the calculation of the normalizing factor, which is what ensures the expression above corresponds to a legitimate probability distribution. For even a trivially sized directed network that has only 20 actors, calculating the denominator means summing over $2^{20 \times (20-1)} = 2^{380}$ possible networks, or, to put it another way, more than the total number of atoms in the universe. One of the first approaches to deal with this issue was a computationally fast pseudo-likelihood approach developed by Strauss and Ikeda (1990), but that approach ignores the interdependent nature of observations in relational data structures, as a result, many have argued that the standard errors remain unreliable (Lubbers and Snijders, 2007; Robins et al., 2007; Van Duijn, Gile and Handcock, 2009). Additionally, there is no asymptotic theory underlying this approach on which to base the construction of confidence intervals and hypothesis tests (Kolaczyk, 2009). The pseudo-likelihood approach has became increasingly unpopular in recent years among those in the network analysis community, particularly, as simulation based techniques have developed. But it has not disappeared. One favored approach in the literature is to approximate the MLE using Markov Chain Monte Carlo techniques, also referred to as MCMC-MLE (Geyer and Thompson, 1992; Snijders, 2002; Handcock, 2003). MCMC-MLE is based on a stochastic approximation of the log-likelihood and a maximization of the approximation, the **ergm** R package developed by Hunter et al. (2008) provides for the estimation of this type of model.

The MCMC-MLE approach is certainly an advancement but notable problems remain. Bhamidi, Bresler and Sly (2008) and Chatterjee and Diaconis (2013) have shown that MCMC procedures can take an exponential time to converge for broad classes of ERGMs unless the dyadic observations are independent. This is a result of the fact that MCMC procedures can still visit an infinitesimally small portion of the set of possible graphs, \mathcal{Y} . A related issue when estimating ERGMs is that the model can become degenerate for certain combinations of estimated parameters. This means that the model is placing a large amount of probability on a small subset of networks that fall in the set of obtainable networks, \mathcal{Y} , but share little resemblance with the observed network, \mathbf{Y} (Schweinberger, 2011). Some have noted that model degeneracy is simply a result of model misspecification (Handcock, 2003; Goodreau et al., 2008; Handcock et al., 2008). This points to an important caveat in interpreting the implications of the Hammersley-Clifford theorem. Though this theorem ensures that any network can be represented through an ERGM, it says nothing about the complexity of the sufficient statistics, $S(y)$, required to do so. Failure to properly account for higher order dependence structures through an appropriate specification can at best lead to model degeneracy, which provides an obvious indication that the specification needs to be altered, and at worst deliver a result that converges but does not appropriately capture the interdependencies in the network. The consequence of the latter case is a set of inferences that will continue to be biased as a result of unmeasured heterogeneity, thus defeating the major motivation for pursuing an inferential network model in the first place.

A recent handbook on using network approaches to research political issues is found in Victor, Montgomery and Lubell (2016). Recent research that uses exponential random graph models specifically includes Victor and Ringe (2009), Berardo and Scholz (2010), Calvo and Leiras (2012), Lubell, Berardo and Robbins (2012), Robbins, Lewis and Wang (2012), Alemán and Calvo (2013), Heaney (2014), and Kirkland and Williams (2014).

1.3.2. **Latent Space Models.** Given the computational and inferential difficulties that go along with utilizing ERGMs, an alternative approach has been utilized by political scientists in studying relational data with third order dependence patterns. This approach is referred to as latent space models.¹⁸

The utilization of latent space models for network analysis is quickly becoming a popular approach for modeling relational data in a variety of fields as diverse as computer science to the social sciences. One reason for their increased usage is that they enable researchers to capture and visualize third-order dependencies in a way that other approaches are not able to replicate. Additionally, the conditional independence assumption that these models are able to provide implies that model degeneracy is not an issue, facilitating the testing of a variety of nodal and dyadic level theories, and providing a range of computational advantages (Hunter, Krivitsky and Schweinberger, 2012).

Three major latent space approaches have been developed to attempt to handle third order dependencies in relational data: latent class model, latent distance model, and the latent factor model. Each of these approaches can be incorporated into an undirected version of the framework that we have been constructing through the inclusion of an additional term, $\alpha(\mu_i, \mu_j)$, that captures latent third order characteristics of a network. General definitions for how $\alpha(\mu_i, \mu_j)$ is defined for these latent space models are shown in Equations 5. One other point of note about each of these latent space approaches is that researchers have to specify a value for K . In the case of the latent distance and factor models, a value of K equal to two or three is typically large enough to account for third order dependencies in relational data. In the next section, we will discuss a set of diagnostic that help researchers to make this choice.

¹⁸We are not aware of any one that has used the Euclidean approach, but the bi-linear latent space approach has been used by a variety of scholars including Hoff and Ward (2004), Ward, Siverson and Cao (2007), Ward and Hoff (2007), Cao (2009), Breunig, Cao and Luedtke (2012), Cao (2012), Ward, Ahlquist and Rozenas (2012), Ward and Hoff (2008), Metternich, Minhas and Ward (2015), and Cao and Ward (2014).

Latent class model

$$\alpha(\mu_i, \mu_j) = m_{\mu_i, \mu_j}$$

$$\mu_i \in \{1, \dots, K\}, i \in \{1, \dots, n\}$$

M a $K \times K$ symmetric matrix

Latent distance model

(5)
$$\alpha(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j) = -|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j|$$

$$\boldsymbol{\mu}_i \in \mathbb{R}^K, i \in \{1, \dots, n\}$$

Latent factor model

$$\alpha(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j) = \boldsymbol{\mu}_i^T \Lambda \boldsymbol{\mu}_j$$

$$\boldsymbol{\mu}_i \in \mathbb{R}^K, i \in \{1, \dots, n\}$$

Λ a $K \times K$ diagonal matrix

In the latent class model, also referred to as the stochastic block model, each node i is a member of some unknown latent class, $\mu_i \in (1, \dots, K)$. A probability distribution is used to describe the relationships between classes (Nowicki and Snijders, 2001), the implication of this is that the probability of a tie between i and j is purely a function of the classes to which they have been assigned. The placement of nodes into classes is done in such a way that actors in the same class are stochastically equivalent, meaning that the probability distribution for the relations that i has are the same as the relations that j has if i and j are in the same class. Given that the probability of a tie between a pair of actors is wholly dependent upon the class to which they belong, nodes in the same class may have small or high probability of ties. A graph such as the one depicted in the left panel of Figure 1 cannot be represented adequately through this

type of approach. To do so, would require a large number of classes, K , that would not be particularly cohesive or distinguishable from one another.¹⁹

A latent space approach that can characterize homophily is the latent distance model developed by Hoff, Raftery and Handcock (2002). In this approach, each node i has some unknown latent position in K dimensional space, $\mu_i \in \mathbb{R}^K$, and the probability of a tie between a pair ij is a function of the negative Euclidean distance between them: $-|\mu_i - \mu_j|$. Hoff, Raftery and Handcock (2002) show that because latent distances for a triple of actors obey the triangle inequality, this formulation models the tendencies toward homophily commonly found in social networks. This approach has been operationalized in the **latentnet** package developed by Krivitsky and Handcock (2015). However, this approach also comes with an important shortcoming that leads it to confound stochastic equivalence and homophily. Consider two nodes i and j that are proximate to one another in K dimensional Euclidean space, this suggests not only that $|\mu_i - \mu_j|$ is small but also that $|\mu_i - \mu_l| \approx |\mu_j - \mu_l|$, the result being that nodes i and j will by construction assumed to possess the same relational patterns with other actors such as l (i.e., that they are stochastically equivalent).²⁰

Now the last approach that we introduce here is similar to the dominant method used in political science and that is the latent factor model. An early iteration of this approach was presented in Hoff (2005) and introduced to political science by Hoff and Ward (2004), but this approach is motivated by an eigenvalue decomposition of a network.²¹ The motivation for this alternative framework stems from the fact that many real networks exhibit varying degrees of stochastic equivalence and homophily. In

¹⁹At the same time it is important to note that the characteristics of the latent class model make it ideal for other inferential goals such as community detection.

²⁰Hoff (2008) shows that the only way to account for a network that exhibits stochastic equivalence through a latent distance model is by setting the number of latent dimensions, K , to be on the order of the class membership size.

²¹An important difference in the earlier approaches such as the GBME compared to the model that we present here is that Λ was taken to be the identity matrix. This approach should also not be confused with the projection model introduced in Hoff, Raftery and Handcock (2002).

these situations, using either the latent distance or class model would end up only representing only a part of the network structure. In the latent factor model, each actor has an unobserved vector of characteristics, $\mu_i = \{\mu_{i,1}, \dots, \mu_{i,K}\}$, which describe their behavior as an actor in the network. The probability of a tie from i to j depends on the extent to which μ_i and μ_j are “similar” (i.e., point in the same direction) and on whether the entries of Λ are greater than or less than zero.

More specifically, the similarity in the latent factors, $\mu_i \approx \mu_j$, corresponds to how stochastically equivalent a pair of actors are and whether or not there is a positive association determines whether the network exhibits positive or negative homophily. For example, say that we estimate a rank-one latent factor model (i.e., $K = 1$), in this case μ_i is represented by a scalar $\mu_{i,1}$, similarly, $\mu_j = \mu_{j,1}$, and Λ will have just one diagonal element λ_k . The average effect this will have on y_{ij} is simply $\lambda_k \times \mu_i \times \mu_j$, where a value of $\lambda_k > 0$ indicates homophily and $\lambda_k < 0$ anti-homophily. Hoff (2008) shows that such a model can represent both homophily and stochastic equivalence, and that the alternative latent space approaches can be represented as a latent factor model but not vice versa. In the directed version of this approach, we use the singular value decomposition,²² here actors in the network have a vector of latent characteristics to describe their behavior as a sender, denoted by μ , and as a receiver, \mathbf{v} : $\mu_i, \mathbf{v}_j \in \mathbb{R}^K$ (Hoff, 2009). These again can alter the probability, or in the continuous case value, of an interaction between ij additively: $\mu_i^T \mathbf{D} \mathbf{v}_j$, where \mathbf{D} is an $n \times n$ diagonal matrix.

The latent factor model is incorporated into the AMEN approach as a multiplicative effect to account for third order dependencies (Hoff, 2009; Hoff et al., 2015). As stated in the beginning of this section incorporating any of these approaches into the additive effects probit framework is possible through the addition of a term that captures third

²²The singular value decomposition is a model based analogue to the eigenvalue decomposition for directed networks.

order interdependencies. In the **latentnet** package this is done by directly incorporating $|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j|$ as follows: $\theta_{ij} = \boldsymbol{\beta}^T \mathbf{X}_{ij} - |\boldsymbol{\mu}_i - \boldsymbol{\mu}_j|$.²³ However, incorporating the term in this way can affect our estimation of the linear relationship between the exogenous nodal and dyadic covariates. This results from collinearity between that set of exogenous attributes and the nodal positions of actors in the latent space. The intuition behind why collinearity occurs is not surprising given our discussion above, the latent space is essentially used to capture dependencies that can result from shared attributes between nodes. Thus if a particular exogenous covariate is actually predictive of relations between ij , due to homophily, this effect will be correlated with the nodal positions of actors in a K dimensional Euclidean space.

One motivation for pursuing network based approaches is that there are a variety of dependencies between observations in relational data and not accounting for those effects leads to biased parameter estimates and uncalibrated confidence intervals. The estimation procedure taken in the **latentnet** package, though extremely useful for understanding and visualizing some third order dependence patterns in relational data, does not address that motivation. Thus the AMEN approach considers the regression model shown in Equation 6:

²³The **latentnet** package also allows for the specification of a bilinear latent space that is closely related to the projection model introduced in Hoff, Raftery and Handcock (2002). This approach, however, is not equivalent to the latent factor approach used in AMEN, both the calculation of nodal positions and general estimation procedure are distinct.

$$y_{ij} = g(\theta_{ij})$$

$$\theta_{ij} = \beta^T \mathbf{X}_{ij} + e_{ij}$$

$$(6) \quad e_{ij} = a_i + b_j + \epsilon_{ij} + \alpha_{\mu_i, \mathbf{v}_j}, \text{ where}$$

$$\alpha_{\mu_i, \mathbf{v}_j} = \boldsymbol{\mu}_i^T \mathbf{D} \mathbf{v}_j = \sum_{k \in K} d_k \mu_{ik} v_{jk}$$

Using this framework, we are able to model the dyadic observations as conditionally independent given θ , while θ depends on the unobserved random effects, e. e is then modeled to account for the potential first, second, and third order dependencies that we have discussed. As described in Equation 3, $a_i + b_j + \epsilon_{ij}$, are the additive random effects in this framework and can capture network covariance through accounting for sender, receiver, and within-dyad dependence. A Bayesian estimation procedure using Gibbs sampling is available in the **amen** package to estimate this type of generalized linear mixed effects model from normal, binomial, ordered probit, and other types of distributions. The quantities to be estimated in this model from the observed data, $\{\mathbf{Y}, \mathbf{X}\}$, are:

- θ : Latent Gaussian variables
- β : Nodal and/or dyadic regression coefficients
- $\{(a_i, b_i)\} \in \{i = 1, \dots, n\}$: Nodal random effects
- $\Sigma_{ab}, \Sigma_\epsilon$: Network covariance

To arrive at posterior values for these parameters we iteratively simulate from their full conditional distributions:²⁴

- $\theta \sim p(\theta | \mathbf{Y}, \mathbf{X}, \beta, \mathbf{a}, \mathbf{b}, \Sigma_\epsilon)$

²⁴Further details on this process can be found in Hoff (2005).

- $\beta \sim p(\beta | \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \Sigma_\epsilon)$
- $\mathbf{a}, \mathbf{b} \sim p(\mathbf{a}, \mathbf{b} | \mathbf{X}, \boldsymbol{\theta}, \beta, \Sigma_{ab}, \Sigma_\epsilon)$
- $\Sigma_\epsilon \sim p(\Sigma_\epsilon | \mathbf{X}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b})$
- $\Sigma_{ab} \sim p(\Sigma_{ab} | \mathbf{a}, \mathbf{b})$

In describing the estimation approach for the multiplicative effects that are used to capture higher-order dependence it is useful to rewrite the directed version of the latent factor model as: $\mathbf{M} = \mathbf{U}^T \mathbf{D} \mathbf{V}$.²⁵ Here \mathbf{M} represents systematic patterns left over in $\boldsymbol{\theta}$ after accounting for any known covariate information and these patterns are being approximated through a model based singular value decomposition (Hoff, 2009). Thus the third order interdependencies captured in the latent factor space of AMEN are those that could not have been explained by the exogenous nodal and dyadic covariates that have already been included in the model. \mathbf{U} and \mathbf{V} are $n \times K$ matrices with orthonormal columns and $\mathbf{D} = \text{diag}\{d_1, \dots, d_K\}$. An MCMC scheme that can be used to construct an empirical distribution around these parameters involves iterating through $k \in 1, \dots, K$:

- $\mathbf{U}_{[,k]} \sim p(\mathbf{U}_{[k]} | \boldsymbol{\theta}, \mathbf{U}_{[,-k]}, \mathbf{D}, \mathbf{V})$
- $\mathbf{V}_{[,k]} \sim p(\mathbf{V}_{[k]} | \boldsymbol{\theta}, \mathbf{U}, \mathbf{D}, \mathbf{V}_{[,-k]})$
- $\mathbf{D}_{[k,k]} \sim p(\mathbf{D}_{[k,k]} | \boldsymbol{\theta}, \mathbf{U}, \mathbf{D}_{[-k,-k]}, \mathbf{V})$

Taken together the additive effects portion of AMEN (described by the SRM) and the multiplicative effects (described by the latent factor model) provide a modeling framework similar to the GLMs that many scholars currently use, and has the benefit of being able to not only deal with interdependencies in relational data but also provide explicit measurements of these dependencies after having taken into account observable information. Specifically, we can obtain degree based effects for actors in the network,

²⁵Framing the problem of accounting for third order interdependencies in this way actually provides a strong motivation for estimating relational data through the type of random effects approach that we are introducing here. See Hoff (2009) for a longer discussion on this topic.

the level of reciprocity between actors, and also visualize the third order interdependencies that remain in the data. This latter point is important to note as effectively using these visualizations may also help users of this approach to determine whether or not the inclusion of some other dyadic or nodal variable is necessary to accounting for patterns such as homophily or stochastic equivalence. In the following section we implement this approach to an application chosen by Cranmer et al. (2016) to highlight the benefits it provides over alternatives such as ERGM and the latent distance model.

2. COMPARISON WITH OTHER APPROACHES

Ingold (2008)

Ingold and Fischer (2014)

Ingold and Leifeld (2014)

Figure 2 highlights how we can expect high levels of sender and receiver heterogeneity.

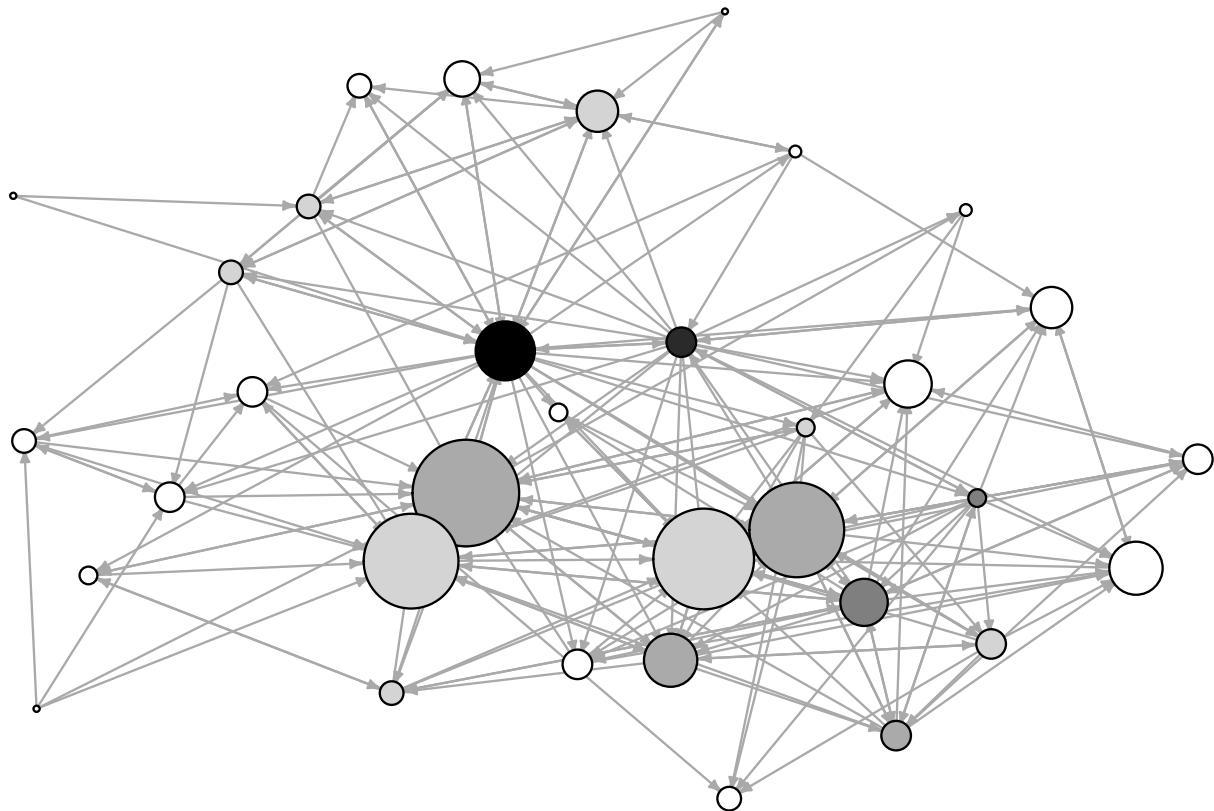


Figure 2. dv net

Krivitsky and Handcock (2015)

2.1. Parameter Estimates.

	Logit	MRQAP	LSM	ERGM	AME
Intercept/Edges	-4.44* (0.34)	-4.24* 	0.94* [0.09; 1.82]	-12.17* (1.40)	-3.39* [-4.38; -2.50]
Conflicting policy preferences					
Business vs. NGO	-0.86 (0.46)	-0.87* 	-1.37* [-2.42; -0.41]	-1.11* (0.51)	-1.37* [-2.44; -0.47]
Opposition/alliance	1.21* (0.20)	1.14* 	0.00 [-0.40; 0.39]	1.22* (0.20)	1.08* [0.72; 1.47]
Preference dissimilarity	-0.07 (0.37)	-0.60 	-1.76* [-2.62; -0.90]	-0.44 (0.39)	-0.79* [-1.55; -0.08]
Transaction costs					
Joint forum participation	0.88* (0.27)	0.75* 	1.51* [0.86; 2.17]	0.90* (0.28)	0.92* [0.40; 1.47]
Influence					
Influence attribution	1.20* (0.22)	1.29* 	0.08 [-0.40; 0.55]	1.00* (0.21)	1.09* [0.69; 1.53]
Alter's influence indegree	0.10* (0.02)	0.11* 	0.01 [-0.03; 0.04]	0.21* (0.04)	0.11* [0.07; 0.15]
Influence absolute diff.	-0.03* (0.02)	-0.06* 	0.04 [-0.01; 0.09]	-0.05* (0.01)	-0.07* [-0.11; -0.03]
Alter = Government actor	0.63* (0.25)	0.68 	-0.46 [-1.08; 0.14]	1.04* (0.34)	0.55 [-0.07; 1.15]
Functional requirements					
Ego = Environmental NGO	0.88* (0.26)	0.99 	-0.60 [-1.32; 0.09]	0.79* (0.17)	0.67 [-0.38; 1.71]
Same actor type	0.74* (0.22)	1.12* 	1.17* [0.63; 1.71]	0.99* (0.23)	1.04* [0.63; 1.50]
Endogenous dependencies					
Mutuality	1.22* (0.21)	1.00* 		0.81* (0.25)	
Outdegree popularity				0.95* (0.09)	
Twopath				-0.04* (0.02)	
GWIdegree (2.0)				3.42* (1.47)	
GWESP (1.0)				0.58* (0.16)	
GWODEGEE (0.5)				8.42* (2.11)	

Table 3. * p < 0.05 (or outside the 95% confidence interval).

2.2. Tie Formation Prediction. The results are displayed in Figure 3 using separation plots and Receiver Operating Characteristic (ROC) curves.

We compare the sensitivity and specificity trade-off for each model using ROC curves. Models that have a better fit according to this test should have curves that follow the left-hand border and then the top border of the ROC space. Here again it is apparent that accounting for the interstate relations and the endogenous network effects leads to noticeable improvements in performance. Last, by calculating the area under the ROC curve (AUC) we can assess the accuracy of each model.

Separation plots provide a visual interpretation of model fit by plotting all observations, in this case country pairs, in the data set according to their predicted value from left (low values) to right (high values). Models with a good fit should have all actual (dark blue) observations towards the right of the separation plot (Greenhill, Ward and Sacks, 2011).

In addition, we also highlight the difference in performance through the utilization of a precision-recall curve. Precision is a measure of result relevancy, while recall is a measure of how many truly relevant results are returned. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate. High scores for both show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall).

Precision is defined as the number of true positives over the number of true positives plus the number of false positives.

Recall is defined as the number of true positives over the number of true positives plus the number of false negatives.

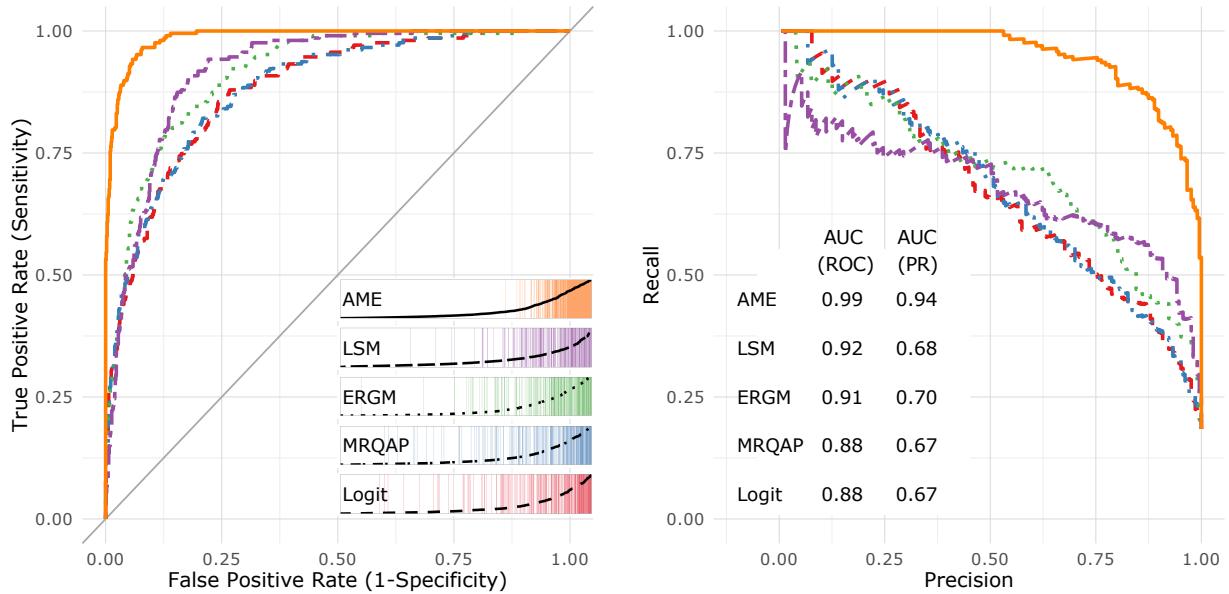


Figure 3. ROC and separation plots

2.3. Capturing Network Attributes. To assess whether the model adequately captures the network parameters of the DV. Here we compare the observed with a set of simulated networks based on certain network statistics (Hunter et al., 2008).

See Morris, Handcock and Hunter (2008) for details on each of these parameters.

- Dyad-wise shared partners - Number of dyads in the network with exactly i shared partners
- Edge-wise shared partners - Similar to above except this counts the number of dyads with the same number of edges
- Geodesic distances - The proportion of pairs of nodes whose shortest connecting path is of length k , for $k = 1, 2, \dots$. Also, pairs of nodes that are not connected are classified as $k = \infty$.
- Incoming k-star - Propensities for individuals to have connections with multiple network partners
- Indegree - degree count is the number of nodes with the same value of the attribute as the ego node

- Outdegree - degree count is the number of nodes with the same value of the attribute as the ego node

Figure 5 give posterior predictive goodness of fit summaries for four network statistics: (1) the empirical standard deviation of the row means; (2) the empirical standard deviation of the column means (heterogeneity of nodes with incoming activity); (3) the empirical within-dyad correlation; (4) a normalized measure of triadic dependence (Hoff et al., 2015).

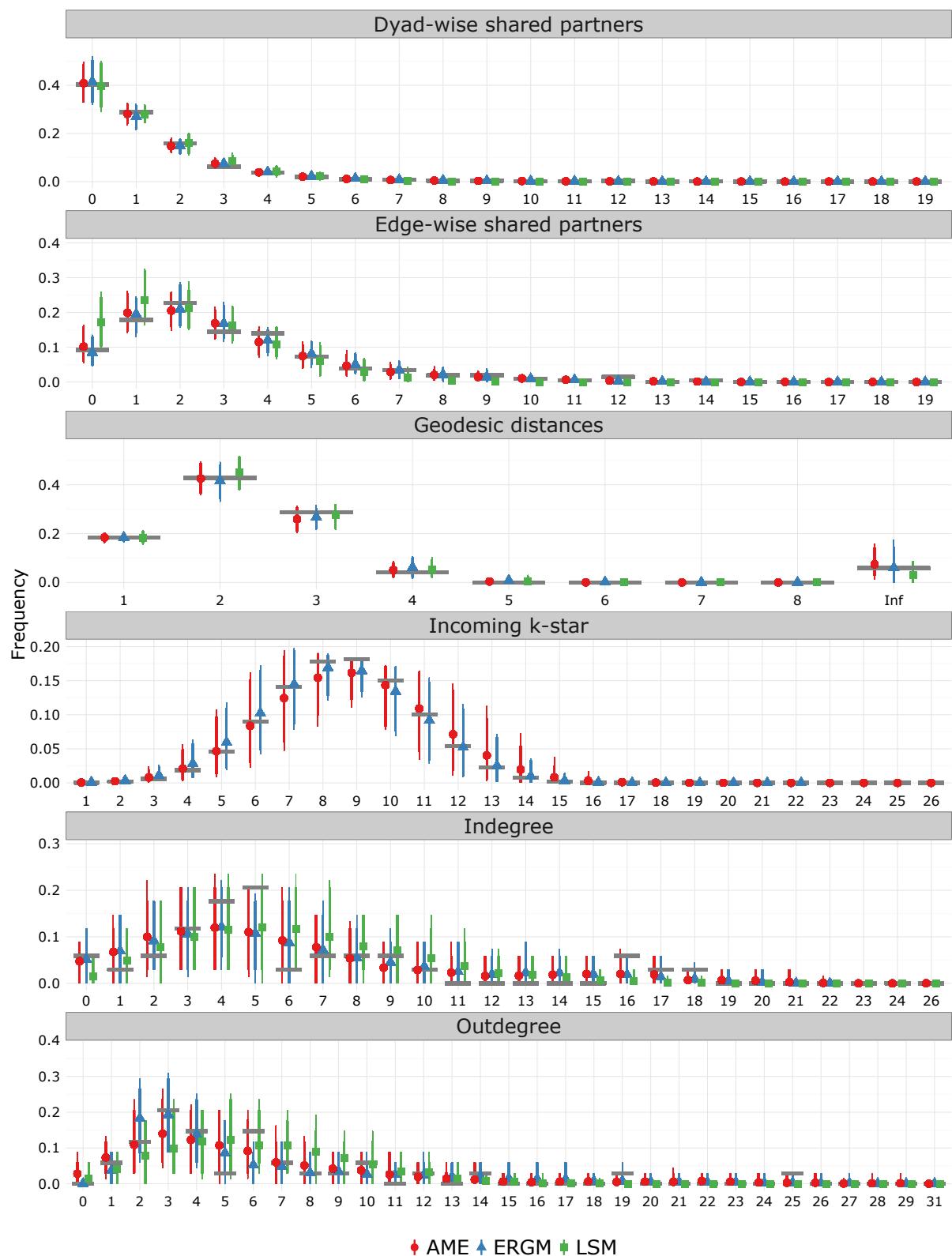
For a given summary statistic $g()$ we first simulate $\mathbf{Y}_{sim} \approx p(\mathbf{Y}_{sim}|\mathbf{Y}_{obs}) = \int p(\mathbf{Y}_{sim}|\theta)p(d\theta|\mathbf{Y}_{obs})$ and then we compare $g(\mathbf{Y}_{sim})$ to $g(\mathbf{Y}_{obs})$. Histograms represent predicted value of statistics under the model and red dash line represents the observed value.

Proportion of ties that are reciprocated.

$$(7) \quad t(Y) = \frac{\sum_{i \neq j} y_{i,j} y_{j,i}}{\sum_{i \neq j} y_{i,j}}$$

Number of transitive triplets, number of triangles in network, number of times ijk are all connected.

$$(8) \quad t(Y) = \sum_{i \neq j \neq k} y_{i,j} y_{i,k} y_{j,k}$$

**Figure 4.** network stats

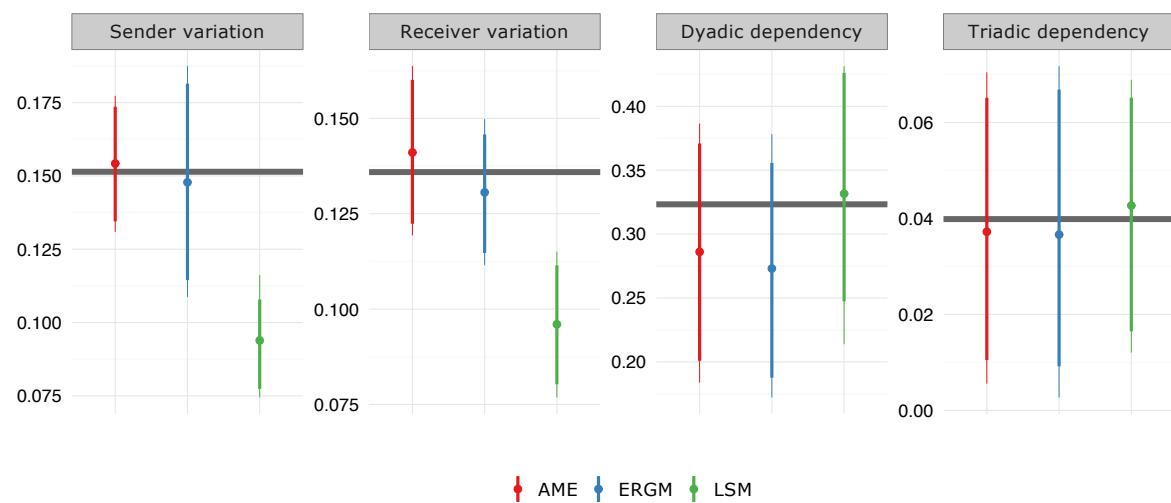


Figure 5. Posterior predictive goodness of fit summary

3. CONCLUSION

4. APPENDIX

4.1. AMEN Model Convergence. MCMC chain analysis for AMEN model presented in paper.

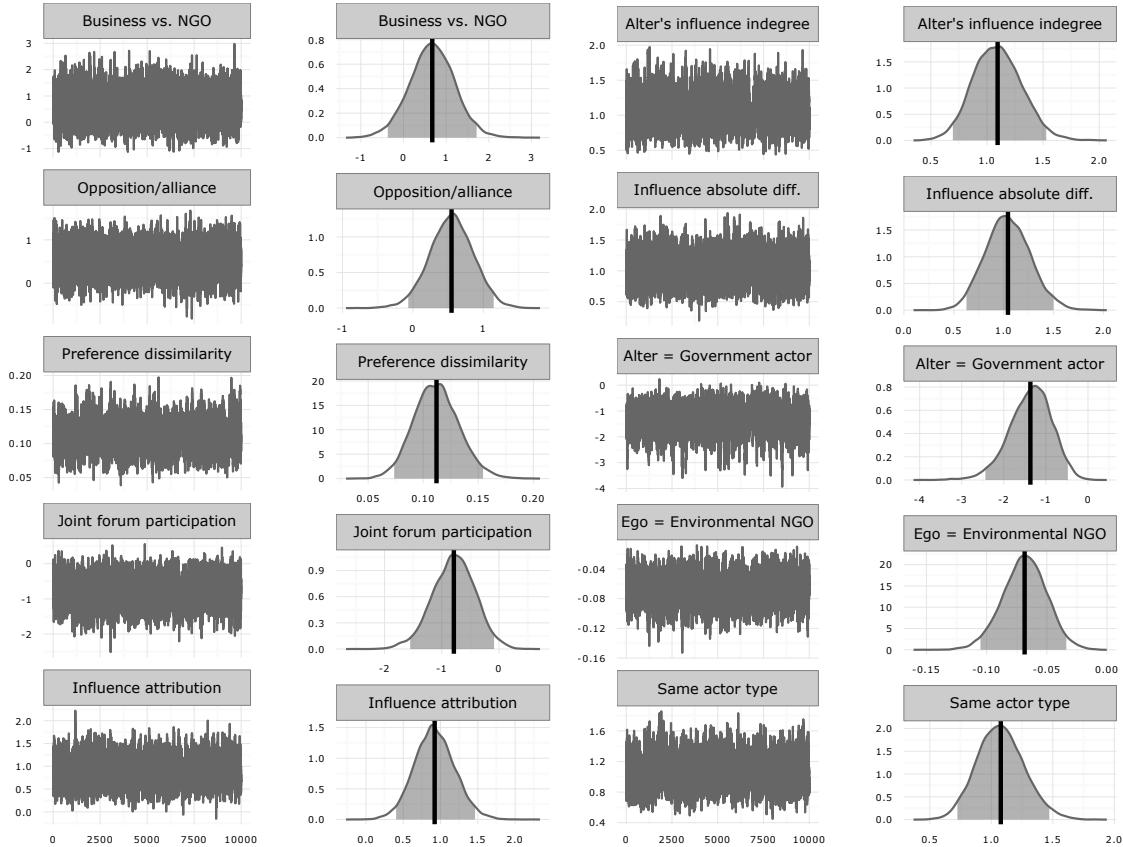
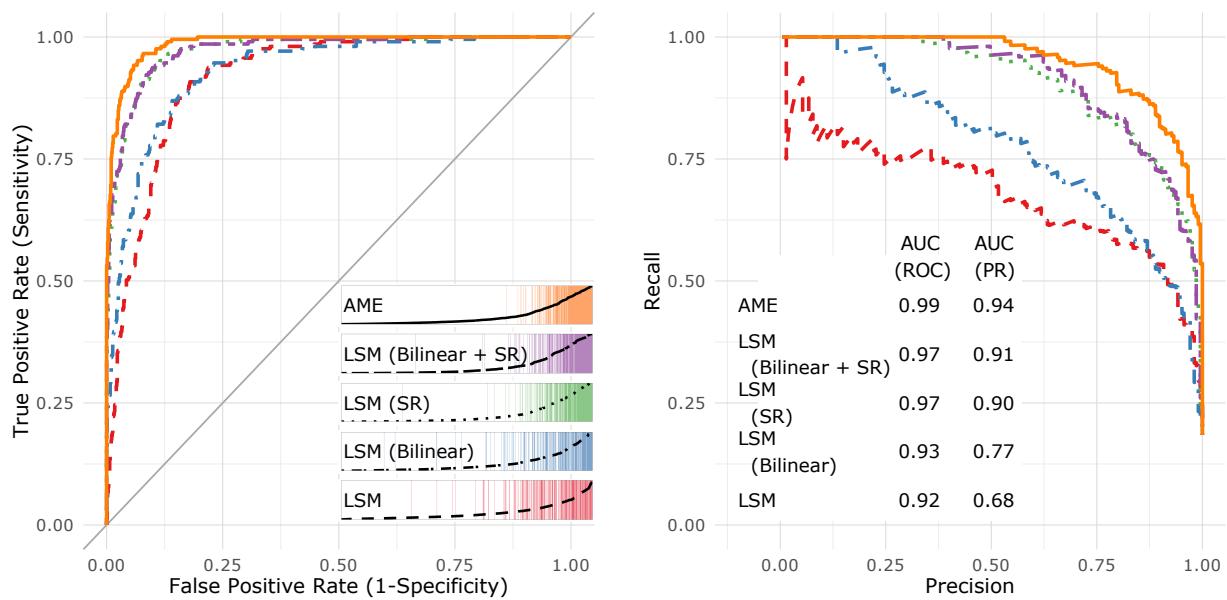
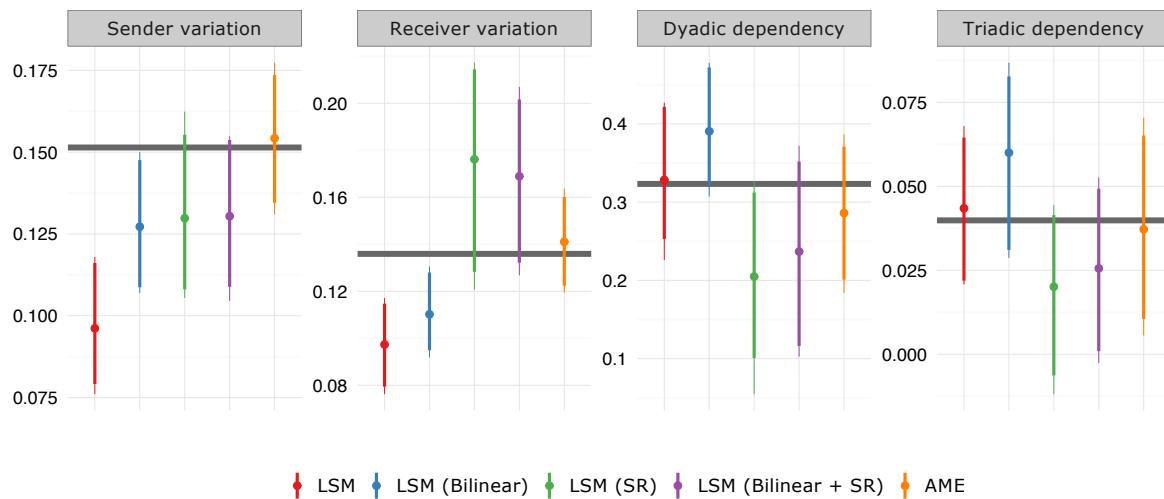


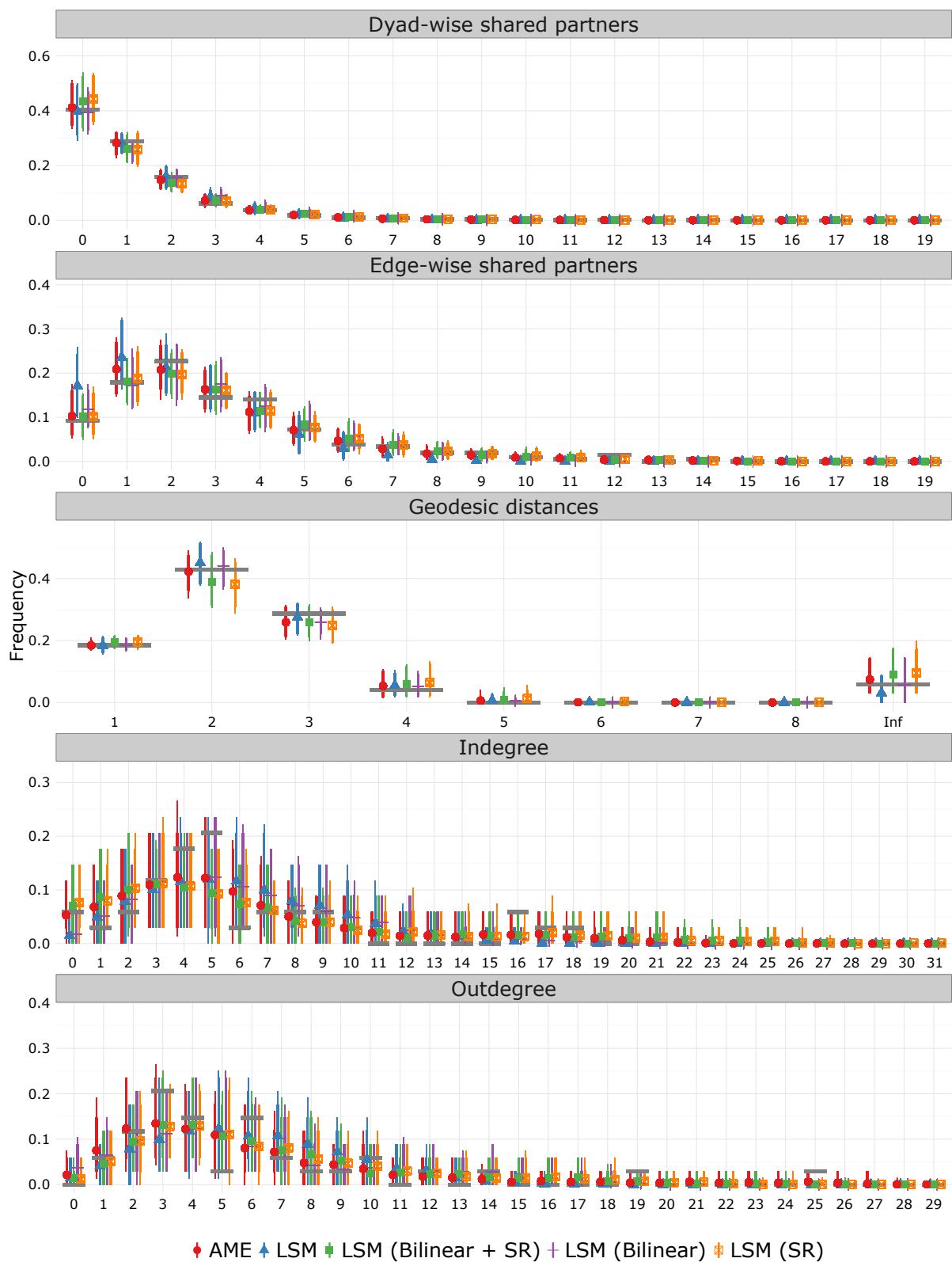
Figure 6. MCMC chain for AMEN model presented in paper. In this model, we utilize the SRM to account for first and second-order dependence. To account for third order dependencies we use the latent factor approach discussed earlier with $K = 2$.

4.2. Comparison of amen & latentnet R Packages. Here we provide a comparison of the AMEN model we presented in the paper with a variety of parameterizations from the **latentnet** package.

	LSM	LSM (Bilinear)	LSM (SR)	LSM (Bilinear + SR)	AME
Intercept/Edges	0.94* [0.09; 1.82]	-2.66* [-3.53; -1.87]	0.60 [-1.10; 2.37]	-2.50* [-4.14; -0.88]	-3.39* [-4.38; -2.50]
Conflicting policy preferences					
Business vs. NGO	-1.37* [-2.42; -0.41]	-2.64* [-4.61; -0.96]	-3.07* [-4.77; -1.56]	-2.87* [-4.63; -1.29]	-1.37* [-2.44; -0.47]
Opposition/alliance	0.00 [-0.40; 0.39]	0.04 [-0.44; 0.54]	0.31 [-0.24; 0.86]	0.24 [-0.36; 0.82]	1.08* [0.72; 1.47]
Preference dissimilarity	-1.76* [-2.62; -0.90]	-2.00* [-3.01; -1.03]	-1.88* [-3.07; -0.68]	-2.20* [-3.46; -0.96]	-0.79* [-1.55; -0.08]
Transaction costs					
Joint forum participation	1.51* [0.86; 2.17]	1.24* [0.53; 1.93]	1.56* [0.69; 2.41]	1.62* [0.70; 2.52]	0.92* [0.40; 1.47]
Influence					
Influence attribution	0.08 [-0.40; 0.55]	-0.08 [-0.62; 0.46]	0.30 [-0.37; 0.96]	0.28 [-0.42; 0.97]	1.09* [0.69; 1.53]
Alter's influence indegree	0.01 [-0.03; 0.04]	-0.05* [-0.09; -0.01]	0.06 [-0.03; 0.14]	0.05 [-0.04; 0.13]	0.11* [0.07; 0.15]
Influence absolute diff.	0.04 [-0.01; 0.09]	0.02 [-0.03; 0.07]	-0.08* [-0.14; -0.02]	-0.08* [-0.14; -0.02]	-0.07* [-0.11; -0.03]
Alter = Government actor	-0.46 [-1.08; 0.14]	-0.80 [-1.67; 0.04]	-0.11 [-1.91; 1.76]	-0.20 [-2.14; 1.74]	0.55 [-0.07; 1.15]
Functional requirements					
Ego = Environmental NGO	-0.60 [-1.32; 0.09]	-1.90* [-3.10; -0.86]	-1.69 [-3.74; 0.23]	-1.84 [-4.02; 0.11]	0.67 [-0.38; 1.71]
Same actor type	1.17* [0.63; 1.71]	1.40* [0.85; 1.95]	1.82* [1.10; 2.54]	1.90* [1.19; 2.62]	1.04* [0.63; 1.50]

Table 4. * p < 0.05 (or o outside the 95% confidence interval).

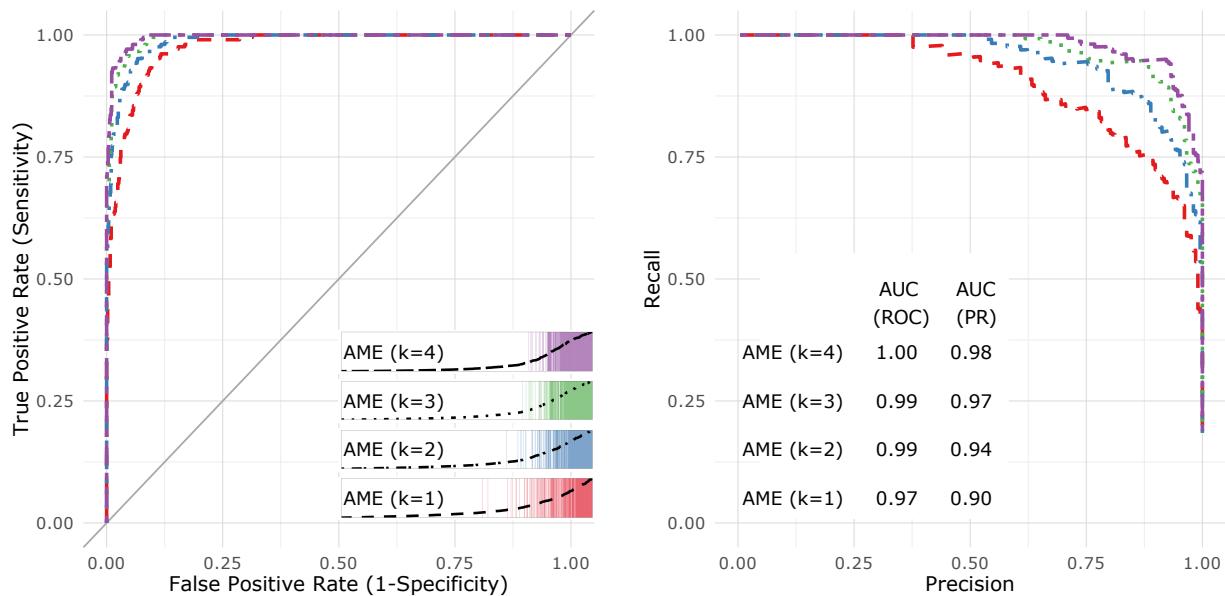
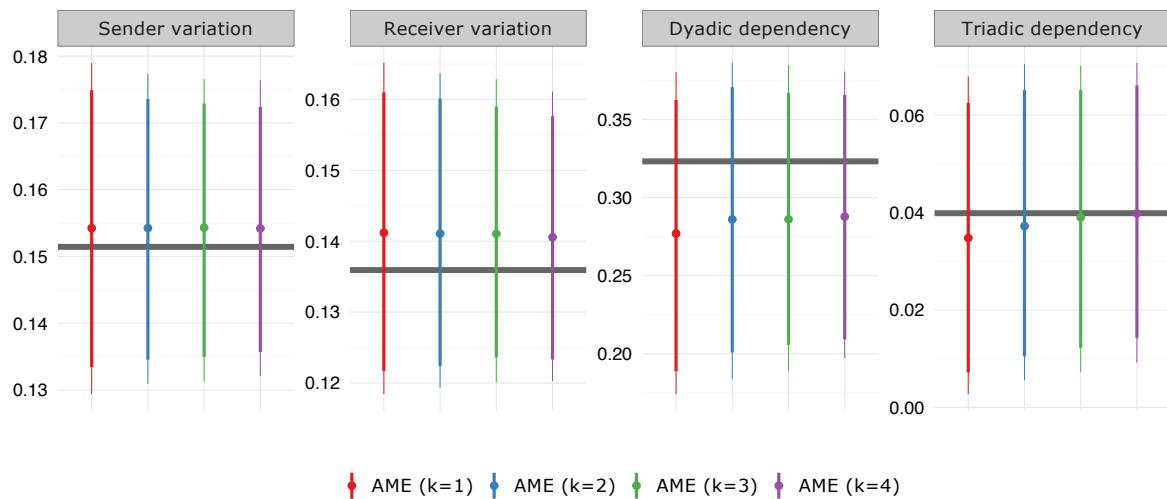
**Figure 7.** ROC and separation plots**Figure 8.** Posterior predictive goodness of fit summary

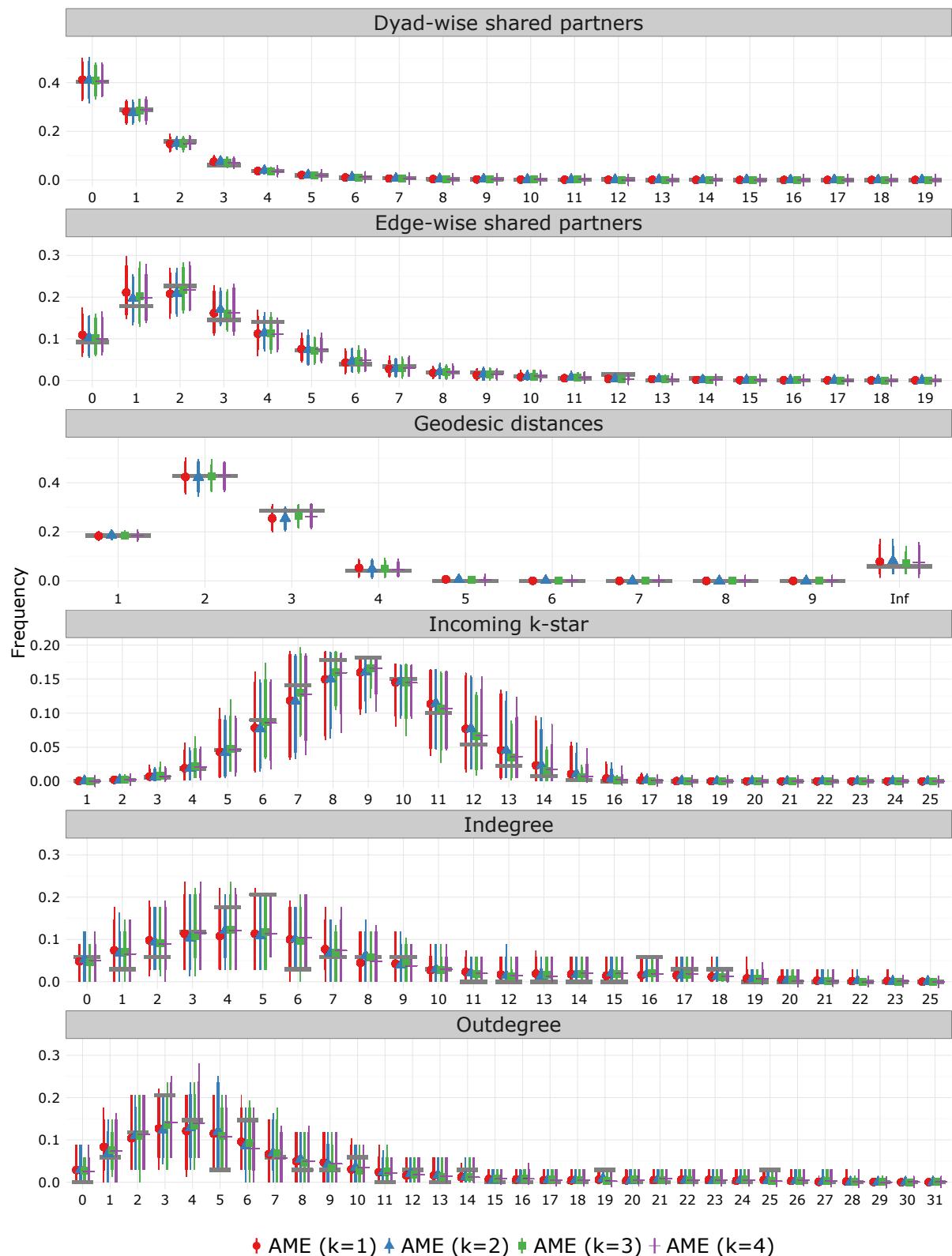
**Figure 9.** network stats

4.3. Comparison with other AME Parameterizations. Here we provide a comparison of the AMEN model we presented in the paper that uses $K = 2$ for multiplicative effects and show how results change when we use $K = /1, 3, 4/$.

	AME (k=1)	AME (k=2)	AME (k=3)	AME (k=4)
Intercept/Edges	-3.08* [-3.91; -2.30]	-3.39* [-4.38; -2.50]	-3.72* [-4.84; -2.73]	-3.93* [-5.12; -2.87]
Conflicting policy preferences				
Business vs. NGO	-1.28* [-2.20; -0.47]	-1.37* [-2.44; -0.47]	-1.48* [-2.63; -0.49]	-1.51* [-2.69; -0.47]
Opposition/alliance	0.95* [0.64; 1.27]	1.08* [0.72; 1.47]	1.19* [0.80; 1.64]	1.28* [0.86; 1.77]
Preference dissimilarity	-0.65* [-1.30; -0.03]	-0.79* [-1.55; -0.08]	-0.89* [-1.71; -0.12]	-0.95* [-1.80; -0.14]
Transaction costs				
Joint forum participation	0.84* [0.38; 1.31]	0.92* [0.40; 1.47]	1.01* [0.44; 1.62]	1.06* [0.43; 1.72]
Influence				
Influence attribution	1.00* [0.63; 1.39]	1.09* [0.69; 1.53]	1.21* [0.75; 1.71]	1.28* [0.80; 1.84]
Alter's influence indegree	0.10* [0.07; 0.14]	0.11* [0.07; 0.15]	0.12* [0.08; 0.17]	0.13* [0.09; 0.18]
Influence absolute diff.	-0.06* [-0.10; -0.03]	-0.07* [-0.11; -0.03]	-0.07* [-0.12; -0.04]	-0.08* [-0.12; -0.04]
Alter = Government actor	0.52 [-0.04; 1.07]	0.55 [-0.07; 1.15]	0.60 [-0.07; 1.27]	0.64 [-0.07; 1.35]
Functional requirements				
Ego = Environmental NGO	0.61 [-0.31; 1.56]	0.67 [-0.38; 1.71]	0.76 [-0.38; 1.90]	0.80 [-0.40; 2.04]
Same actor type	0.97* [0.60; 1.35]	1.04* [0.63; 1.50]	1.11* [0.64; 1.59]	1.17* [0.68; 1.68]

Table 5. * $p < 0.05$ (or o outside the 95% confidence interval).

**Figure 10.** ROC and separation plots**Figure 11.** Posterior predictive goodness of fit summary

**Figure 12.** network stats

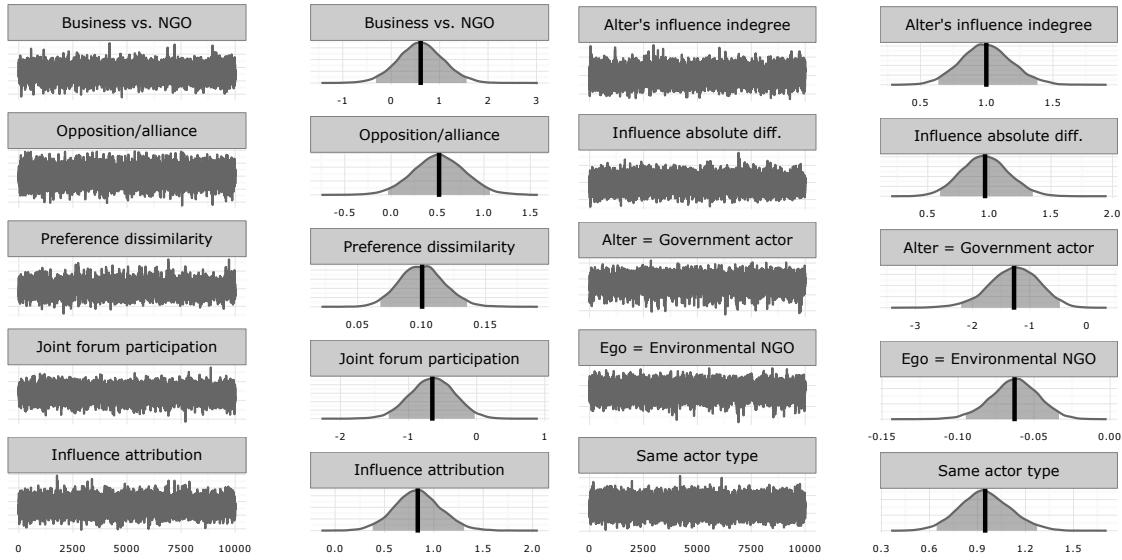


Figure 13. MCMC chain for AMEN model in which we utilize the SRM to account for first and second-order dependence. To account for third order dependencies we use the latent factor approach discussed earlier with $K = 1$.

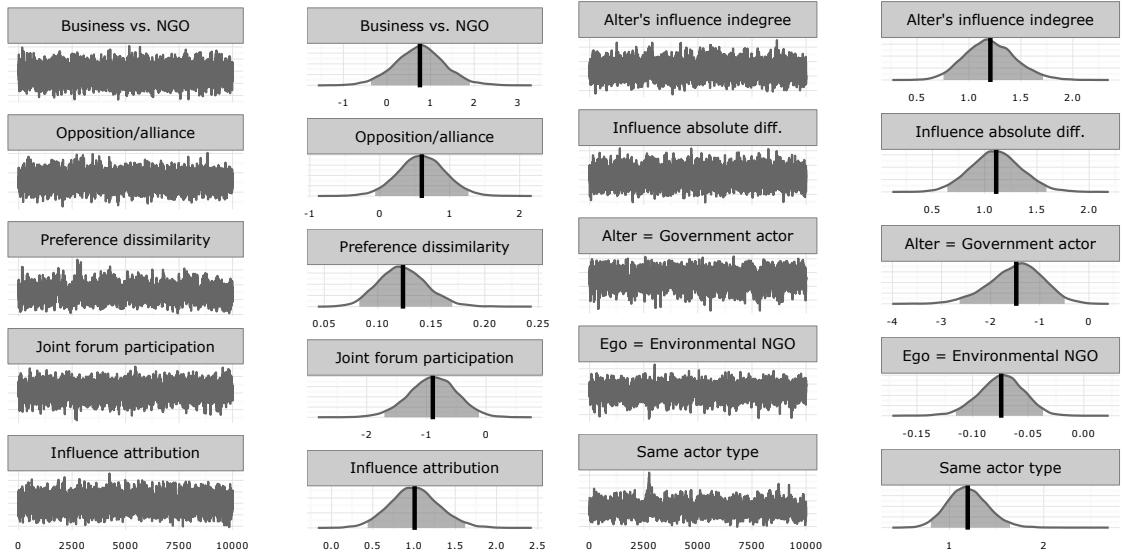


Figure 14. MCMC chain for AMEN model in which we utilize the SRM to account for first and second-order dependence. To account for third order dependencies we use the latent factor approach discussed earlier with $K = 3$.

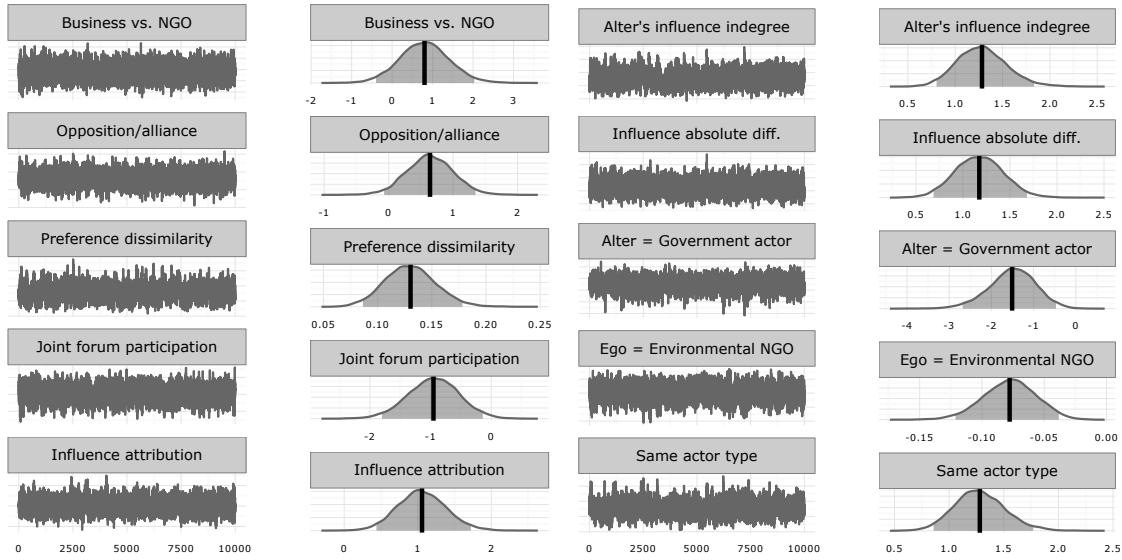


Figure 15. MCMC chain for AMEN model in which we utilize the SRM to account for first and second-order dependence. To account for third order dependencies we use the latent factor approach discussed earlier with $K = 4$.

REFERENCES

- Alemán, Eduardo and Ernesto Calvo. 2013. "Explaining Policy Ties in Presidential Congresses: A Network Analysis of Bill Initiation Data: Policy Ties in Presidential Congresses." *Political Studies* 61(2):356–377.
- Anderson, Carolyn J, Stanley Wasserman and Katherine Faust. 1992. "Building stochastic blockmodels." *Social networks* 14(1):137–161.
- Barabási, Albert-László and Albert Réka. 1999. "Emergence of Scalings in Random Networks." *Science* 286:509–510.
- Beck, Nathaniel and Jonathan N. Katz. 1995. "What to Do (and Not to Do) With Pooled Time-Series Cross-Section Data." *American Political Science Review* 89(3):634–647.
- Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable." *American Journal of Political Science* 42(2):1260–1288.

- Berardo, Ramiro and John T. Scholz. 2010. "Self-Organizing Policy Networks: Risk, Partner Selection, and Cooperation in Estuaries." *American Journal of Political Science* 54(3):632–649.
- Besag, Julian E. 1977. "Efficiency of Pseudolikelihood Estimation for Simple Gaussian Fields." *Biometrika* 64:616.
- Bhamidi, Shankar, Guy Bresler and Allan Sly. 2008. Mixing time of exponential random graphs. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*. IEEE pp. 803–812.
- Bolton, Gary E., Jordi Brandts and Axel Ockenfels. 1998. "Measuring motivations for the reciprocal responses observed in a simple dilemma game." *Experimental Economics* 1(3):207–219.
- Bonabeau, Eric. 2002. "Agent-based modeling: Methods and techniques for simulating human systems." *Proceedings of the National Academy of Sciences* 99(suppl 3):7280–7287.
- Brandes, Ulrik and Thomas Erlebach. 2005. *Network analysis: methodological foundations*. Vol. 3418 Springer Science & Business Media.
- Brandt, Patrick T., Michael Colaresi and John R. Freeman. 2008. "The Dynamics of Reciprocity, Accountability, and Credibility." *Journal of Conflict Resolution* 52(3):343–374.
- Breunig, Christian, Xun Cao and Adam Luedtke. 2012. "Global Migration and Political Regime Type: A Democratic Disadvantage." *British Journal of Political Science* 42(4):tba.
- Calvo, Ernesto and Marcelo Leiras. 2012. "The Nationalization of Legislative Collaboration: Territory, Partisanship, and Policymaking in Argentina." *Revista Ibero-Americana de Estudios Legislativos* 1(2):2–19.
- Cao, Xun. 2009. "Networks of intergovernmental organizations and convergence in domestic economic policies." *International Studies Quarterly* 53(4):1095–1130.
- Cao, Xun. 2010. "Networks as Channels of Policy Diffusion: Explaining Worldwide Changes in Capital Taxation, 1998-2006." *International Studies Quarterly* 54(3):823–854.

- Cao, Xun. 2012. "Global networks and domestic policy convergence: A network explanation of policy changes." *World Politics* 64(03):375–425.
- Cao, Xun and Michael D. Ward. 2014. "Do Democracies Attract Portfolio Investment?" *International Interactions* 40(2):216–245.
- Carnegie, Allison. 2014. "States held hostage: Political hold-up problems and the effects of international institutions." *American Political Science Review* 108(01):54–70.
- Chatterjee, Sourav and Persi Diaconis. 2013. "Estimating and understanding exponential random graph models." *The Annals of Statistics* 41(5):2428–2461.
- Choucri, Nazli and Robert C. North. 1972. "Dynamics of International Conflict." *World Politics* 24(2):80–122.
- Cox, James C., Daniel Friedman and Steven Gjerstad. 2007. "A tractable model of reciprocity and fairness." *Games and Economic Behavior* 59(1):17–45.
- Cranmer, Skyler and Bruce A. Desmarais. 2011. "Inferential Network Analysis with Exponential Random Graph Models." *Political Analysis* 19(1):66–86.
- Cranmer, Skyler J., Philip Leifeld, Scott D. McClurg and Meredith Rolfe. 2016. "Navigating the Range of Statistical Tools for Inferential Network Analysis." *American Journal of Political Science*.
- Cranmer, Skyler J., Tobias Heinrich and Bruce A. Desmarais. 2014. "Reciprocity and the structural determinants of the international sanctions network." *Social Networks* 36:5–22.
- Dafoe, Allan. 2011. "Statistical Critiques of the Democratic Peace: Caveat Emptor." *American Journal of Political Science* 55(2):247–262.
- Diehl, Paul F. and Thorin M. Wright. 2016. "A Conditional Defense of the Dyadic Approach." *International Studies Quarterly*.
- Dixon, William. 1983. "Measuring interstate affect." *American Journal of Political Science* pp. 828–851.

- Dorff, Cassy and Michael D Ward. 2013. "Networks, Dyads, and the Social Relations Model." *Political Science Research and Methods* 1(02):159–178.
- Dorff, Cassy and Shahryar Minhas. 2016. "When Do States Say Uncle? Network Dependence and Sanction Compliance." *International Interactions* .
- Erdős, Paul and Alfréd Rényi. 1959. "On random graphs." *Publicationes Mathematicae* 6:290–297.
- Erikson, Robert S., Pablo M. Pinto and Kelly T. Rader. 2014. "Dyadic analysis in international relations: A cautionary tale." *Political Analysis* 22(4):457–463.
- Frank, Ove. 1971. *Statistical Inference in Graphs*. Stockholm: FOA.
- Frank, Ove and David Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81(395):832–842.
- Franzese, Robert and Jude C. Hayes. 2007. "Spatial Econometric Models for the Analysis of TSCS Data in Political Science." *Political Analysis* 15(15):2.
- Fuhrmann, Matthew and Todd S. Sechser. 2014. "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence." *American Journal of Political Science* 58(4):919–935.
- Geyer, Charles J. and Elizabeth A. Thompson. 1992. "Constrained Monte Carlo Maximum Likelihood for Dependent Data, (with Discussion)." *Journal of the Royal Statistical Society, Series B, Methodological* 54:657–699.
- Goldstein, Joshua S. and John R. Freeman. 1990. *Three Way Street: Strategic Reciprocity in World Politics*. Chicago, IL: University of Chicago Press.
- Goldstein, Joshua S. and John R. Freeman. 1991. "U.S.-Soviet-Chinese Relations: Routine, Reciprocity, or Rational Expectations?" *American Political Science Review* 85(1):pp. 17–35.
- Goodreau, Steven M, Mark S Handcock, David R Hunter, Carter T Butts and Martina Morris. 2008. "A statnet Tutorial." *Journal of Statistical Software* 24(9):1.

- Goyal, Sanjeev. 2012. *Connections: an introduction to the economics of networks*. Princeton University Press.
- Greenhill, Brian D. 2015. *Transmitting rights; International Organizations and the Diffusion of Human Rights Practices*. Oxford, UK: Oxford University Press.
- Greenhill, Brian, Michael D. Ward and Audrey Sacks. 2011. "The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Data." *American Journal of Political Science* 55(4):991-1002.
- Hammersley, John M and Peter Clifford. 1971. "Markov fields on finite graphs and lattices."
- Handcock, Mark. 2003. Assessing degeneracy in statistical models of social networks. Technical report. Working Paper 39, Center for Statistics and the Social Sciences, University of Washington.
- Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau and Martina Morris. 2008. "statnet: Software tools for the representation, visualization, analysis and simulation of network data." *Journal of Statistical Software* 24(1):1548.
- Heaney, Michael T. 2014. "Multiplex Networks and Interest Group Influence Reputation: An Exponential Random Graph Model." *Social Networks* 36(1):66–81.
- Hoff, Peter. 2008. Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems*. pp. 657–664.
- Hoff, Peter. 2009. "Multiplicative latent factor models for description and prediction of social networks." *Computational and Mathematical Organization Theory* 15(4):261–272.
- Hoff, Peter, Bailey Fosdick, Alex Volfovsky and Yanjun He. 2015. *amen: Additive and Multiplicative Effects Models for Networks and Relational Data*. R package version 1.1.
URL: <https://CRAN.R-project.org/package=amen>
- Hoff, Peter D. 2005. "Bilinear Mixed-Effects Models for Dyadic Data." *Journal of the American Statistical Association* 100(4690):286–295.

- 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.
- Hoff, Peter D. and Michael D. Ward. 2004. "Modeling Dependencies in International Relations Networks." *Political Analysis* 12(2):160–175.
- Hunter, David, Mark Handcock, Carter Butts, Steven M Goodreau and Martina Morris. 2008. "ergm: A package to fit, simulate and diagnose exponential-family models for networks." *Journal of Statistical Software* 24(3):1–29.
- Hunter, David, Pavel Krivitsky and Michael Schweinberger. 2012. "Computational statistical methods for social network models." *Journal of Computational and Graphical Statistics* 21(4):856–882.
- Ingold, Karin. 2008. *Les mécanismes de décision: Le cas de la politique climatique Suisse. Politikanalysen*. Rüegger Verlag, Zürich.
- Ingold, Karin and Manuel Fischer. 2014. "Drivers of collaboration to mitigate climate change: An illustration of Swiss climate policy over 15 years." *Global Environmental Change* 24:88–98.
- Ingold, Karin and Philip Leifeld. 2014. "Structural and institutional determinants of influence reputation: A comparison of collaborative and adversarial policy networks in decision making and implementation." *Journal of Public Administration Research and Theory* p. muu043.
- Jackson, Matthew. 2014. "Networks in the understanding of economic behaviors." *The Journal of Economic Perspectives* 28(4):3–22.
- Kenny, David A., Deborah A. Kashy and William L. Cook. 2006. *Dyadic Data Analysis*. New York: Guilford Press.
- Keohane, Robert O. 1989. "Reciprocity in international relations." *International Organization* 40(1).

Kinne, Brandon J. 2013. "Network dynamics and the evolution of international cooperation." *American Political Science Review* 107(04):766–785.

Kirkland, Justin H. and R. Lucas Williams. 2014. "Partisanship and Reciprocity in Cross-Chamber Legislative Interactions." *Journal of Politics* 6(3):754–769.

Kolaczyk, Eric D. 2009. *Statistical Analysis of Network Data: Methods and Models*. Berlin: Springer Verlag.

Krivitsky, Pavel N. and Mark S. Handcock. 2015. *latentnet: Latent Position and Cluster Models for Statistical Networks*. The Statnet Project (<http://www.statnet.org>). R package version 2.7.1.

URL: <http://CRAN.R-project.org/package=latentnet>

Lai, David. 1995. "A Structural Approach to Alignment: A Case Study of the China- Soviet-U.S. Strategic Triangle, 1971-1988." *International Interactions* 20(4):349–374.

Lemke, Douglas and William Reed. 2001. "War and Rivalry among Great Powers." *American Journal of Political Science* 45(2):457–469.

Li, Heng and Eric Loken. 2002. "A Unified Theory of Statistical Analysis and Inference for Variance Component Models For Dyadic Data." *Statistica Sinica* 12(2):519–535.

Lubbers, Miranda J and Tom AB Snijders. 2007. "A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes." *Social Networks* 29(4):489–507.

Lubell, Mark and Scholz, John, Ramiro Berardo and Garry Robbins. 2012. "Testing Policy Theory with Statistical Models of Networks." *Policy Studies Journal* 40(3):351–374.

Manger, Mark S., Mark A. Pickup and Tom A.B. Snijders. 2012. "A Hierarchy of Preferences: A Longitudinal Network Analysis Approach to PTA Formation." *Journal of Conflict Resolution* 56(5):852–877.

Mansfield, Edward, Helen V. Milner and B. Peter Rosendorff. 2000. "Free to Trade? Democracies, Autocracies, and International Trade Negotiations." *American Political Science Review* 94(2):305–321.

- Maoz, Zeev, Ranan D. Kuperman, Lesley Terris and Ilan Talmud. 2006. "Structural Equivalence and International Conflict: A Social Networks Analysis." *Journal of Conflict Resolution* 50(5):664–689.
- Metternich, Nils, Shahryar Minhas and Michael Ward. 2015. "Firewall? or Wall on Fire? A Unified Framework of Conflict Contagion and the Role of Ethnic Exclusion." *Journal of Conflict Resolution* .
- Minhas, Shahryar, Peter D. Hoff and Michael D. Ward. 2016. "A New Approach to Analyzing Coevolving Longitudinal Networks in International Relations." *Journal of Peace Research* 53(3):491–505.
- Mitchell, Sara McLaughlin. 2002. "A Kantian system? Democracy and third-party conflict resolution." *American Journal of Political Science* pp. 749–759.
- 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(4):1554–7660.
- Nowicki, Krzysztof and Tom A. B. Snijders. 2001. "Estimation and Prediction for Stochastic Blockstructures." *Journal of the American Statistical Association* 96(455):1077–1087.
- Pattison, Philippa and Stanley Wasserman. 1999. "Logit models and logistic regressions for social networks. II. Multivariate relations." *British Journal of Mathematical and Statistical Psychology* 52:169–194.
- Pawitan, Yudi. 2013. *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. 1st ed. Oxford, England: Oxford University Press.
- Rajmaira, Sheen and Michael D. Ward. 1990. "Evolving Foreign Policy Norms: Reciprocity in the Superpower Triad." *International Studies Quarterly* 34:457–475.
- Réka, Albert, Hawoong Jeong and Albert-László Barabási. 1999. "The Diameter of the WWW." *Nature* 401(6749):130–131.
- Richardson, Lewis F. 1960. *Arms and Insecurity*. Quadrangle.

- Robbins, Garry, Jenny M. Lewis and Peng Wang. 2012. "Statistical Network Analysis for Analyzing Policy Networks." *Policy Studies Journal* 40(3):375–401.
- 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.
- Schweinberger, Michael. 2011. "Instability, sensitivity, and degeneracy of discrete exponential families." *Journal of the American Statistical Association* 106(496):1361–1370.
- Shalizi, Cosma Rohilla and Andrew C. Thomas. 2011. "Homophily and contagion are generically confounded in observational social network studies." *Sociological Methods & Research* 40(2):211–239.
- Signorino, Curtis. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 92(2):279–298.
- 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(1):99–153.
- Snijders, Tom AB. 1996. "Stochastic actor-oriented models for network change." *Journal of Mathematical Sociology* 21(1-2):149–172.
- Snijders, Tom A.B. 2002. "Markov Chain Monte Carlo Estimation of Exponential Random, Graph Models." *Journal of Social Structure* 3(2):via web:, zeeb.library.cmu.edu: 7850/JoS/Snijders/Mcpstar.pdf.
- Snijders, Tom A.B. 2011. "Statistical Models for Social Networks." *Annual Review of Sociology* 37:131–53.
- Strauss, Davis and Michael Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85:204–212.
- Van Duijn, Marijtje AJ, 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.

- Victor, Jennifer Nicoll, Alexander H. Montgomery and Mark Lubell, eds. 2016. *The Oxford Handbook of Political Networks*. Oxford, UK: Oxford University Press.
- Victor, Jennifer and Nils Ringe. 2009. "The Social Utility of Informal Institutions: Caucuses as Networks in the 100th U.S. House of Representatives." *American Politics Research* 37(5):742–66.
- Ward, Michael D., John S. Ahlquist and Arturas Rozenas. 2012. "Gravity's Rainbow: A Dynamic Latent Space Model for the World Trade Network." *Network Science* 1(1):95–118.
- Ward, Michael D. and Kristian Skrede Gleditsch. 2008. *Spatial Regression Models*. Vol. 155 Sage.
- Ward, Michael D. and Peter D. Hoff. 2007. "Persistent Patterns of International Commerce." *Journal of Peace Research* 44(2):157–175.
- Ward, Michael D. and Peter D. Hoff. 2008. Analyzing Dependencies in Geo-Politics and Geo-Economics. In *Contributions to Conflict Management, Peace Economics, and Development: Volume 6: War, Peace, and Security*, ed. Jacques Fontanel and Manas Chatterji. Amsterdam: Elsevier Science pp. 133–160.
- Ward, Michael D., Randolph M. Siverson and Xun Cao. 2007. "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace." *American Journal of Political Science* 51(3):583–601.
- Ward, Michael D. and Sheen Rajmaira. 1992. "Reciprocity and Norms in U.S.–Soviet Foreign Policy." *Journal of Conflict Resolution* 36:342–368.
- Wasserman, Stanley and Phillipa Pattison. 1996. "Logit, models and logistic regression for social networks: I. An introduction to Markov graphs and p^* ." *Psychometrika* 61:401–425.
- Wong, George Y. 1982. "Round robin analysis of variance via maximum likelihood." *Journal of the American Statistical Association* 77(380):714–724.

SHAHRYAR MINHAS: DEPARTMENT OF POLITICAL SCIENCE

Current address: Michigan State University

E-mail address: s7.minhas@gmail.com

PETER D. HOFF: DEPARTMENT OF STATISTICS

Current address: Duke University

E-mail address: peter.hoff@duke.edu

MICHAEL D. WARD: DEPARTMENT OF POLITICAL SCIENCE

Current address: Duke University

E-mail address: michael.d.ward@duke.edu