LET'S SAY 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).

Date: September 24, 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 IR for dealing with data structures of this sort, 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 amongst themselves.[2] An even harder assumption to maintain 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 continue to favor the dyadic approach is that many events are not multilateral (Diehl & Wright, 2016), thus alleviating the need for an approach that incorporates interdependencies between observations. The network perspective, however, is that even the bilateral events we study are taking place within a broader system, and what takes place in one part of that system may be dependent upon another.

The potential for interdependence between observations poses a challenge to statistical modeling as the assumption made by standard approaches used across the social

[1]To 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 et al. (2000); Lemke & Reed (2001); Mitchell (2002); Dafoe (2011); Fuhrmann & Sechser (2014); Carnegie (2014).

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

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

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.[4] More relevant is the fact that a wealth of research from other disciplines would argue that carrying the independence assumption into a study with relational data is misguided and likely to lead to biased inferences.[5]

Despite the hesitation among some in the discipline to adopt network analytic approaches, in recent years we have at least seen 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 piece by Cranmer et al. (2016) that provides an overview and comparison of a handful of network based inferential models, 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 et al. (2002). Their discussion around the differences in these approaches and their empirical comparison of them is extremely valuable and necessary, at the same time, they overlook a decade worth of developments that latent variable models have undergone. This is particularly relevant in the context of providing an overview for the field as by focusing on the results from one early attempt at a latent variable model, they end up overlooking much of the work that has actually been done using this type of approach in political science. 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 & Ward (2004); Ward et al. (2007); Metternich

[4] For example, see Beck et al. (1998); Signorino (1999); Hoff & Ward (2004); Franzese & Hayes (2007); Cranmer & Desmarais (2011); Erikson et al. (2014).

[5] From Computer Science see: Bonabeau (2002); Brandes & Erlebach (2005). From Economics see: Goyal (2012); Jackson (2014). From Psychology see: Pattison & Wasserman (1999); Kenny et al. (2006). From Statistics see: Snijders (1996); Hoff et al. (2002).

et al. (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.

In this paper, we introduce the additive and multiplicative effects network model (AMEN). To highlight the benefits of this approach we estimate this model on the application presented in Cranmer et al. (2016) and compare it to the other models presented in that paper, when 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). The rest of this paper 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 interdependence while still allowing scholars to test 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.

[6]The 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/.

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

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

Table 1. Structure of datasets used in canonical design.

$\longrightarrow$

	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 (Ward & Ahlquist, 2010). The stochastic component reflects our assumptions about the probability distribution from which the data is generated: $y_{ij} \overset{\text{iid}}{\sim} \mathcal{F}(\theta_{ij})$, where $\mathcal{F}$ represents a probability distribution or mass function such as normal or binomial, and $\overset{\text{iid}}{\sim}$ 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 $\mathbf{X}_{ij}$ the parameters of our distribution each of the dyadic observations are conditionally independent.

The usage of this assumption becomes clearer if we go a bit further 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.

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

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

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

Having constructed the likelihood we can proceed by solving through maximization or numerical analysis. However, the important point to note here 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 y_{ij} is independent of 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 complicates the a priori assumption of observational independence, and without this assumption the joint density function cannot be written in the way described above and we cannot produce a valid likelihood.[10] Thus inferences drawn from models that ignore potential interdependencies between dyadic observations are likely to have a number of issues such as biased effect estimation, 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 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. In undirected data an event cannot be

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

[9]For example, see Ward et al. (2007); Cranmer et al. (2014); Dorff & Minhas (2016).

[10]This problem has been noted in works such as Lai (1995); Manger et al. (2012); Kinne (2013).

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

attributed to a specific sender or receiver rather it is just an indication of a relationship between a pair of countries, an example of this that commonly arises in the IR literature involves models of alliance relationships. In directed matrices, the off-diagonal values are not symmetric and there is a clear sender and receiver as in the case of exports or aid flows.

A common type of structural interdependency that arises in relational data structures is "preferential attachment" (Réka et al., 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 $\{y_{ij}, y_{ik}, y_{il}\}$, 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 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.

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.[12] More specifically, this concept actually has deep roots in political science theories of cooperation and the evolution of norms between states (Richardson, 1960; Choucri & North, 1972; Keohane, 1989). 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, and to do so Warner et al. (1979) developed the social relational model (SRM), a type of ANOVA decomposition technique.[13] The SRM is of particular note as it provides the error 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 across column means (in-degree), correlation between row and column means, and correlations within dyads. Wong (1982) and Li & Loken (2002) provide a random effects representation of the SRM:

[12] For example, see Bolton et al. (1998); Cox et al. (2007).

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

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

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 linear mixed effects model: $\boldsymbol{\beta}^T\mathbf{X}_{ij} + a_i + b_j + \epsilon_{ij}$, where $\boldsymbol{\beta}^T\mathbf{X}_{ij}$ accommodates the inclusion of dyadic, sender, and receiver covariates. Through this approach we can effectively incorporate row, column, and within-dyad dependence in a regression framework. Further this approach can be extended to a 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: $\theta_{ij} = \boldsymbol{\beta}^T \mathbf{X}_{ij} + e_{ij}$. 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: $y_{ij} = I(\theta_{ij} > 0)$.

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 form 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.[14]

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

[14]Homophily 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 & Thomas (2011) for a more detailed discussion on the concept of homophily.

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, one, not independent of the ties that already exist between those actors and i, and, second, higher than the probability that either of those actors might form a tie with another actor, l with whom they have no shared attributes.

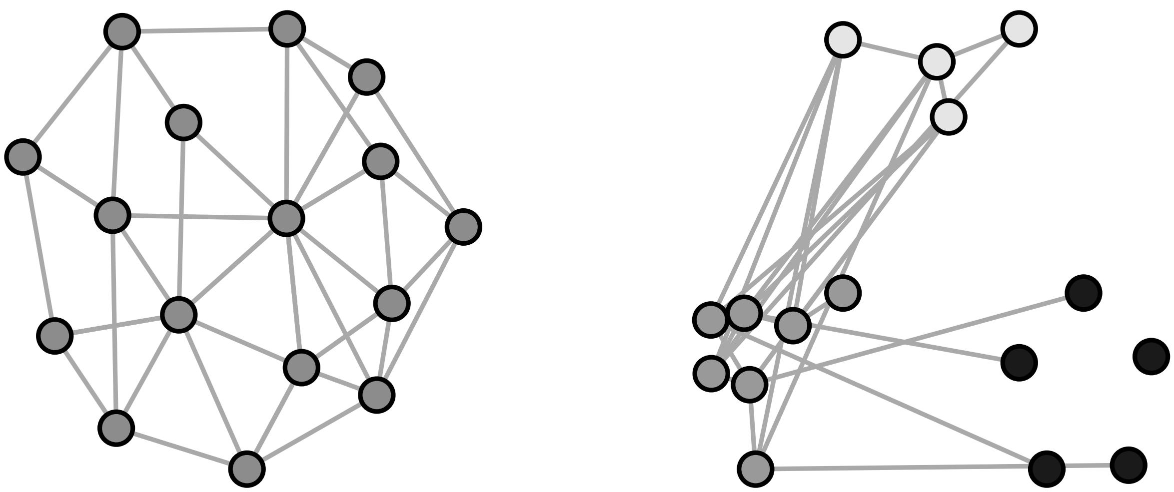

Figure 1. Graph on the left is a representation of an undirected network that exhibits a high degree of homophily, while on the left we show an undirected network that exhibits stochastic equivalence.

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 ij 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 et al., 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 et al. (2012) theorize and estimate a stochastic equivalence structure to explain the formation of preferential trade agreements (PTAs). Specifically, they theorize that PTA formation is related to differences in per capita income levels between countries. Countries falling into high, middle, and low income per capita levels will have patterns of PTA formation that are determined by the groups they fall into. 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 point here is that the behavior of actors in a network can at times be governed by group level dynamics, and failing to account for patterns such as this could lead to discounting an important part of the data generating process.

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 is likely enough to justify 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.[15] There are 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 an assumption such as this is, in spirit, the same reason why we no longer model time-series cross-sectional data without including country level fixed or random effects.

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. ERGMs, originally developed by Frank & Strauss (1986) and Wasserman & Pattison (1996), are particularly useful when researchers are interested in, and have a theoretical justification for, 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

[15]The main function in the **amen** package is titled "ame" and by default it runs a model assuming that no multiplicative effects are necessary.

in a network, balanced triads, reciprocal pairs and so on.[16] 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:

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

$\boldsymbol{\beta}$ 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 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 & Clifford (1971).

However, 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 & Ikeda (1990), but that approach ignores the interdependent nature of observations in

[16] Morris et al. (2008) and Snijders et al. (2006) provide a detailed list of network statistics that can be included in an ERGM model specification.

relational data structures, as a result, many have argued that the standard errors remain unreliable (Lubbers & Snijders, 2007; Robins et al., 2007; Van Duijn et al., 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. One favored approach in the literature is to approximate the MLE using Markov Chain Monte Carlo techniques, also referred to as MCMC-MLE (Geyer & 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** 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 et al. (2008) and Chatterjee & Diaconis (2013) have shown that MCMC procedures can take an exponential time to converge for broad classes of ERGMs unless the dyadic observations are essentially 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 very related issue when estimating ERGMs is the model becoming degenerate, meaning 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. 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 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 what we see to be a major motivation for pursuing an inferential network model in the first place.

1.3.2. ***Latent Space Models.*** Given the computational and inferential difficulties that go along with utilizing ERGMs, an alternative approach that has been utilized by political scientists in studying relational data with third order dependence patterns are latent space models. 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 obvious reason for their increased usage is that they enable us 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 et al., 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.

$$
\begin{aligned}
&\text{Latent class model} \\
&\qquad \alpha(\mu_i, \mu_j) = m_{\mu_i,\mu_j} \\
&\qquad \mu_i \in \{1, \ldots, K\},\ i \in \{1, \ldots, n\} \\
&\qquad M \text{ a } K \times K \text{ symmetric matrix} \\
&\text{Latent distance model} \\
&\qquad \alpha(\boldsymbol{\mu_i}, \boldsymbol{\mu_j}) = -|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j| \\
&\qquad \boldsymbol{\mu}_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\} \\
&\text{Latent factor model} \\
&\qquad \alpha(\boldsymbol{\mu_i}, \boldsymbol{\mu_j}) = \boldsymbol{\mu}_i^T \Lambda \boldsymbol{\mu}_j \\
&\qquad \boldsymbol{\mu}_i \in \mathbb{R}^K,\ i \in \{1, \ldots, n\} \\
&\qquad \Lambda \text{ a } K \times K \text{ diagonal matrix}
\end{aligned}
\tag{5}
$$

In the latent class model, also referred to as the stochastic block model, each node i is a member of some unknown latent class, $\mu_i \in (1, \ldots, K)$. A probability distribution is used to describe the relationships between classes (Nowicki & 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 wholely 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.[17]

A latent space approach that can characterize homophily is the latent distance model developed by Hoff et al. (2002). In this approach, each node i has some unknown latent position in K dimensional space, $\boldsymbol{\mu}_i \in \mathbb{R}^K$, and the probability of a tie between a pair ij is a function of the negative Euclidean distance between them: $-|\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j|$. Hoff et al. (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 & 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 $|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j|$ is small but also that $|\boldsymbol{\mu}_i - \boldsymbol{\mu}_l| \approx |\boldsymbol{\mu}_j - \boldsymbol{\mu}_l|$, 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).[18]

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 & Ward (2004), but this approach is motivated by an eigenvalue decomposition of a network.[19]

[17] At 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.

[18] Hoff (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.

[19] An 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 et al. (2002).

The motivation for this alternative framework stems from the fact that many real networks exhibit varying degrees of stochastic equivalence and homophily. In 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, $\boldsymbol{\mu}_i = \{\mu_{i,1}, \ldots, \mu_{i,K}\}$, 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 $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ 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, $\boldsymbol{\mu}_i \approx \boldsymbol{\mu}_j$, 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 that we estimate a rank-one latent factor model (i.e., $K = 1$), in this case $\boldsymbol{\mu}_i$ is represented by a scalar $\mu_{i,1}$, similarly, $\boldsymbol{\mu}_j = \mu_{j,1}$, 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,[20] here actors in the network have a vector of latent characteristics to describe their behavior as a sender, denoted by $\boldsymbol{\mu}$, and as a receiver, $\mathbf{v}$: $\boldsymbol{\mu}_i, \mathbf{v}_j \in \mathbb{R}^K$ (Hoff, 2009). These again can alter the probability, or in the continuous case value, of an interaction between ij additively: $\boldsymbol{\mu}_i^T \mathbf{D} \mathbf{v}_j$, 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

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

effects probit framework is possible through the addition of a term that captures third order interdependencies. In the **latentnet** package this is done by directly incorporating $|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j|$ as follows: $\theta_{ij} = \boldsymbol{\beta}^T \mathbf{X}_{ij} - |\boldsymbol{\mu}_i - \boldsymbol{\mu}_j|$.[21] 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.

A motivation this paper began with 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:

[21]The **latentnet** package also allows for the specification of a bilinear latent space that is closely related to the projection model introduced in Hoff et al. (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.

$$
\begin{aligned}
y_{ij} &= g(\theta_{ij}) \\
\theta_{ij} &= \boldsymbol{\beta}^T \mathbf{X}_{ij} + e_{ij} \\
e_{ij} &= a_i + b_j + \epsilon_{ij} + \alpha_{\boldsymbol{\mu}_i, \mathbf{v}_j} \text{ , where} \\
\alpha_{\boldsymbol{\mu}_i, \mathbf{v}_j} &= \boldsymbol{\mu}_i^T \mathbf{D} \mathbf{v}_j = \sum_{k \in K} d_k \mu_{ik} v_{jk}
\end{aligned} \tag{6}
$$

Using this framework, we are able to model the dyadic observations as conditionally independent given $\boldsymbol{\theta}$, while $\boldsymbol{\theta}$ depends on the the unobserved random effects, $\mathbf{e}$. $\mathbf{e}$ 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:

- $\boldsymbol{\theta}$: Latent Gaussian variables
- $\boldsymbol{\beta}$: Nodal and/or dyadic regression coefficients
- $\{(a_i, b_i)\} \in \{i = 1, \ldots, n\}$: Nodal random effects
- Σ_{ab}, Σ_ϵ: Network covariance

To arrive at posterior values for these parameters we iteratively simulate from their full conditional distributions:[22]

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

[22]Further details on this process can be found in Hoff (2005).

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

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: $\mathbf{M} = \mathbf{U}^T\mathbf{D}\mathbf{V}$.[23] 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 we have already included in the model. $\mathbf{U}$ and $\mathbf{V}$ are $n \times K$ matrices with orthnormal columns and $\mathbf{D} = diag\{d_1, \ldots, d_K\}$. An MCMC scheme that can be used to construct an empirical distribution around these parameters involves iterating through $k \in 1, \ldots, K$:

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

Taken together the additive effects portion of AMEN (described by the SRM) and the multiplicative effects (described by the latent factor model) provide a modelling 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,

[23] Framing 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 & Fischer (2014)

Ingold & Leifeld (2014)

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

Krivitsky & Handcock (2015)

2.1. Parameter Estimates.

2.2. 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 et al. (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

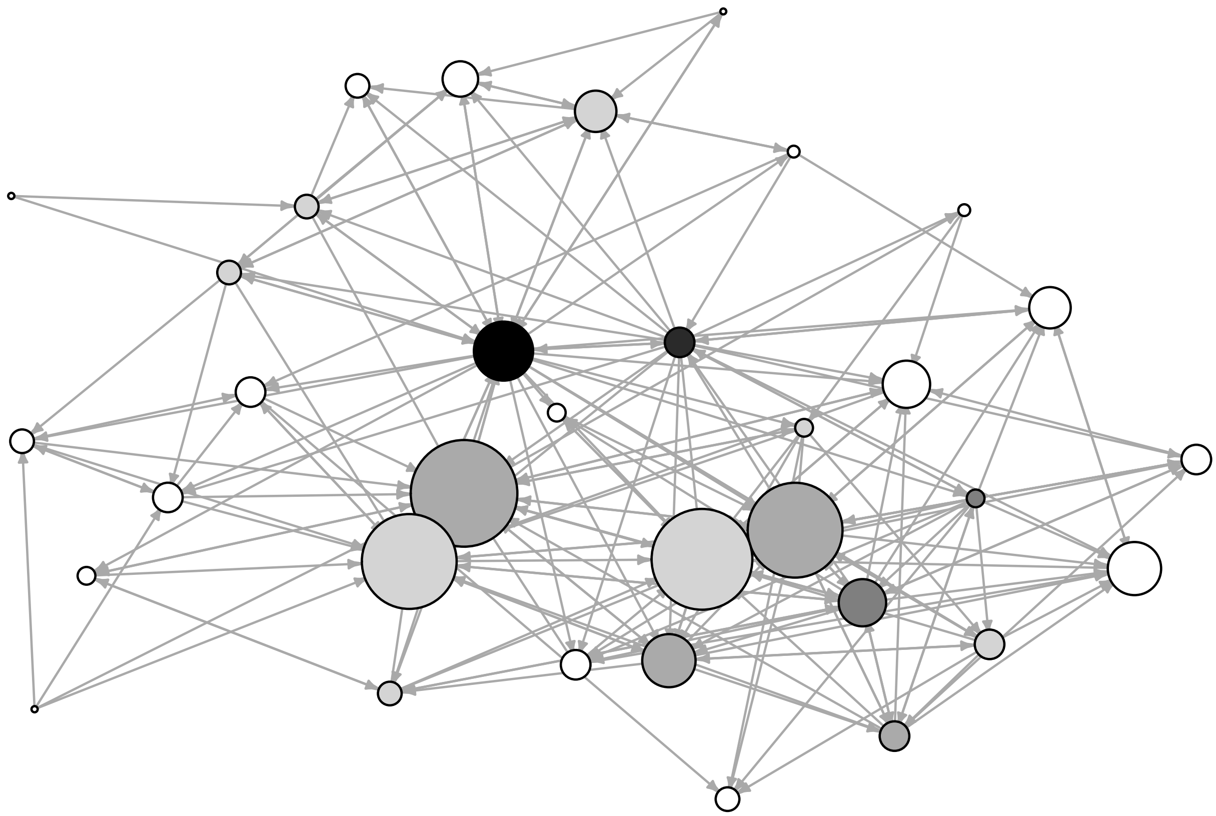

Figure 2. dv net

- Geodesic distances - The proportion of pairs of nodes whose shortest connecting path is of length k, for $k = 1, 2, \ldots$ 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 4 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.

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

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

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

2.3. **Tie Formation Prediction.** The results are displayed in Figure 5 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 should have all actual (dark blue) observations towards the right of the separation plot (Greenhill et al., 2011).

Beger (2015)

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.

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

Table 3. * $p < 0.05$ (or 0 outside the 95% confidence interval).

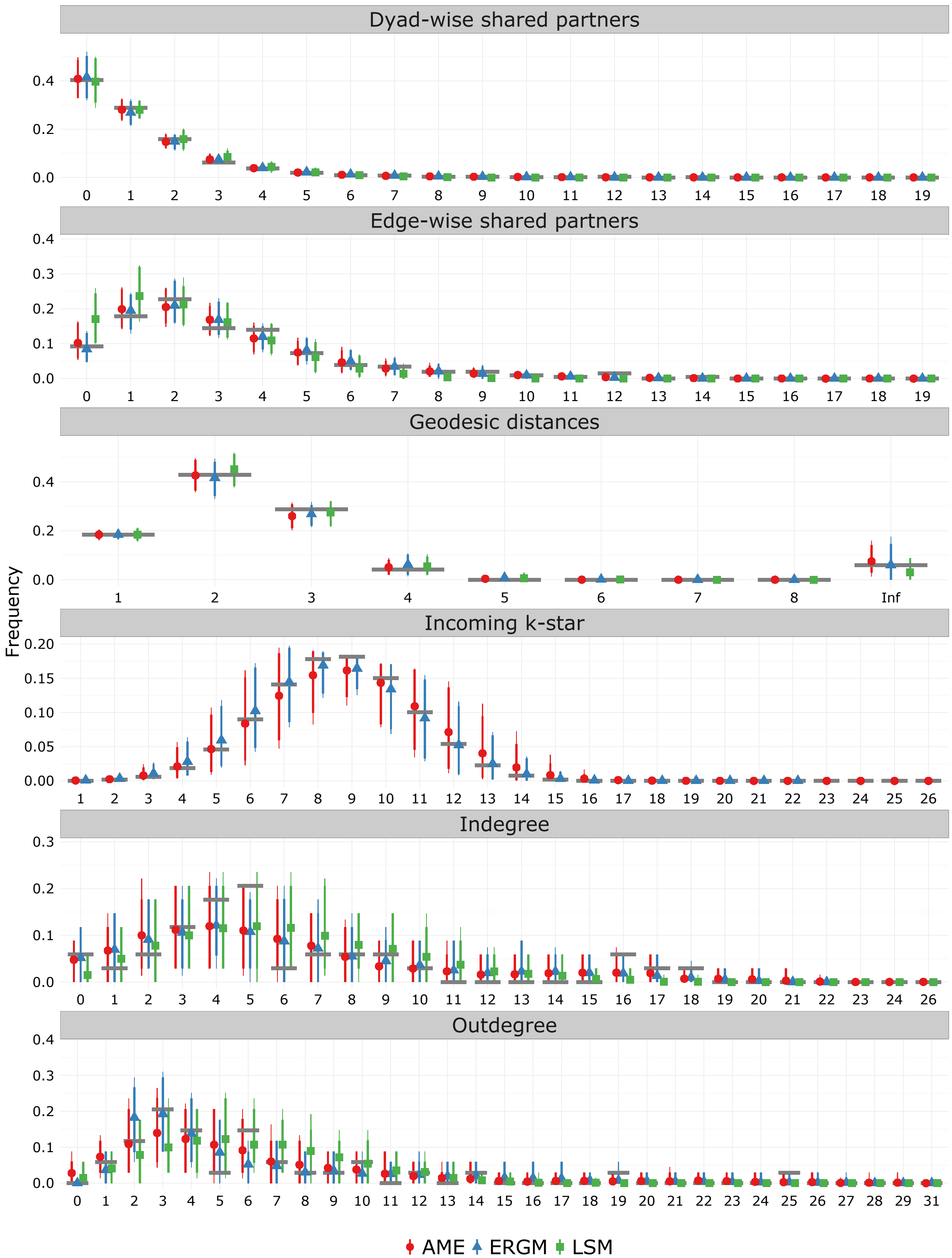

Figure 3. network stats

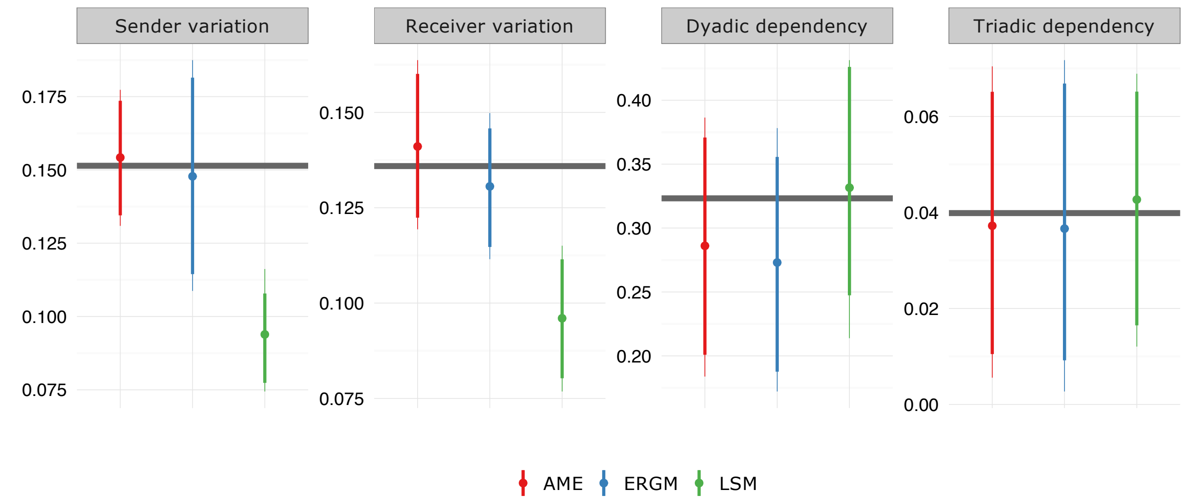

Figure 4. Posterior predictive goodness of fit summary

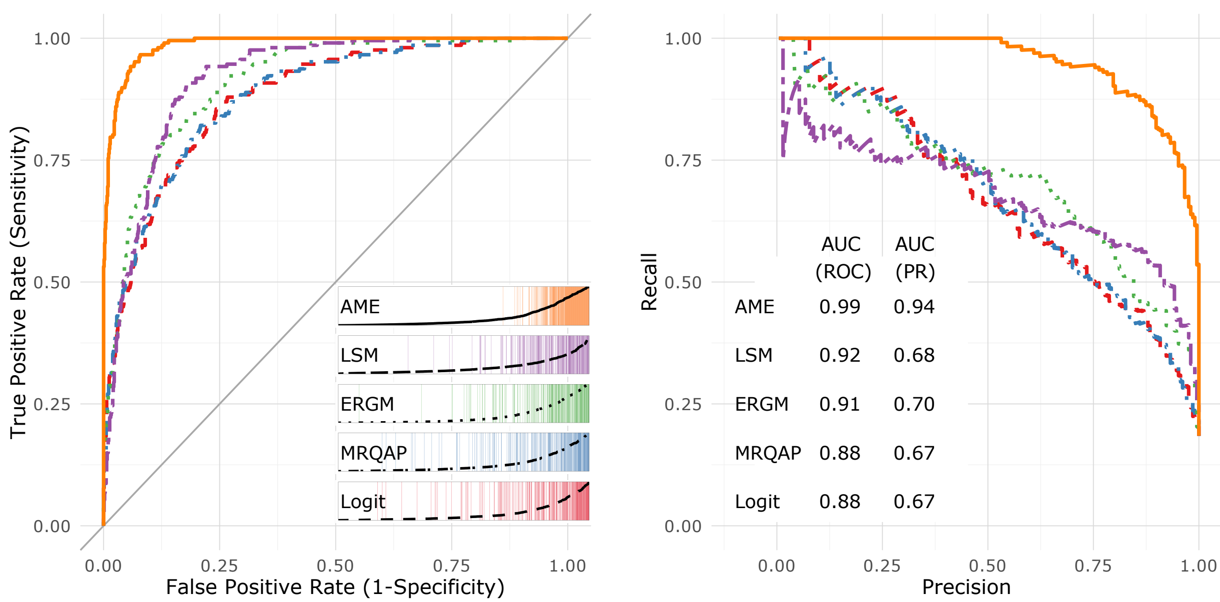

Figure 5. ROC and separation plots

3. Conclusion

4. Appendix

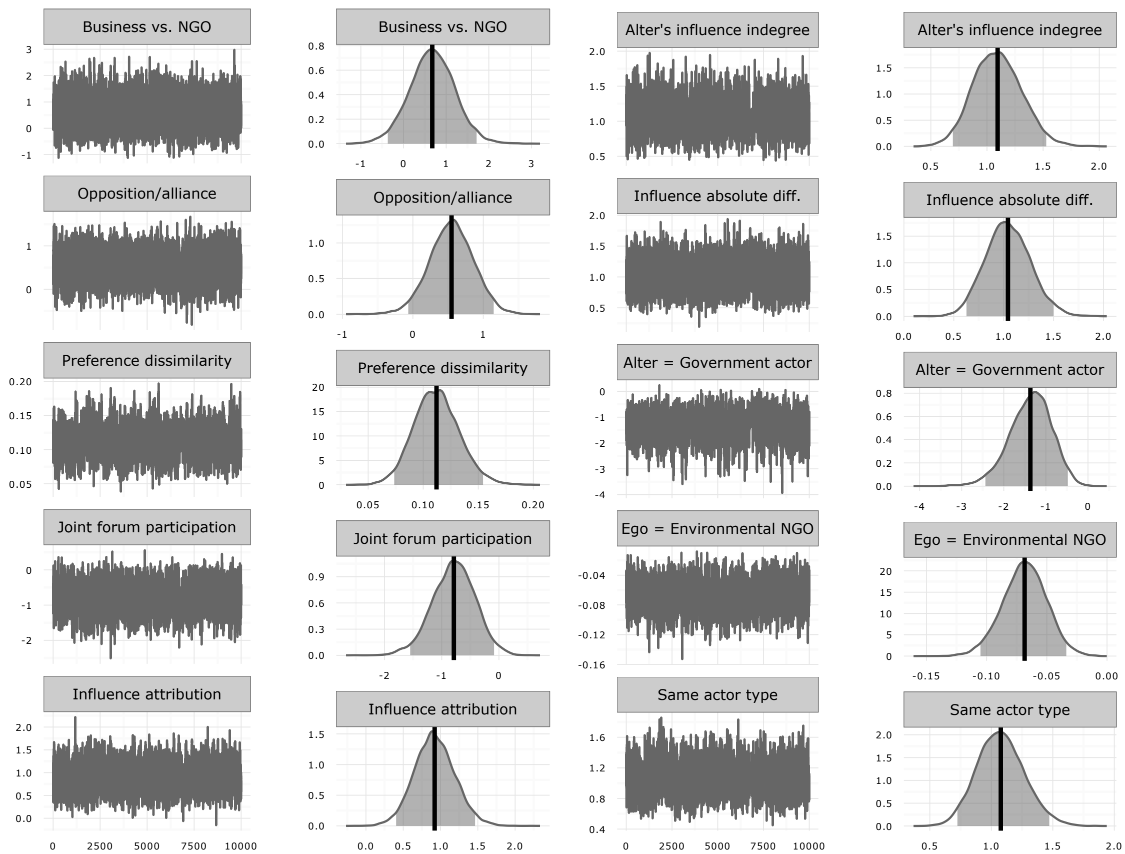

Figure 6. ame convergence k = 2

4.1. AMEN Model Convergence.

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

Table 4. * $p < 0.05$ (or 0 outside the 95% confidence interval).

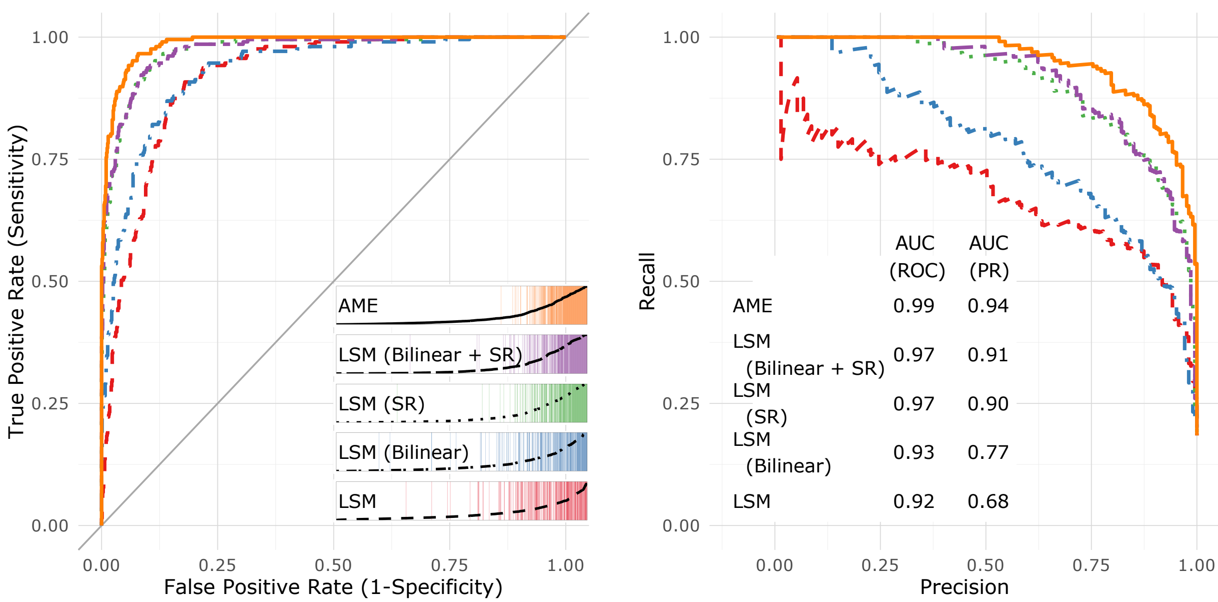

Figure 7. ROC and separation plots

4.2. Comparison of amen & latentnet $\mathcal{R}$ Packages.

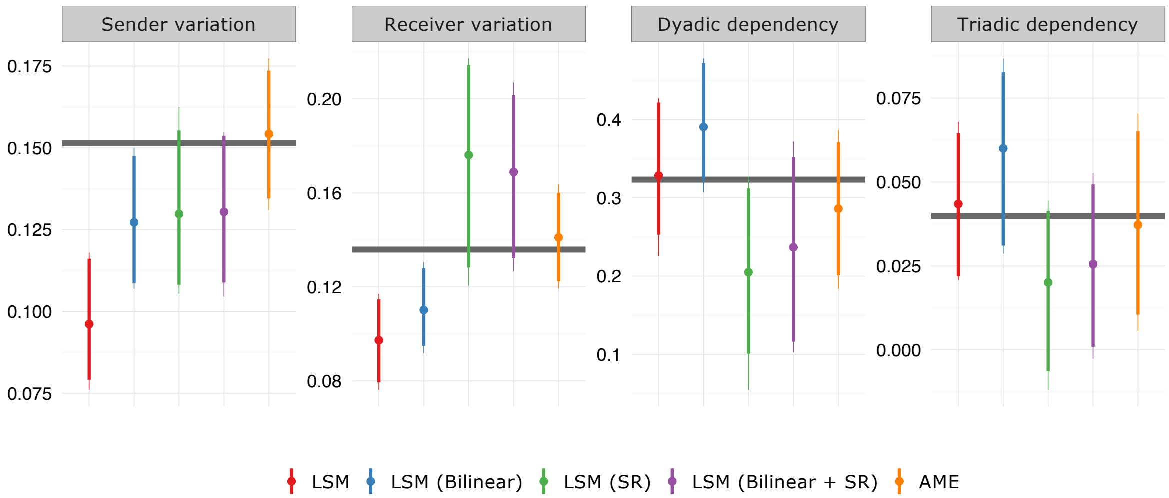

Figure 8. Posterior predictive goodness of fit summary

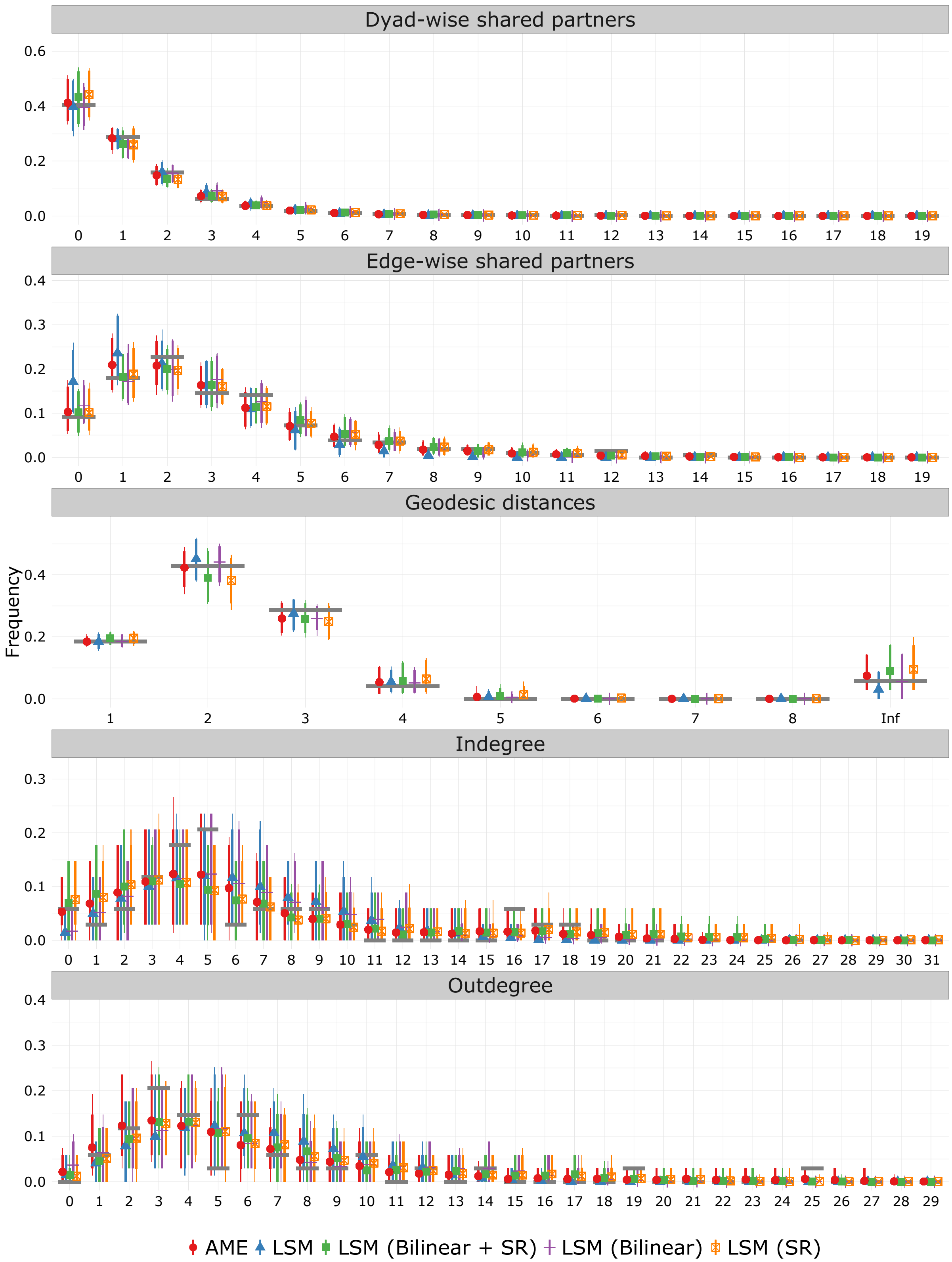

Figure 9. network stats

	AME (k=0)	AME (k=1)	AME (k=2)	AME (k=3)	AME (k=4)
Intercept/Edges	-2.75*	-3.08*	-3.39*	-3.72*	-3.93*
	[-3.43; -2.09]	[-3.91; -2.30]	[-4.38; -2.50]	[-4.84; -2.73]	[-5.12; -2.87]
Conflicting policy preferences					
Business vs. NGO	-1.08*	-1.28*	-1.37*	-1.48*	-1.51*
	[-1.82; -0.41]	[-2.20; -0.47]	[-2.44; -0.47]	[-2.63; -0.49]	[-2.69; -0.47]
Opposition/alliance	0.83*	0.95*	1.08*	1.19*	1.28*
	[0.57; 1.10]	[0.64; 1.27]	[0.72; 1.47]	[0.80; 1.64]	[0.86; 1.77]
Preference dissimilarity	-0.49	-0.65*	-0.79*	-0.89*	-0.95*
	[-1.06; 0.06]	[-1.30; -0.03]	[-1.55; -0.08]	[-1.71; -0.12]	[-1.80; -0.14]
Transaction costs					
Joint forum participation	0.73*	0.84*	0.92*	1.01*	1.06*
	[0.34; 1.12]	[0.38; 1.31]	[0.40; 1.47]	[0.44; 1.62]	[0.43; 1.72]
Influence					
Influence attribution	0.88*	1.00*	1.09*	1.21*	1.28*
	[0.57; 1.19]	[0.63; 1.39]	[0.69; 1.53]	[0.75; 1.71]	[0.80; 1.84]
Alter's influence indegree	0.09*	0.10*	0.11*	0.12*	0.13*
	[0.06; 0.12]	[0.07; 0.14]	[0.07; 0.15]	[0.08; 0.17]	[0.09; 0.18]
Influence absolute diff.	-0.06*	-0.06*	-0.07*	-0.07*	-0.08*
	[-0.08; -0.03]	[-0.10; -0.03]	[-0.11; -0.03]	[-0.12; -0.04]	[-0.12; -0.04]
Alter = Government actor	0.49	0.52	0.55	0.60	0.64
	[-0.01; 0.99]	[-0.04; 1.07]	[-0.07; 1.15]	[-0.07; 1.27]	[-0.07; 1.35]
Functional requirements					
Ego = Environmental NGO	0.54	0.61	0.67	0.76	0.80
	[-0.28; 1.36]	[-0.31; 1.56]	[-0.38; 1.71]	[-0.38; 1.90]	[-0.40; 2.04]
Same actor type	0.88*	0.97*	1.04*	1.11*	1.17*
	[0.55; 1.21]	[0.60; 1.35]	[0.63; 1.50]	[0.64; 1.59]	[0.68; 1.68]

Table 5. * $p < 0.05$ (or 0 outside the 95% confidence interval).

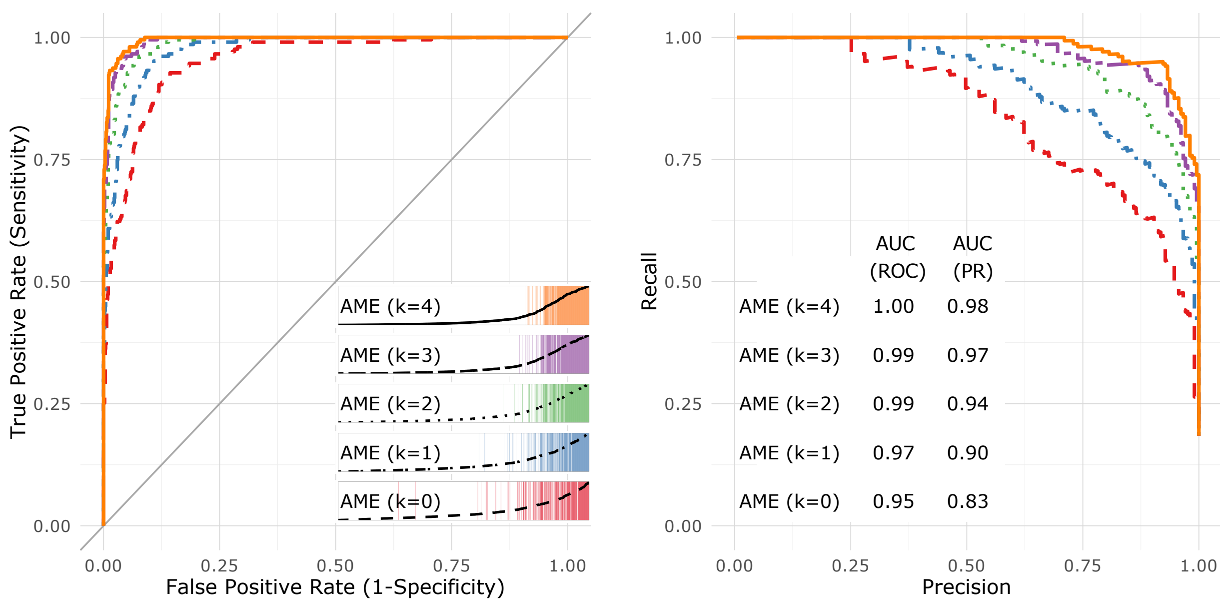

Figure 10. ROC and separation plots

4.3. Comparison with other AME Parameterizations.

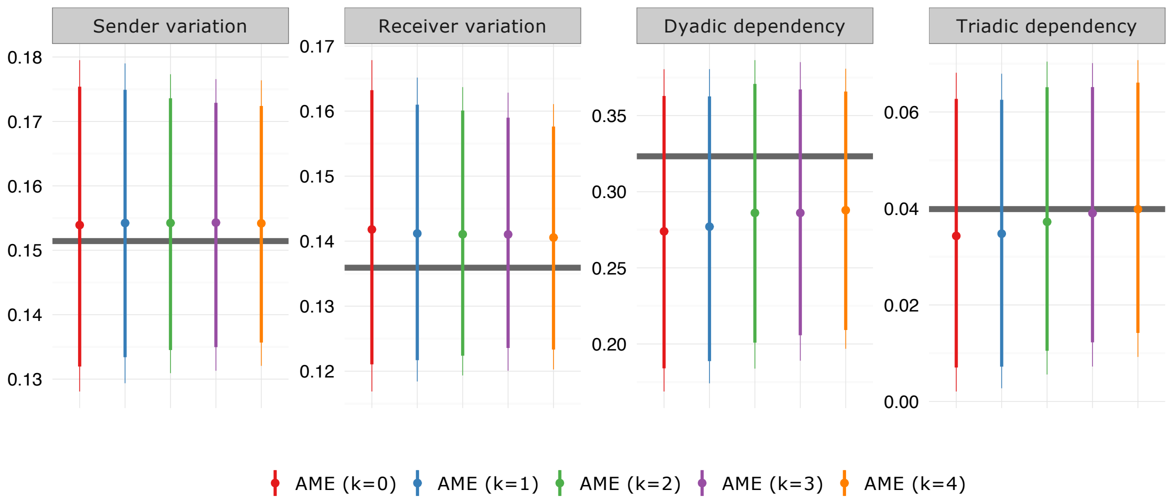

Figure 11. Posterior predictive goodness of fit summary

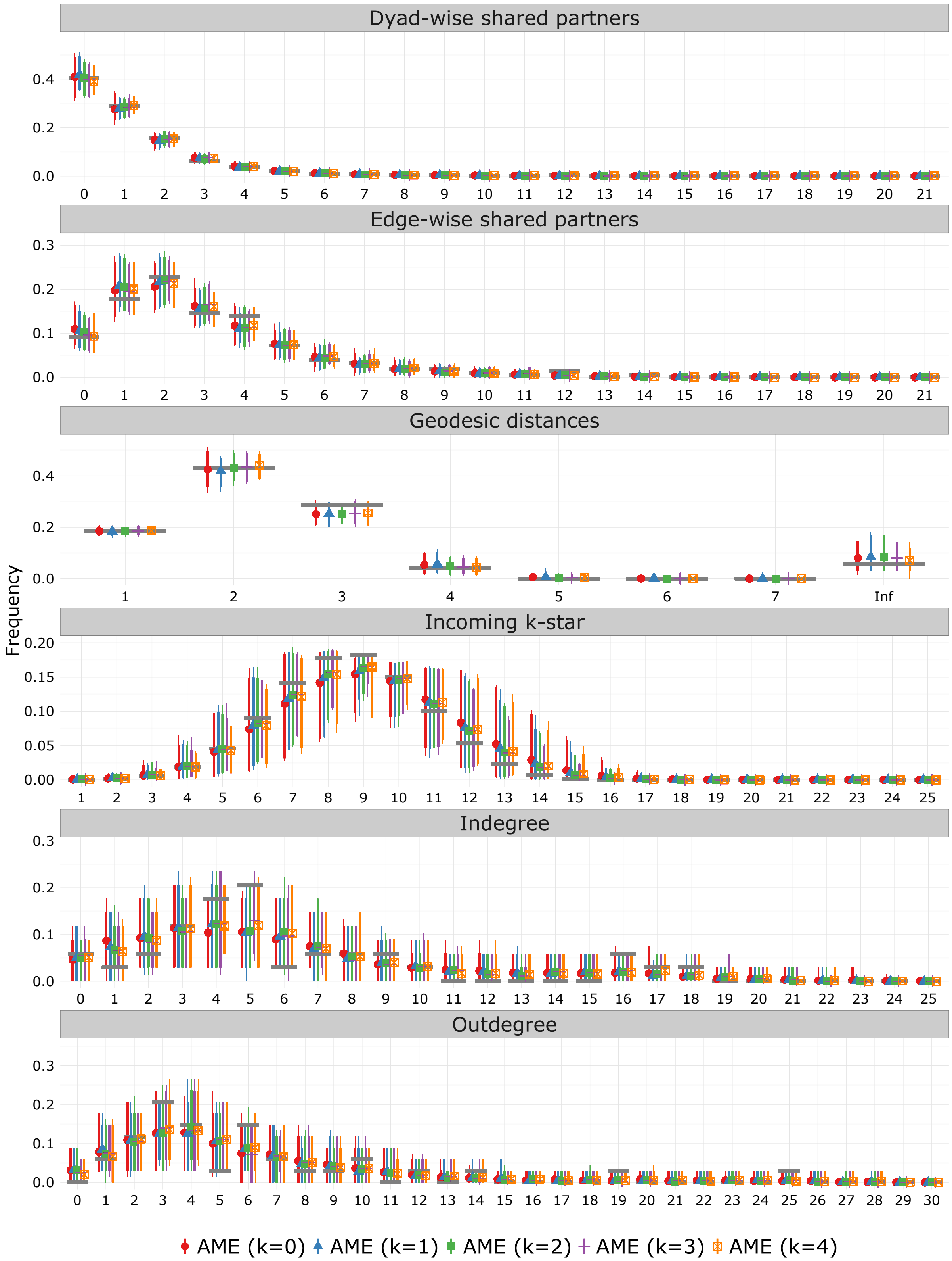

Figure 12. network stats

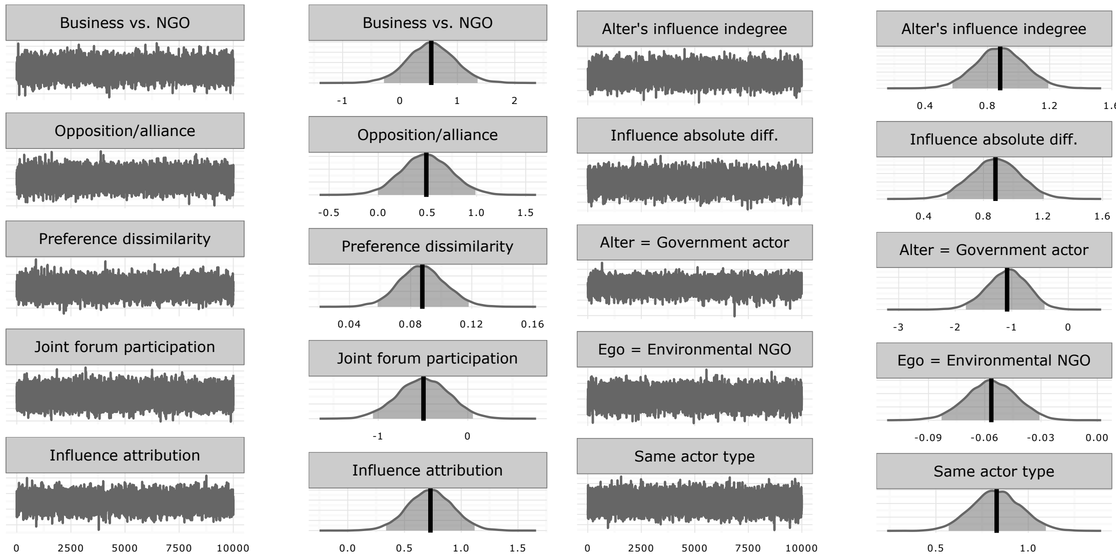

Figure 13. ame convergence k = 0

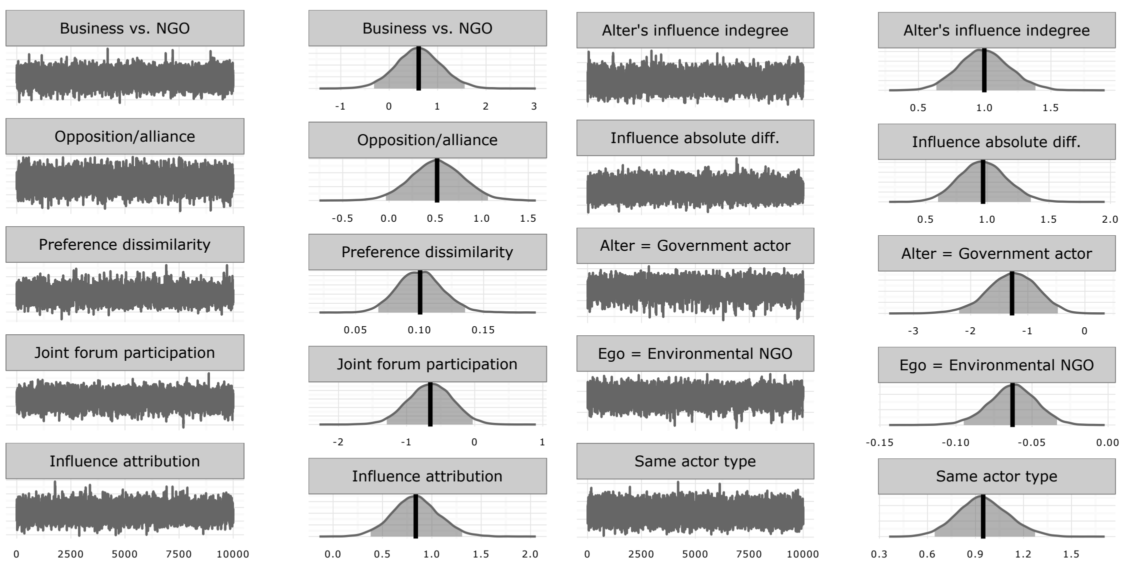

Figure 14. ame convergence k=1

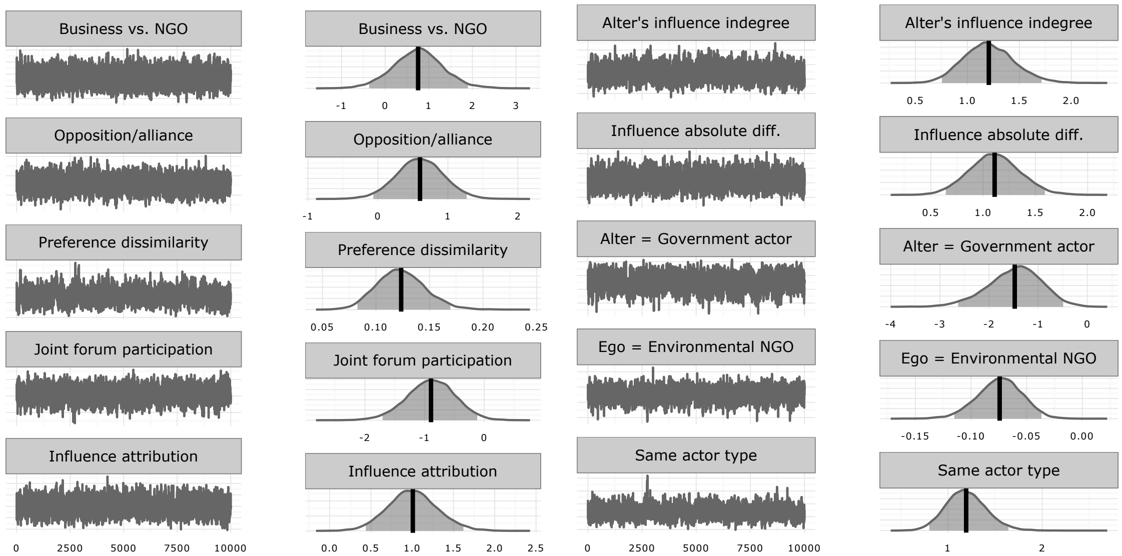

Figure 15. ame convergence k=3

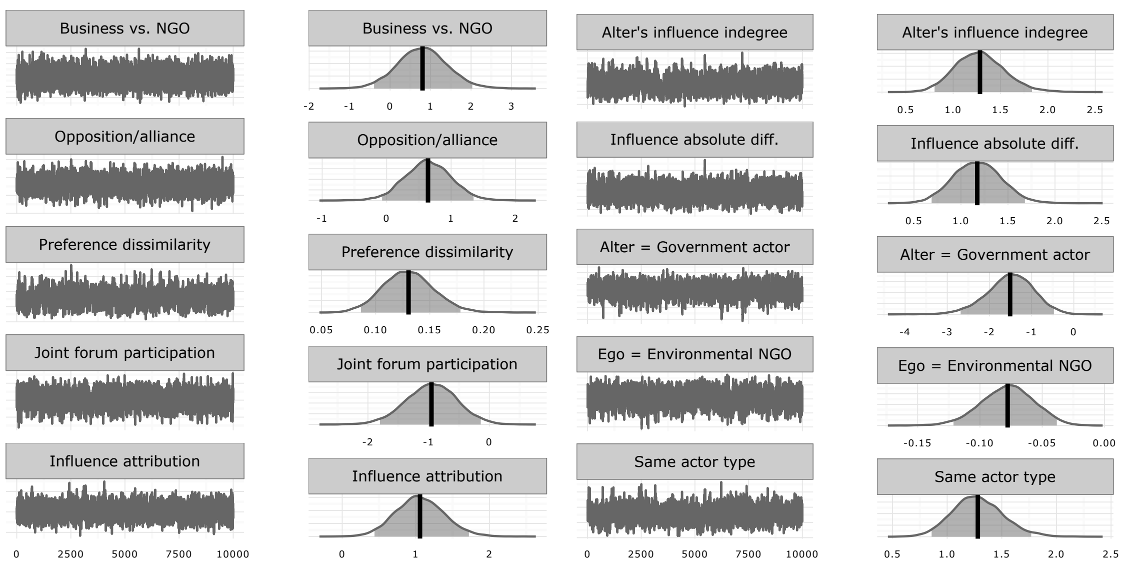

Figure 16. ame convergence k = 4

References

Anderson, Carolyn J; Stanley Wasserman & Katherine Faust (1992) Building stochastic blockmodels. *Social networks* 14(1): 137–161.

Beck, Nathaniel & 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 & 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.

Beger, Andreas (2015) Precision-recall curves. Blog post (http://andybeger.com/2015/03/16/precision-recall-curves/).

Bhamidi, Shankar; Guy Bresler & Allan Sly (2008). Mixing time of exponential random graphs. In: Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on , 803–812. IEEE.

Bolton, Gary E; Jordi Brandts & 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 & Thomas Erlebach (2005) *Network analysis: methodological foundations* volume 3418. Springer Science & Business Media.

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 & Persi Diaconis (2013) Estimating and understanding exponential random graph models. *The Annals of Statistics* 41(5): 2428–2461.

Choucri, Nazli & Robert C North (1972) Dynamics of International Conflict. *World Politics* 24(2): 80–122.

Cox, James C; Daniel Friedman & Steven Gjerstad (2007) A tractable model of reciprocity and fairness. *Games and Economic Behavior* 59(1): 17–45.

Cranmer, Skyler & Bruce A Desmarais (2011) Inferential Network Analysis with Exponential Random Graph Models. *Political Analysis* 19(1): 66–86.

Cranmer, Skyler J; Tobias Heinrich & Bruce A Desmarais (2014) Reciprocity and the structural determinants of the international sanctions network. *Social Networks 36*: 5–22.

Cranmer, Skyler J; Philip Leifeld; Scott D McClurg & Meredith Rolfe (2016) Navigating the range of statistical tools for inferential network analysis. *American Journal of Political Science*.

Dafoe, Allan (2011) Statistical Critiques of the Democratic Peace: Caveat Emptor. *American Journal of Political Science* 55(2): 247–262.

Diehl, Paul F & 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*: 828–851.

Dorff, Cassy & Shahryar Minhas (2016) When do states say uncle? Network dependence and sanction compliance. *International Interactions*.

Dorff, Cassy & Michael D Ward (2013) Networks, dyads, and the social relations model. *Political Science Research and Methods* 1(02): 159–178.

Erikson, Robert S; Pablo M Pinto & Kelly T Rader (2014) Dyadic analysis in international relations: A cautionary tale. *Political Analysis* 22(4): 457–463.

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

Franzese, Robert & Jude C Hayes (2007) Spatial econometric models for the analysis of TSCS data in political science. *Political Analysis* 15(15): 2.

Fuhrmann, Matthew & 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 & 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.

Goodreau, Steven M; Mark S Handcock; David R Hunter; Carter T Butts & 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; Michael D Ward & 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 & 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 & Martina Morris (2008) statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software* 24(1): 1548.

Hoff, Peter (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In: Advances in Neural Information Processing Systems , 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 & Yanjun He (2015). amen: Additive and Multiplicative Effects Models for Networks and Relational Data. R package version 1.1.

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 & Mark S Handcock (2002) Latent space approaches to social network analysis. *Journal of the American Statistical association* 97(460): 1090–1098.

Hoff, Peter D & 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 & 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 & 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.

Ingold, Karin & 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 & 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*: muu043.

Jackson, Matthew (2014) Networks in the understanding of economic behaviors. *The Journal of Economic Perspectives* 28(4): 3–22.

Jones, Zachary; Matthew Denny; Bruce Desmarais & Hannah Wallach (2016) Inference on the effects of observed features in latent space models for networks. Presented at the Summer Meetings of the Political Methodology Society.

Kenny, David A; Deborah A Kashy & William L Cook (2006) *Dyadic data analysis*. 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.

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

Krivitsky, Pavel N & 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.

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 & William Reed (2001) War and rivalry among great powers. *American Journal of Political Science* 45(2): 457–469.

Li, Heng & Eric Loken (2002) A unified theory of statistical analysis and inference for variance component models for dyadic data. *Statistica Sinica*: 519–535.

Lubbers, Miranda J & Tom A 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.

Manger, Mark S; Mark A Pickup & Tom A 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 & B. P 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 & Ilan Talmud (2006) Structural Equivalence and International Conflict: A Social Networks Analysis. *Journal of Conflict Resolution* 50(5): 664–689.

Metternich, Nils; Shahryar Minhas & 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 & 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 M (2002) A kantian system? democracy and third-party conflict resolution. *American Journal of Political Science*: 749–759.

Morris, Martina; Mark S Handcock & 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 & Tom A. B Snijders (2001) Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* 96(455): 1077–1987.

Pattison, Philippa & Stanley Wasserman (1999) Logit models and logistic regressions for social networks. ii. multivariate relations. *British Journal of Mathematical and Statistical Psychology 52*: 169–194.

Réka, Albert; Hawoong Jeong & A.-L Barabási (1999) The diameter of the www. *Nature* 401(6749): 130–131.

Richardson, Lewis F (1960) *Arms and Insecurity*. Quadrangle.

Robins, Garry; Tom Snijders; Peng Wang; Mark Handcock & 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 R & 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 (1996) Stochastic actor-oriented models for network change. *Journal of mathematical sociology* 21(1-2): 149–172.

Snijders, Tom A (2002) Markov chain monte carlo estimation of exponential random, graph models. *Journal of Social Structure* 3(2): via web:, zeeb.library.cmu.edu: 7850/JoSS/snijders/Mcpstar.pdf.

Snijders, Tom A (2011) Statistical models for social networks. *Annual Review of Sociology 37*: 131–53.

Snijders, Tom A. B; Philippa E Pattison; Garry L Robins & Mark S Handcock (2006) New specifications for exponential random graph models. *Sociological Methodology* 36(1): 99–153.

Strauss, Davis & Michael Ikeda (1990) Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association 85*: 204–212.

Van Duijn, Marijtje A; Krista J Gile & Mark S Handcock (2009) A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks* 31(1): 52–62.

Ward, Michael D & John S Ahlquist (2010) Maximum Likelihood Strategies for Social Science, version 0.8. pdf typescript, Duke University.

Ward, Michael D & Kristian S Gleditsch (2008) *Spatial Regression Models* volume 155. Sage.

Ward, Michael D; Randolph M Siverson & Xun Cao (2007) Disputes, democracies, and dependencies: A reexamination of the kantian peace. *American Journal of Political Science* 51(3): 583–601.

Warner, Rebecca; David Kenny & Michael Stoto (1979) A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology 37*: 1742–1757.

Wasserman, Stanley & 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: pdhoff@duke.edu

MICHAEL D. WARD: DEPARTMENT OF POLITICAL SCIENCE

Current address: Duke University

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