

# A Dynamic Additive and Multiplicative Effects Model with Application to the United Nations Voting Behaviors

Bomin Kim<sup>1</sup>, Xiaoyue Niu<sup>1</sup>, David Hunter<sup>1</sup>, and Xun Cao<sup>2</sup>

<sup>1</sup>Department of Statistics, The Pennsylvania State University

<sup>2</sup>Department of Political Science, The Pennsylvania State University

## Abstract

In this paper, we introduce a statistical regression model for discrete-time networks that are correlated over time. Our model is a dynamic version of a Gaussian additive and multiplicative effects (DAME) model which extends the latent factor network model of Hoff (2009) and the additive and multiplicative effects model of Minhas et al. (2016a), by incorporating the temporal correlation structure into the prior specifications of the parameters. The temporal evolution of the network is modeled through a Gaussian process (GP) as in Durante and Dunson (2013), where we estimate the unknown covariance structure from the dataset. We analyze the United Nations General Assembly voting data from 1983 to 2014 (Voeten et al., 2016) and show the effectiveness of our model at inferring the dyadic dependence structure among the international voting behaviors as well as allowing for a varying number of nodes over time. Overall, the DAME model shows significantly better fit to the dataset compared to alternative approaches. Moreover, after controlling for other dyadic covariates such as geographic distances and bilateral trade between countries, the model-estimated additive effects, multiplicative effects, and their movements reveal interesting and meaningful foreign policy positions and alliances of various countries.

## 1 Introduction

In recent decades, social network analysis has been well-established and is widely used in a variety of applications, ranging from friendship and collaboration networks to disease transmission. Because a network naturally evolves over time, there has been a growing need for methods of modeling networks that change over time. A number of models have been suggested that are extensions of static network models, such as the temporal exponential random graph model (TERGM) (Hanneke et al., 2010) and the dynamic stochastic blockmodel (Xu and Hero III, 2013), or new models for network dynamics, such as the stochastic actor oriented model (SAOM) (Snijders et al., 2010). On the other hand, there are dynamic network models that give consideration to the unobserved latent space (Hoff et al., 2002)—the structure of the network that is not explained through the use of exogenous node and dyad covariates. To provide novel insights into the latter, we extend existing latent factor models (Hoff, 2005, 2009; Hoff et al., 2014; Minhas et al., 2016a) and develop the dynamic additive and multiplicative effects model (DAME) for discrete-time networks, with emphasis on the latent structures—unmeasured attributes of nodes for tie formations and their changes over time.

Hoff et al. (2002) introduces a class of models where the probability of a relation between actors depends on the positions of individuals in an unobserved “social space”. There are two specifications in the latent space model: (i) “the latent distance model” which is built upon the latent Euclidean space; and (ii) “the latent factor model” which stems from the projection model. Specifically, there are several versions of the latent projection model in which the probability of a tie between nodes  $i$  and  $j$  is determined by the normalized inner product of the two nodes’ latent positions  $f(\mathbf{v}_i, \mathbf{v}_j) = (\mathbf{v}_i' \mathbf{v}_j) / |\mathbf{v}_j|^{-1}$ . Hoff (2005) introduces the symmetric multiplicative interaction effect  $(\mathbf{v}_i' \mathbf{v}_j)$  into the network generalized linear model in order to capture third-order dependence patterns—often described by the three features, transitivity, balance, and clusterability. Hoff (2008)

parameterizes the multiplicative effects via eigendecomposition  $\mathbf{u}_i^T \mathbf{A} \mathbf{u}_j$ , and demonstrates that the latent eigenmodel is able to represent a wide array of patterns in the data due to its unrestricted low-rank approximation to the symmetric relational data. Hoff (2009) extends the framework to model asymmetric social networks using the singular value decomposition  $\mathbf{u}_i^T \mathbf{D} \mathbf{v}_j$ . Finally, Hoff et al. (2014) and Minhas et al. (2016a) combine the additive and multiplicative effects to model the second-order (or reciprocity) and third-order dependencies and establish the AME—additive and multiplicative effects—regression model for dyadic response data  $y_{ij}$ , in which

$$\begin{aligned} y_{ij} &= g(\theta_{ij}), \\ \theta_{ij} &= \boldsymbol{\beta}^T \mathbf{X}_{ij} + e_{ij}, \text{ and} \\ e_{ij} &= a_i + b_j + \epsilon_{ij} + \alpha(\mathbf{u}_i, \mathbf{v}_j), \end{aligned} \tag{1}$$

where  $\alpha(\mathbf{u}_i, \mathbf{v}_j) = \mathbf{u}_i^T \mathbf{D} \mathbf{v}_j$  and  $g(\theta_{ij})$  is a link function used to model various types of relational matrices such as continuous, binary, ordinal, or rank-based responses.

As dynamic network analysis has become an emergent scientific field, there has been a growing number of dynamic network models that incorporate the latent space models in the last decade (Kim et al., 2017). For instance, latent distance models have been a strong motivation for various dynamic network models proposed by many authors (Sarkar and Moore, 2005; Sarkar et al., 2007; Sewell and Chen, 2015, 2016; Friel et al., 2016), providing ample references for the reader interested in these specific problems. On the other hand, a comparatively small literature employs the AME framework to develop dynamic network models. Ward and Hoff (2007) first introduce the concept of dynamic latent factors, and this idea is expanded by Ward et al. (2013) to analyze bilateral trade using the generalized bi-linear mixed effect model (Hoff, 2005). These models allow time-varying parameters for edge covariates and latent factors to extensively investigate the temporal evolution of networks. More recently, He and Hoff (2017) develop a coevolution model for the analysis of longitudinal network and nodal attribute data, including latent nodal attributes. This multiplicative coevolution regression (MCR) model provides the benefit of allowing nodes to change their nodal (or latent) attributes  $\mathbf{X}_t$  depending on their past relations as well as the evolution of the network  $\mathbf{Y}_t$ , however, the contagion of the network is limited to a first-order autoregressive, or AR(1), model. There also exists a series of papers for modeling a tensor representation of network or multi-way array (Hoff, 2011; Hoff et al., 2011; Hoff, 2015; Minhas et al., 2016b), with longitudinal networks serving as an example of the general model. Still, the latent variable structure is not the main focus of those papers.

Whereas the earlier works rely on Markovian assumptions—i.e., the network at the next timestep depends only on the network at present state, not on the sequence of networks that preceded it—there are dynamic latent space models which relax the Markovian assumption and instead take advantage of temporal dependence in the longer history. Focusing on the temporal aspect of networks, Durante and Dunson (2013) proposed the dynamic latent space model for binary symmetric matrices, which assumes that the latent factors are evolving in continuous time via Gaussian processes (GP). Precisely, the model formulation follows

$$\begin{aligned} y_{ij,t} | \pi_{ij}(t) &\sim \text{Bern}(\pi_{ij}(t)), \\ \pi_{ij}(t) &= \frac{1}{1 + e^{-s_{ij}(t)}}, \\ s_{ij}(t) &= \mu(t) + \mathbf{x}_i(t)' \mathbf{x}_j(t) \end{aligned} \tag{2}$$

for  $i < j$ , with  $\mathbf{x}_{ih}(\cdot) \sim \text{GP}(0, \tau_h^{-1} c_X)$  and  $\mu(\cdot) \sim \text{GP}(0, c_\mu)$ , where  $\mathbf{x}_i(t) = [x_{i1}(t), \dots, x_{iH}(t)]'$  for  $i = 1, \dots, V$  are the latent vectors of node  $i$ . The variance multipliers  $\tau_h^{-1}$  for  $h = 1, \dots, H$  are shrinkage parameters with a gamma prior, and  $c_\mu$  and  $c_x$  are the squared exponential functions  $c_X(t, t') = \exp(-\kappa_X \|t - t'\|_2^2)$  and  $c_\mu(t, t') = \exp(-\kappa_\mu \|t - t'\|_2^2)$ . This modeling approach is based on nonparametric Bayesian inference and has the advantage of learning the number of latent dimensions  $H$  in the model (Bhattacharya and Dunson, 2011). In addition, the non-Markovian property of the model allows networks with unequal time intervals. Although the model has been successfully applied to different types of longitudinal networks (Durante and Dunson, 2014a,b), it lacks several

benefits of the AME model. First, model (2) does not include the additive effects  $a_i$  and  $b_j$ , which can capture significant heterogeneity in activity levels across nodes. Second, model (2) uses the term  $\mathbf{x}_i(t)' \mathbf{x}_j(t)$  to represent multiplicative latent factor effects. However, the form  $\mathbf{u}_i(t)^T \mathbf{D}(t) \mathbf{u}_j(t)$  used in the AME model allows the flexibility of negative eigenvalues  $d_r(t)$ . According to Hoff (2008), the parameterization of  $\mathbf{u}_i(t)^T \mathbf{D}(t) \mathbf{u}_j(t)$  can represent both positive or negative transitivity in varying degrees; on the other hand, the parameterization of  $\mathbf{v}_i(t)^T \mathbf{v}_j(t)$  is not able to explain negative transitivity or stochastic equivalence, where nodes with the same or similar latent vectors do not have strong relationships with one another.

Here we propose a new model, the dynamic additive and multiplicative effects (DAME) model, combining the advantages of the AME model and the dynamic latent space model. We use the same formulation as the AME model, while incorporating the time-varying prior structures of Durante and Dunson (2013, 2014a,b). Additionally, the DAME model employs two innovations. First, we learn the temporal correlation of networks by estimating Gaussian process length parameters instead of using fixed covariance structure. Our method does not require any initial guess about correlations, and it further enables efficient estimation of the fixed and random effect parameters. Second, in order to increase flexibility and accuracy of the model, the DAME model allows the number of nodes to change over time by allowing a special case of missing values, which is referred to as “structural missingness” in this paper. In what follows, we introduce the DAME model by describing how we take advantage of temporal correlation and deriving the sampling equations for hierarchical Bayesian inference (Section 2), present simulation studies to show some advantages of the DAME model over alternative approaches (Section 3), and apply the DAME model to the United Nations General Assembly voting network (Section 4).

## 2 Dynamic Additive and Multiplicative Effects Model

### 2.1 Model Formulation

Our main goal is to simultaneously model the sequence of  $N \times N$  time-varying symmetric matrices  $\mathbf{Y} = \{\mathbf{Y}^1, \dots, \mathbf{Y}^T\}$ , where the entry  $y_{ij}^t$  denotes any relational data corresponding to the node pair  $(i, j)$  at timepoint  $t$ , using the observed covariate arrays  $\mathbf{X} = \{\mathbf{X}^1, \dots, \mathbf{X}^T\}$  where  $\mathbf{X}^t = \{\mathbf{X}_1^t, \dots, \mathbf{X}_P^t\}$ . For  $i = 2, \dots, N, j = 1, \dots, i - 1$ , and  $t = 1, \dots, T$ , we assume

$$y_{ij}^t = \sum_{p=1}^P \beta_p^t X_{ijp}^t + z_{ij}^t, \quad (3)$$

where  $X_{ijp}^t$  is the  $p$ th edge covariate,  $\beta_p^t$  is the corresponding unknown coefficient, and  $z_{ij}^t$  is the unobserved random effect with the additive and multiplicative form

$$z_{ij}^t = \theta_i^t + \theta_j^t + \mathbf{u}_i^{t'} \mathbf{D}^t \mathbf{u}_j^t + \epsilon_{ij}^t, \quad (4)$$

where  $\theta_i^t$  and  $\theta_j^t$  are the node-specific additive random effects,  $\mathbf{D}^t$  denotes the  $R \times R$  diagonal matrix of eigenvalues  $d_1^t, \dots, d_R^t$ , and  $\mathbf{u}_i^t = [u_{i1}^t, \dots, u_{iR}^t]'$  denotes the  $R$ -length vector of latent coordinates of node  $i$ , and  $\epsilon_{ij}^t$  is the random error.

To model the temporal dependence in the networks beyond Markovian assumptions, we adopt the prior specifications in Bhattacharya and Dunson (2011) and Durante and Dunson (2013). Specifically, we assume independent Gaussian process (GP) priors for the parameters  $\boldsymbol{\beta}, \boldsymbol{\theta}$  and  $\mathbf{d}$ :

1. For  $p = 1, \dots, P$ ,

$$\begin{aligned} \boldsymbol{\beta}_p &\sim \mathcal{N}_T(\mathbf{0}, \Sigma_p^\beta), \\ \Sigma_p^\beta &= \tau_p^\beta f(\kappa_p^\beta), \end{aligned}$$

where  $\boldsymbol{\beta}_p = (\beta_p^1, \dots, \beta_p^T)$  is a  $T$ -dimensional vector and  $\Sigma_p^\beta$  is a  $T \times T$  covariance matrix with variance parameter  $\tau_p^\beta$  and range parameter  $\kappa_p^\beta$ .

2. For  $i = 1, \dots, N$ ,

$$\begin{aligned}\boldsymbol{\theta}_i &\sim \mathcal{N}_T(\mathbf{0}, \Sigma^\theta) \\ \Sigma^\theta &= \tau^\theta f(\kappa^\theta),\end{aligned}$$

where  $\boldsymbol{\theta}_i = (\theta_i^1, \dots, \theta_i^T)$  is a  $T$ -dimensional vector and  $\Sigma^\theta$  is a  $T \times T$  covariance matrix with variance parameter  $\tau^\theta$  and range parameter  $\kappa^\theta$ .

3. For  $r = 1, \dots, R$ ,

$$\begin{aligned}\mathbf{d}_r &\sim \mathcal{N}_T(\mathbf{0}, \Sigma_r^d), \\ \Sigma_r^d &= \tau_r^d f(\kappa_r^d),\end{aligned}$$

where  $\mathbf{d}_r = (d_r^1, \dots, d_r^T)$  is a  $T$ -dimensional vector and  $\Sigma_r^d$  is a  $T \times T$  covariance matrix with variance parameter  $\tau_r^d$  and range parameter  $\kappa_r^d$ .

The key part of the Gaussian process is the formulation of covariance matrices  $\Sigma^\beta$ ,  $\Sigma^\theta$ , and  $\Sigma^d$ . Among a number of common covariance functions (Rasmussen, 2004), we use the standard Exponential (or Ornstein–Uhlenbeck) function such that the  $(t, t')$ th element of  $f(\kappa)$  is

$$f_{tt'}(\kappa) = \exp\left(-\frac{|t - t'|}{\kappa}\right),$$

where  $|t - t'|$  is the one-dimensional Euclidean distance between the two timepoints  $t$  and  $t'$ . Alternatively, we can replace the distance term by  $|t - t'|^2$  and use the squared Exponential covariance function when smooth functions are required. Existing works (Bhattacharya and Dunson, 2011; Durante and Dunson, 2013) fix the parameter  $\kappa$  that characterizes the length-scale of the process, however, prior knowledge on how much the networks are correlated over time is unavailable in practice. To avoid the challenge of choosing an appropriate value of  $\kappa$ , we jointly estimate  $\tau$  and  $\kappa$  assuming inverse-Gamma and half-Cauchy priors— $\tau \sim \mathcal{IG}(a, b)$  and  $\kappa \sim \text{half-Cauchy}(\gamma)$ —across the parameters  $(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d})$ .

For the remaining parameters  $\mathbf{u}$  and  $\epsilon$ , we assign simple independent Normal and inverse-Gamma priors:

4. For  $i = 1, \dots, N; r = 1, \dots, R$ , and  $t = 1, \dots, T$ ,

$$u_{ir}^t \sim \mathcal{N}(0, \tau_{rt}^u),$$

where  $\tau_{rt}^u \sim \mathcal{IG}(a, b)$ .

5. For  $i = 1, \dots, N; j = 1, \dots, N$ , and  $t = 1, \dots, T$ ,

$$\epsilon_{ij}^t \sim \mathcal{N}(0, \sigma_e^2),$$

where  $\sigma_e^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$ , and the  $\epsilon_{ij}^t$  are independent given  $\sigma_e^2$ .

## 2.2 Posterior Computation

We take a Bayesian approach to infer the parameters in the DAME model. Our posterior computation is performed via a Gibbs sampler to update the vector of time-varying regression coefficients and the vector of additive and multiplicative latent factors, along with the use of a Metropolis-Hastings (MH) algorithm to sample the variance and length GP parameters  $(\tau, \kappa)$ . This section outlines the steps and sampling equations for MCMC updates of the DAME model, where the derivations of each step can be found in the supplementary material.

To begin with, let  $\mathbf{E}^t$  denote the  $N \times N$  matrix of random noise, where the  $(i, j)^{th}$  entry is defined as  $E_{ij}^t = y_{ij}^t - \left( \sum_{p=1}^P \beta_p^t X_{ijp}^t + \theta_i^t + \theta_j^t + u_i^{t'} \mathbf{D}^t u_j^t \right)$ . Given that the distribution of the observed

network  $\mathbf{Y} = \{\mathbf{Y}^1, \dots, \mathbf{Y}^T\}$  conditional on all the parameters can be written as the product of Normal probability density functions (pdf)

$$P(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{u}, \sigma_e^2, \boldsymbol{\tau}^\beta, \boldsymbol{\tau}^\theta, \boldsymbol{\tau}^d, \boldsymbol{\tau}^u, \boldsymbol{\kappa}^\beta, \boldsymbol{\kappa}^\theta, \boldsymbol{\kappa}^d) \propto \prod_{t=1}^T \prod_{i>j} (\sigma_e^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} \|E_{ij}^t\|^2\right\}, \quad (5)$$

we sequentially update each parameter from its full conditional distribution in the following sampling steps:

1. Sample  $\sigma_e^2 \sim \text{IG}\left(\frac{T \cdot N(N-1)}{4} + a_\sigma, \frac{1}{2} \sum_{t=1}^T \sum_{i>j} (E_{ij}^t)^2 + b_\sigma\right)$ ;
2. For each  $p = 1, \dots, P$  in a random order, sample  $\boldsymbol{\beta}_p$  as follows:
  - (a) Sample  $(\tau_p^\beta, \kappa_p^\beta)$  using a MH algorithm (refer to Equation (3) in the supplementary material)
  - (b) Sample  $\boldsymbol{\beta}_p \sim \mathcal{N}_T(\tilde{\mu}_{\beta_p}, \tilde{\Sigma}_{\beta_p})$  with

$$\tilde{\Sigma}_{\beta_p} = \left( (\tau_p^\beta c_p^\beta)^{-1} + \frac{\text{diag}(\{\sum_{i>j} X_{ijp}^{t2}\}_{t=1}^T)}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{\beta_p} = \left( \frac{\{\sum_{i>j} (E_{ij[-p]}^t X_{ijp}^t)\}_{t=1}^T}{\sigma_e^2} \right) \tilde{\Sigma}_{\beta_p},$$

where  $E_{ij[-p]}^t = E_{ij}^t + \beta_p^t X_{ijp}^t$ .

3. For each  $i = 1, \dots, N$  in a random order, sample  $\boldsymbol{\theta}_i$  as follows:
  - (a) Sample  $(\tau^\theta, \kappa^\theta)$  using a MH algorithm (refer to Equation (4) in the supplementary material)
  - (b) Sample  $\boldsymbol{\theta}_i \sim \mathcal{N}_T(\tilde{\mu}_{\theta_i}, \tilde{\Sigma}_{\theta_i})$  with

$$\tilde{\Sigma}_{\theta_i} = \left( (\tau^\theta c^\theta)^{-1} + \frac{(N-1)I_T}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{\theta_i} = \left( \frac{\{\sum_{i=i, j \neq i} E_{ij[-i]}^t\}_{t=1}^T}{\sigma_e^2} \right) \tilde{\Sigma}_{\theta_i},$$

where  $E_{ij[-i]}^t = E_{ij}^t + \theta_i^t$ .

4. For each  $r = 1, \dots, R$  in a random order, sample  $\mathbf{d}_r$  as follows:
  - (a) Sample  $(\tau_r^d, \kappa_r^d)$  using a MH algorithm (refer to Equation (5) in the supplementary material)
  - (b) Sample  $\mathbf{d}_r \sim \mathcal{N}_T(\tilde{\mu}_{d_r}, \tilde{\Sigma}_{d_r})$  with

$$\tilde{\Sigma}_{d_r} = \left( (\tau_r^d c_r^d)^{-1} + \frac{\text{diag}(\{\sum_{i>j} (u_{ir}^t u_{jr}^t)^2\}_{t=1}^T)}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{d_r} = \left( \frac{\{\sum_{i>j} (E_{ij[-r]}^t u_{ir}^t u_{jr}^t)\}_{t=1}^T}{\sigma_e^2} \right) \tilde{\Sigma}_{d_r},$$

where  $E_{ij[-r]}^t = E_{ij}^t + u_{ir}^{t'} d_r^t u_{jr}^t$ .

5. For each  $t = 1, \dots, T$  and  $i = 1, \dots, N$  in a random order, sample  $\mathbf{u}_i^t$  as follows:
  - (a) For each  $r = 1, \dots, R$ , sample  $\tau_{rt}^u \sim \mathcal{IG}(\frac{N}{2} + a_u, \frac{1}{2} \sum_{i=1}^N (u_{ir}^t)^2 + b_u)$  and construct the  $R \times R$  covariance matrix  $\boldsymbol{\tau}_t^u = \text{diag}(\tau_{1t}^u, \dots, \tau_{Rt}^u)$
  - (b) Sample  $\mathbf{u}_i^t \sim \mathcal{N}_R(\tilde{\mu}_{u_i^t}, \tilde{\Sigma}_{u_i^t})$  with

$$\tilde{\Sigma}_{u_i^t} = \left( (\boldsymbol{\tau}_t^u)^{-1} + \frac{\sum_{j \neq i} \mathbf{D}^t u_j^t u_j^{t'} \mathbf{D}^t}{\sigma_e^2} \right)^{-1} \text{ and } \tilde{\mu}_{u_i^t} = \left( \frac{\sum_{i=i, j \neq i} (E_{ij[-u]}^t u_j^{t'} \mathbf{D}^t)'}{\sigma_e^2} \right) \tilde{\Sigma}_{u_i^t},$$

where  $E_{ij[-u]}^t = E_{ij}^t + u_i^{t'} \mathbf{D}^t u_j^t$ .

Note that after steps 2 through 5,  $\mathbf{E} = \{\mathbf{E}^1, \dots, \mathbf{E}^T\}$  has to be calculated again using the previously updated values, so that any update is conditioned on the current values of all the other parameters.

### 2.3 Handling Missing Data

By the nature of longitudinal networks, new nodes can join the network and existing nodes can disappear at any timepoint. Consequently, any missing data (or edge) could be either missing at random or missing not at random, where the former straightforwardly suggests that the propensity for a data point to be missing is completely random (i.e., no relationship between a missing data point and any values in the data set) and the latter implies that the missingness is specifically related to what is missing. Specifically, missing not at random often occurs in longitudinal networks when a node has not yet joined or has dropped out. If we treat the two cases identically and ignore or impute the entire missings, it would be problematic since we may end up with a small number of nodes losing a large amount of information or introduce large bias into the estimation from imprecise imputation, respectively. While some continuous-time network models (Butts, 2008; Vu et al., 2011) naturally address this issue since they exploit survival analysis, allowing for a varying number of nodes is not a trivial issue in the modeling of discrete-time networks.

In the DAME model, we handle the two types of missing data—“random missing” and “structural missing”—using the approach similar to Snijders et al. (2010), which uses the known information on ‘joiners’ and ‘leavers’—i.e., identification on who are absent at a given timepoint. More precisely, we define the  $N \times T$  matrix of availability  $\mathbf{A}$  as an input to the model, where the  $(n, t)$ th element is defined as

$$A_{nt} = \begin{cases} 1, & \text{node } n \text{ is available at timepoint } t \\ 0, & \text{node } n \text{ is not available at timepoint } t, \end{cases}$$

and  $N$  is the number of actors who are part of the network at any time  $1 \leq t \leq T$ . We then assume that missing edges corresponding to nodes  $n$  and times  $t$  for which  $A_{nt} = 1$  are missing at random, while those for which  $A_{nt} = 0$  are structural missings. Following common practice, missing at random values are imputed from the current estimates of parameter values at each MCMC iteration. On the contrary, we leave out structural missings from the entire estimation procedure by estimating the parameters without including the structural missing values. Although this method is only applicable when we have prior knowledge about the availability of nodes at each timepoint, it is still a novel and natural solution to handle all missing edges and allow a varying number of nodes under the Bayesian setting.

## 3 Simulation Study

We provide a simulation study to evaluate the performance of our proposed model on its ability to capture some important properties of the true data and correctly reconstruct the true underlying processes from the model estimates. There are two objectives in this simulation study: 1) show that understanding the correct covariance structure plays a key role in the model performance, in the case of modeling a network that is highly correlated across time, and 2) demonstrate that the eigendecomposition formulation of the multiplicative random effects (i.e.  $\mathbf{u}'\mathbf{D}\mathbf{u}$ ) has the benefit of revealing various types of transitivity effects.

### 3.1 Estimating Strong Correlations

We generate a set of relational data  $\mathbf{Y}$  for  $N = 20$  and  $T = 10$  according to the generative process in Section 2.1, with  $P = 1$ ,  $R = 2$ ,  $(a, b) = (2, 1)$ , and  $(\kappa^\beta, \kappa^\theta, \kappa^d) = (10, 10, 10)$  so that the resulting dynamic network is highly correlated across timepoints with higher-order serial correlations. For example, the lag 1 correlation of the parameters is  $f_{01}(\kappa = 10) = 0.905$ , and the lag 9 correlation of the parameters is  $f_{09}(\kappa = 10) = 0.407$ . We run 6,000 MCMC iterations which appears to be long enough for full convergence, and then discard the first 1,000 samples with a thinning interval of 10.

To summarize our simulation results, we define a new measure that captures the overall correlation of the dataset and refer to “lagged degree correlations” (DC). For lag  $l = 1, \dots, T - 1$ , we

calculate the Pearson correlation  $\rho(\cdot)$  between the vectors  $v_1$  and  $v_2$ :

$$\begin{aligned} \text{DC}_l &= \rho(v_1, v_2), \text{ with} \\ v_1 &= [\text{degrees}(\hat{\mathbf{Y}}^1), \dots, \text{degrees}(\hat{\mathbf{Y}}^{T-l})], \\ v_2 &= [\text{degrees}(\hat{\mathbf{Y}}^{1+l}), \dots, \text{degrees}(\hat{\mathbf{Y}}^T)], \end{aligned} \quad (6)$$

where  $\text{degrees}(\hat{\mathbf{Y}}^t)$  indicates the vector of degree statistics for all nodes calculated from the posterior estimates of  $\mathbf{Y}^t$ , so that  $v_1$  and  $v_2$  both have length  $N \times (T-l)$ . We use this lagged degree correlation statistic to estimate the  $l$ -lag temporal dependence in discrete-time networks.

Figure 1 compares the posterior distribution of lagged degree correlations (DC) from our model and the “independence model” —the same general framework but without estimating the Gaussian process covariance parameters and instead fixing all  $\kappa = 0$ . The posterior samples of lagged degree correlations are calculated from the degree statistics constructed from their respective posterior estimates of  $\hat{\mathbf{Y}}$ . This comparison highlights the excellent performance of the DAME model in correctly estimating the true temporal correlation across the timepoints. This can be seen by comparing with the true degree correlation statistics, where the posterior DC estimates from the DAME model perfectly recovered all the degree correlations from lag 1 to lag 3. Meanwhile, the independence model always exhibits lower posterior estimates than the true correlations, showing that we may not be able to capture an important aspect of the true network—temporal dependence in long memory history—if we apply network models with temporal independence assumptions (e.g., static network models at each timepoint) or Markovian assumptions (e.g., the AR(1) network models in Section 1).

### 3.2 Capturing Transitivity

As introduced in Section 1, there exist two different types of transivities, positive and negative transitivity, and the parameterization without the  $\mathbf{D}$  term (i.e.  $\mathbf{u}'\mathbf{u}$ ) is not able to capture negative transitivity. We test whether the DAME model can explain both positive and negative transivities by conducting another simulation study. Similar to Section 3.1, we generate  $\mathbf{Y}$  with  $N = 20$ ,  $T = 10$ ,  $P = 1$ ,  $R = 2$ , and  $(a, b) = (2, 1)$ . Considering that our new goal is not to estimate temporal correlations but to represent transitivity effects, this time we fix  $d_r^t = \pm 2$  for  $r = 1, \dots, R$  and  $t = 1, \dots, T$  so that the generated network exhibits positive ( $d_r^t = +2$  for all  $r$  and  $t$ ), mixed ( $d_1^t = -2$  and  $d_2^t = +2$  for all  $t$ ), or negative ( $d_r^t = -2$  for all  $r$  and  $t$ ) transitive features, respectively. Again, we run 6,000 MCMC iterations and discard the first 1,000 with a thinning interval of 10. We fix the range parameter  $\kappa$ 's at their true values (i.e.,  $(\kappa^\beta, \kappa^\theta, \kappa^d) = (0, 0, 0)$ ) and do not estimate the covariance parameters (i.e., we fit the independence models in which all parameters and the resulting dynamic networks are independent across any timepoint), thus the difference in model performance only originates from the multiplicative effects formulations— $\mathbf{u}'\mathbf{D}\mathbf{u}$  and  $\mathbf{u}'\mathbf{u}$ . If we estimated  $\kappa$ 's for

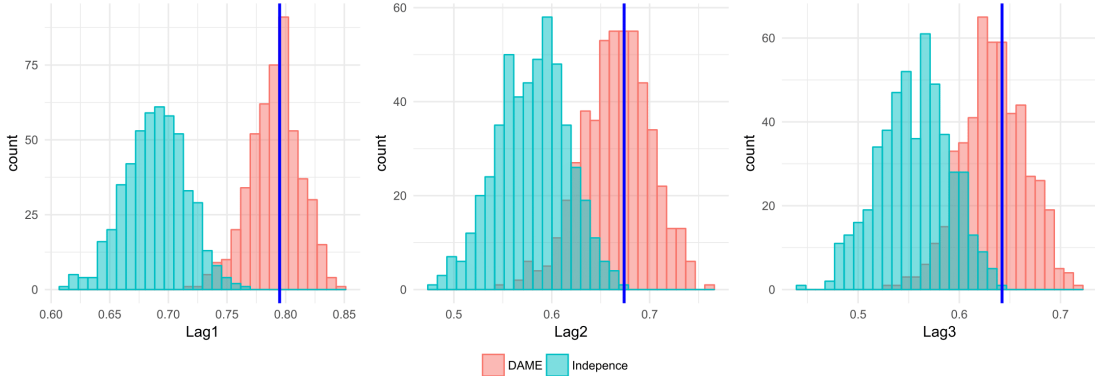


Figure 1: Histogram of posterior lagged degree correlations for  $l = 1, 2, 3$ : the DAME model (red) and independence model (green), with the vertical lines representing observed DC statistics.

this comparison, the difference in results may not only arise from the formulation of multiplicative effects, but also possibly from lack of correlation structure in  $\mathbf{u'u}$  because our modeling framework does not impose any temporal correlation on the latent positions  $\mathbf{u}$  in Section 2.1.

Figure 2 illustrates a graphical comparison between the two formulations of the multiplicative random effects with respect to the degree of the first, second, and third moments of the edge matrix (i.e.,  $\text{degree}(\hat{\mathbf{Y}})$ ,  $\text{degree}(\hat{\mathbf{Y}}^2)$  and  $\text{degree}(\hat{\mathbf{Y}}^3)$ , respectively). We randomly choose a node and show its degree distribution over time. For the case of positive transitivity, our model and its alternative do not show significant differences; both formulations achieve great performance in replicating the degrees of the first, second, and third moments of the edge matrix. On the contrary, when we fit the network with mixed or negative transitivity, the two formulations reveal noticeable differences. While the DAME model can still recover the true degrees of the first to third moments, the alternative model without  $\mathbf{D}$  term shows inaccuracy in simulating a network that is close to the true data. Not only does the alternative  $\mathbf{u'u}$  model introduce bias, but also it yields significantly wider interval estimates, implying lower precision compared to the DAME model. In addition, the evidence of the  $\mathbf{u'u}$  model’s failure to capture the transitivity effects becomes larger as the network tends toward stronger negative transitivity and also as we move to the degree statistics in higher moments. These findings strongly support our choice of the  $\mathbf{u'Du}$  formulation over the  $\mathbf{u'u}$  to model networks with various types of transitivity.

## 4 Analysis of the United Nations Voting Network

### 4.1 Data

Votes in the United Nations General Assembly (UNGA) have been analyzed in many political science papers (Voeten, 2000, 2004; Bearce and Bondanella, 2007; Mattes et al., 2015; Bailey et al., 2017) and have become the standard data sources to study states’ preferences, one of the most important topics in the field of international relations (Wendt, 1994). With regard to policy implications, for instance, states are the key actors on the global stage. Knowing their preferences towards each other and their stances on difference issues helps us predict future foreign policies and state behaviors. Unfortunately, many existing studies ignore three important features of the dataset. First, votes are highly correlated across timepoints, because they are the reflections of history. Bailey et al. (2017) propose a dynamic ordinal spatial IRT (item response theory) model that allows for inter-temporal comparisons, but their model limits the temporal dependence to be lag 1 (i.e., the Markovian assumption—votes at time  $(t + 1)$  are only dependent on votes at time  $t$ ). Second, although the researchers have viewed “voting” as dyadic behavior and have thus used dyadic similarity indicators such as affinity or S scores (Gartzke, 1998; Signorino and Ritter, 1999), to our knowledge, the United Nations voting data have never been analyzed using network models. Third, third-order dependence (e.g., transitivity and clusterability) has not been investigated despite the fact that voting decisions are not limited to dyadic calculations—country A’s decision to vote along with country B might well be influenced by a country C’s decision.

To process data from the United Nations General Assembly votes from 1983 to 2014, we first determine the countries to be included in this analysis by considering countries’ values in predictors such as polity score, or GDP, and then dropping countries with missing values in over 10 years, while the remaining missing values are imputed using the data from previous years. This results in 97 countries in total, and the full list of countries with their abbreviations is provided in Appendix A. The voting data were obtained from Voeten et al. (2016), specifically the subset of the votes called ‘important votes’, identified by the State Department as “votes on issues which directly affected important United States interests and on which the United States lobbied extensively.” For example, in 2001, important votes include ‘Israeli Actions in the Occupied Territories’, ‘Peaceful Settlement of the Question of Palestine’, ‘U.S. Embargo Against Cuba’, and ‘Nuclear Disarmament’. More can be found in <https://www.state.gov/p/io/rls/rpt/>. The number of important votes on average is 12 per year, ranging from 6 to 28. We only use important votes from the original data because non-important votes show high agreement rate over the time period (1983–2014) with few variations. Annual average voting similarity indices (i.e., agreement rates) for non-important votes are provided



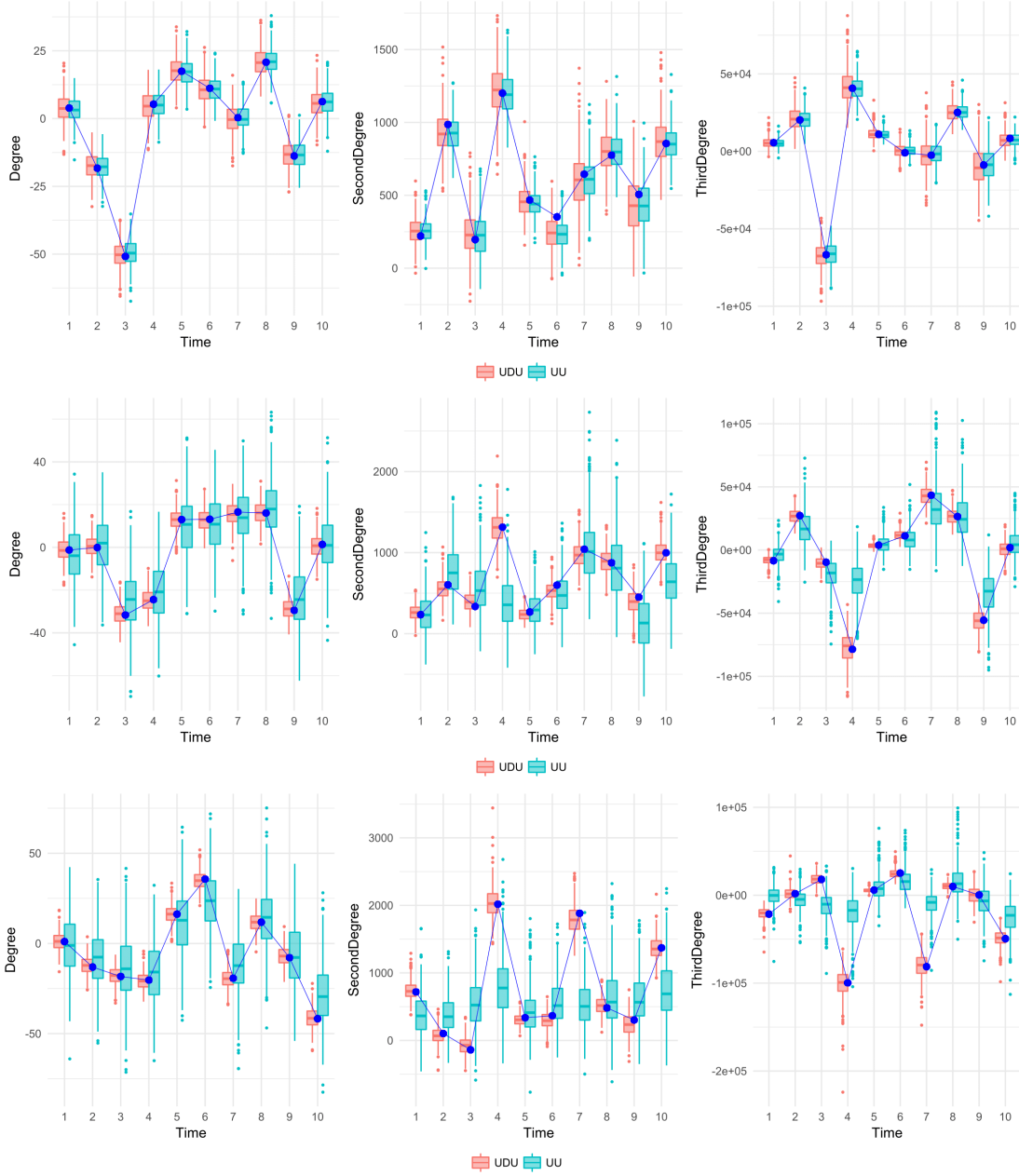


Figure 2: Boxplots of 500 posterior predictive degree statistics in the first (*left*), second (*center*), and third (*right*) moments corresponding to positive (*upper*), mixed (*middle*), and negative (*lower*) transivities: the DAME (red) and the  $\mathbf{u}'\mathbf{u}$  (green) models are shown with the dots representing the observed statistics.

in Appendix B. We then construct the response  $\mathbf{Y} = \{\mathbf{Y}^1, \dots, \mathbf{Y}^{32}\}$ , where each  $\mathbf{Y}^t$  is a  $97 \times 97$  matrix of a voting similarity index from 0 to 1 computed using three-category vote data ( $Y = \text{“yes”}$  or approval for an issue;  $A = \text{abstain}$ ;  $N = \text{“no”}$  or disapproval for an issue). Specifically, voting similarity index between the two countries  $i$  and  $j$  at year  $t$  is calculated as (Number of votes  $i$  and  $j$  agreed at year  $t$ ) / (Number of votes  $i$  and  $j$  both participated at year  $t$ ), which corresponds to the variable ‘agree3unimportant’ in the original dataset. Note that abstention is counted as half-agreement with a yes or no vote (Voeten et al., 2016), while two abstentions is treated as full agreement. For a basic summary of the United Nations voting data for important votes, see Appendix C.

As an exploratory data analysis, Table 1 illustrates the lagged degree correlation (defined in Equation 6) of the observed dataset to measure how strongly the United Nations voting data are correlated over time. There exists strong positive correlation in how the countries vote in the United Nations General Assembly over time, and as the distance between two timepoints becomes larger the correlation tends to be weaker. This provides solid evidence to support the use of a non-Markovian model, since the observed lagged degree correlations are higher than what is expected under the Markovian assumption (e.g., the expected lag  $l$  autocorrelations for AR(1) model are  $(0.732)^2 = 0.536$  for  $l = 2$ ,  $(0.732)^3 = 0.392$  for  $l = 3$ , and so on). Therefore, the DAME model with Gaussian process specifications may be one of the appropriate approaches to account for the strong temporal dependence in this dataset.

Lag	1	2	3	4	5	6	7	8	9	10
DC	0.732	0.623	0.513	0.435	0.395	0.315	0.263	0.158	0.164	0.203

Table 1: Lagged degree correlation (DC) of the United Nations voting data for  $l = 1, \dots, 10$ .

Next, to dynamically model the United Nations voting network in relation to other variables reflecting international relations, we combine  $P = 5$  different dyadic variables from the Correlates of War (COW) data (Gibler, 2008), Polity IV data (Marshall et al., 2014) and the International Monetary Fund (IMF)’s Direction of Trade Statistics (DOTS) and International Financial Statistics (IFS) data, and construct the observed edge covariates  $\mathbf{X}$ . For  $p = 1, \dots, P$ , we set the explanatory variable  $X_{ijp}^t$  as below.

1.  $X_{ij1}^t$ : intercept, included to account for the baseline degree of agreement.
2.  $X_{ij2}^t$ : log of the geographic distance between the capital cities of country  $i$  and country  $j$ .
3.  $X_{ij3}^t$ : 1 if country  $i$  and country  $j$  have a formal alliance including mutual defense pacts, non-aggression treaties, and ententes at time  $t$ , and 0 otherwise.
4.  $X_{ij4}^t$ : absolute difference in polity score between country  $i$  and country  $j$  at time  $t$ <sup>1</sup>.
5.  $X_{ij5}^t$ : index of economic dependence using bilateral trade weighted by each country’s gross domestic product (GDP), as defined in Gartzke (2000). That is,

$$X_{ij5}^t = \min\left(\frac{\text{Trade}_{ijt}}{\text{GDP}_{it}}, \frac{\text{Trade}_{jti}}{\text{GDP}_{jt}}\right).$$

6.  $X_{ij6}^t$ : indicator of whether country  $i$  and country  $j$  share the official language.

By definition, all covariates are symmetric (i.e.,  $X_{ijp}^t = X_{jip}^t$ ). Two variables—log(distance) and common language—are time-invariant covariates, although their coefficients may vary over time. Correlations between the covariates are summarized in Appendix D.

We specify the matrix of availability  $\mathbf{A}$  introduced in Section 2.3 to reflect some countries’ non-participation in the United Nations General Assembly (UNGA):

1. North Korea (PRK) has structural missing values from  $t = 1$  to  $t = 8$  because North Korea did not vote until North Korea and South Korea were simultaneously admitted to the United Nations in 1991.
2. South Korea (ROK) has structural missing values from  $t = 1$  to  $t = 8$  because South Korea did not vote until North Korea and South Korea were simultaneously admitted to the United Nations in 1991.

---

<sup>1</sup>Polity IV data contain coded annual information on the level of democracy for various countries, and a polity score ranges from -10 to +10, with -10 to -6 corresponding to autocracies, -5 to 5 corresponding to anocracies, and 6 to 10 to democracies.

3. Russia (RUS) has structural missing values from  $t = 1$  to  $t = 9$  because Russia succeeded the Soviet Union's seat, including its permanent membership on the Security Council in the United Nations, after the dissolution of the Soviet Union in 1991.
4. Iraq (IRQ) has structural missing values from  $t = 13$  to  $t = 21$  because Iraq did not participate in the UNGA roll-call votes from 1995 to 2003. Under the rule of Saddam, Iraq had been under severe sanctions from the international community, including the United Nations, since 1990.

Any missing values corresponding to a country's missing period are treated as structural missing values. As explained in Section 2.3, other missing values are treated as missing at random and are thus imputed.

## 4.2 Model Validation

To check the fit of the DAME model of Section 2.1 to the data, we fit the model with four different specifications: 1) with additive and multiplicative effects (DAME), 2) one with only multiplicative effects (ME), 3) with only additive effects (AE), and 4) without any random effects (NO). Each of the four specifications uses all six edge covariates in Section 4.1. Figure 3 depicts the degree statistics constructed from 500 different posterior predictive samples (i.e.,  $\text{degree}(\mathbf{Y})$ ). Out of 97 countries, we only present the results for Israel (ISR), which reveal clear differences among the four models. First of all, we see the bias correcting effect of including additive effects (AE), compared to the model with no random effects (NO). Next, when we compare models with only additive effects (AE) and only multiplicative effects (ME), the multiplicative effects model shows significantly narrower width of credible intervals. Lastly, our model with both additive and multiplicative effects (DAME) outperforms the ME model in terms of both accuracy and precision. To be specific, the DAME model corrects the bias in the ME model by incorporating node-specific additive effects. Overall, not only does the DAME model provide the most accurate estimates over time, but it also yields the narrowest 95% credible intervals among the four. These findings emphasize the importance of including both the additive and multiplicative terms to enable the model to capture some features not explicable by fixed effects or only one random effect. More results and interpretation using the full DAME model are presented in Section 4.3, and the posterior predictive plots checking the overall degree distributions, aggregating all nodes and timepoints, are provided in Appendix E.

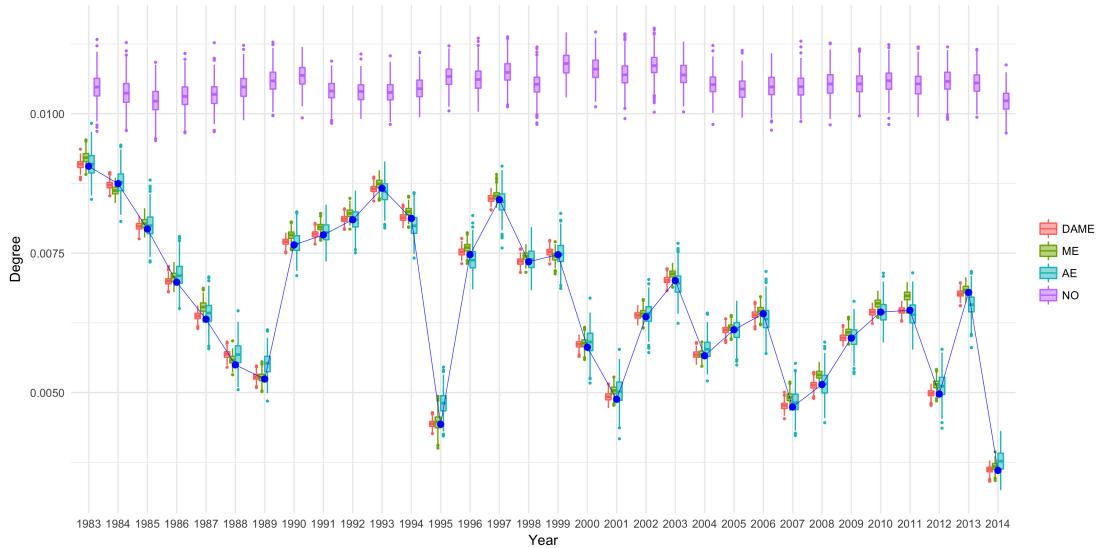


Figure 3: Boxplots of 500 posterior predictive degree statistics for Israel (ISR): the DAME (*red*), ME (*green*), AE (*blue*), and NO (*purple*) models are shown with the dots representing the country's observed statistics.

### 4.3 Parameter Estimation and Interpretation

To apply the DAME model to the United Nations General Assembly voting data, we fix the dimension of the multiplicative effects to be  $R = 2$  based on some preliminary experiments, where increasing the dimension does not significantly improve the model fitting. For instance, the estimated eigenvalue  $d_3^t \approx 0$  for every  $t = 1, \dots, T$ . Note that a unique reduced-rank structure of voting (or agreement) network is briefly explained in Appendix F. In this section, we present the results based on 30,000 Gibbs iterations with a burn-in of 5,000, where we thin by keeping every 50<sup>th</sup> sample. All model parameters, including the GP parameters  $(\kappa, \tau)$ , are estimated according to Section 2.2, using the hyperparameters  $(a, b) = (2, 1)$  and  $\gamma = 5$ .

Figure 4 shows the posterior mean estimates of the fixed effect coefficients  $\{\beta_p\}_{p=1}^P$  with their corresponding 95% credible intervals. Overall, the effects of the covariates on the United Nations voting behavior change substantially over time, especially in the cases of geographic distance and trade-to-GDP ratio. Most importantly, the “critical junction” for these temporal changes seems to be around the end of Cold War, that is, the late 1980s and early 1990s. For instance, the middle panel of the top row reveals the pattern of influence of geographic distance on voting behavior. The gravity model (Leibenstein, 1966; Rodrigue et al., 2009) suggests that the influence of phenomena or populations (e.g., trade and migration) on two countries varies inversely with the distance between them, and we see an overall negative coefficient for geographic distance which is consistent with the gravity model. However, the negative effect of geographic distance is less significant after the early 1990s. It is likely that the votes in the United Nations were much more influenced by the overall ideological conflicts between the Soviet Union plus its satellites and the Western camp so the effect of geographical distance was weakened during the Cold War. Moreover, regarding the effect of polity—or distance in polity as we operationalize this variable—the result suggests that in general the political regime similarity does not often result in higher agreement in the United Nations General Assembly, at least for a few time periods included in the study, e.g., 1990–1995, 1997–2002, and 2005–2010. Scholars in the liberal tradition of international relations have long been arguing for shared norms, values and preferences between democracies; what the result here suggests that such similarity in preferences seem to be not sufficiently strong enough to sway countries’ votes in the United Nations General As-

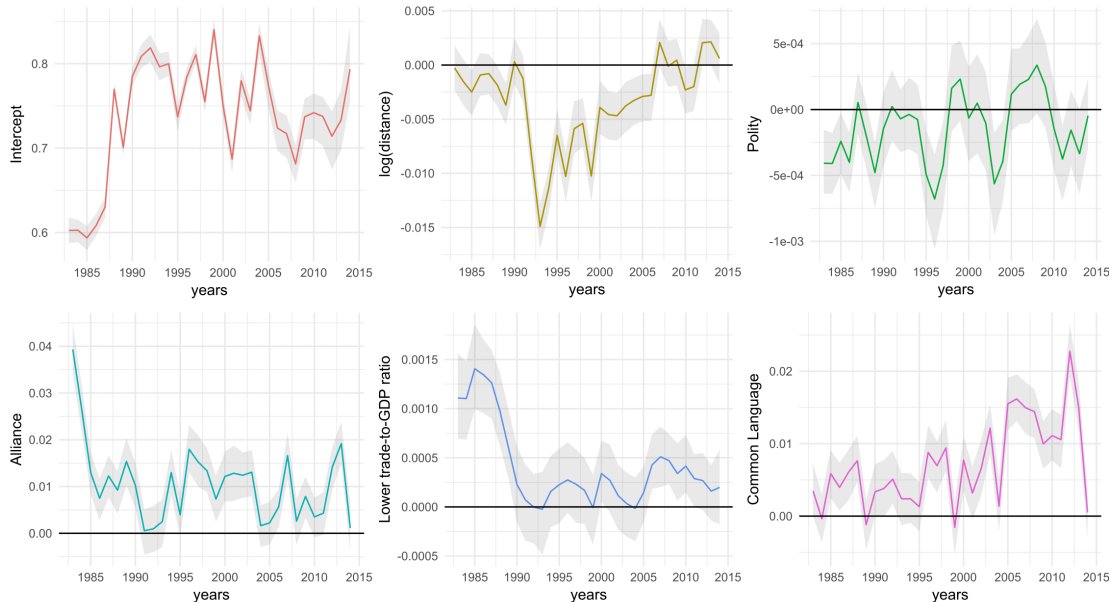


Figure 4: Posterior mean estimates for the fixed effect coefficients  $\beta$  (colored line): Intercept,  $\log(\text{distance})$ , Alliance, Polity difference, Lower trade-to-GDP ratio, and Common Language, and their corresponding 95% credible intervals (grey areas).

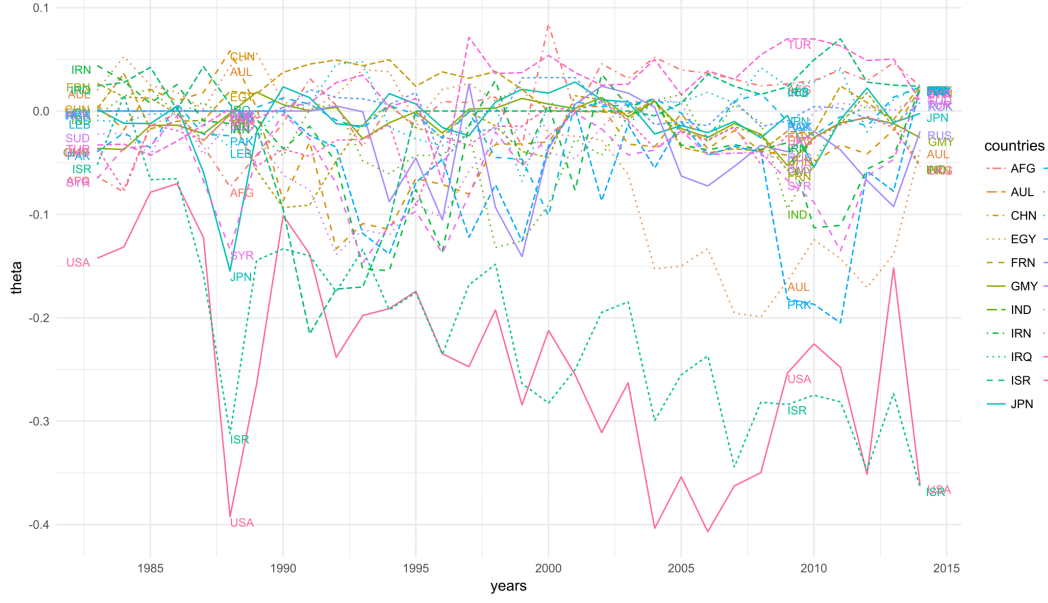


Figure 5: Posterior mean estimates for the additive random effect estimates  $\theta$ .

sembly, at least in the case of important votes and when we account for the other factors in the model.

After controlling for the observed covariates, we move on to the analysis of random effects, both additive and multiplicative ones. For clear visualization, we only present the result from 21 countries, where the countries are chosen based on the most active countries during the ten year period from 2004 to 2014 (Hoff, 2015). Here, the action types include negative material actions, positive material actions, negative verbal actions and positive verbal actions. The 21 most active countries are marked with \* in Appendix A. Figure 5 shows the posterior mean estimates of each country’s additive random effects, that is, its node-specific time-varying intercepts  $\theta_i^t$ . Here, the United States (USA) and Israel (ISR) stand out with large negative additive random effects, suggesting that these two countries are less likely overall to cast the same votes as the rest of the countries. Considering that the majority of votes are “yes”, the two countries are more likely to vote for “no” in general.

Finally, we provide the estimated latent positions of the 21 countries. To determine the posterior distribution of  $\mathbf{u}$  without identifiability issues, we calculate an eigendecomposition on every posterior sample of the multiplicative effect matrix  $\mathbf{u}'\mathbf{D}\mathbf{u}$ , and let  $\mathbf{D}$  be the diagonal matrix of eigenvalues and  $\mathbf{u}$  be the corresponding eigenvectors. We then apply a Procrustes transformation on each posterior estimate of  $\mathbf{u}$  and multiply by  $\sqrt{\mathbf{D}}$ , and obtain the posterior mean estimates and 95% credible regions of  $\mathbf{u}_i\sqrt{\mathbf{D}}$ , for all  $i = 1, \dots, N$ . In Figure 6, we see clear patterns of clustering of the countries. For example, in 1986, we observe a clear cluster including the USA and its Cold War allies—Japan (JPN), France (FRN), Germany (GMY), United Kingdom (UKG), Australia (AUL) and Israel (ISR). Moreover, it is interesting that Israel (ISR) is always close to the United States (ISR) in the latent space. Specifically, in 2014, USA and Israel seem to have drifted away from other countries including the USA’s traditional allies in Europe and Asia, which indicates that the two countries’ alliance is beyond those observed variables such as economic factors.

## 5 Discussion

As an extension of the additive and multiplicative effects (AME) model, the dynamic additive and multiplicative network effects (DAME) model can flexibly learn the underlying time-varying structure in dynamic networks, while inferring the effects of node-specific and dyad-specific latent variables. Accounting for the correlation structure of the networks makes better use of dynamic networks than

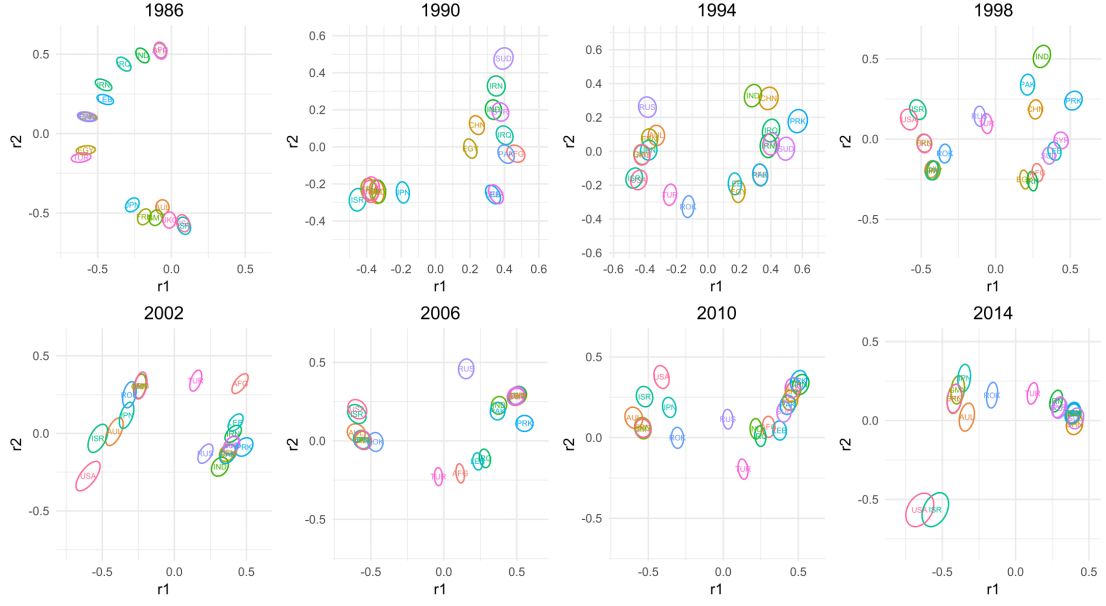


Figure 6: Posterior mean estimates for the multiplicative random effects and their corresponding 95% credible regions for the 21 selected countries for every 4 years between 1983 and 2014.

modeling them as separate network snapshots. Our algorithm eliminates the need to assume an arbitrary user-defined covariance structure, making it easier to learn the temporal dependence from the data itself. Information on temporal correlation thus leads to more accurate and precise inference. Further, the visualization of the model-estimated time-varying parameters provides an effective temporal trend analysis of dynamic networks, as well as the descriptive visualization of higher-order dependencies over time.

We have demonstrated effectiveness of our model by modeling the United Nations voting networks. The estimated additive and multiplicative effects and their changes over time reveal that the United Nations voting behavior reflects interesting and meaningful foreign policy positions and alliances of various countries, even after controlling for other edge covariates that are considered critical in the studies of international relations. Although we illustrate the entire framework in the context of symmetric or undirected networks, our model can be easily extended to allow directed networks, following the additive and multiplicative effects model for the directed network (Minhas et al., 2016a). Furthermore, the approach can be applied to binary and ordinal network data with appropriate link functions, while we currently only provide the application to continuous-valued networks. Finally, considering the recent explosion of network dataset with large numbers of timepoints, our model has a broad range of applicability, suggesting a promising approach that can accommodate huge networks that span long periods of time.

## References

- Bailey, M. A., Strezhnev, A., and Voeten, E. (2017). Estimating dynamic state preferences from United Nations voting data. *Journal of Conflict Resolution*, 61(2):430–456.
- Bearce, D. H. and Bondanella, S. (2007). Intergovernmental organizations, socialization, and member-state interest convergence. *International Organization*, 61(4):703–733.
- Bhattacharya, A. and Dunson, D. B. (2011). Sparse bayesian infinite factor models. *Biometrika*, 98(2):291–306.
- Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, 38(1):155–200.
- Durante, D. and Dunson, D. (2014a). Bayesian logistic gaussian process models for dynamic networks. In *Artificial Intelligence and Statistics*, pages 194–201.
- Durante, D. and Dunson, D. B. (2013). Nonparametric bayes dynamic modeling of relational data. *arXiv preprint arXiv:1311.4669*.
- Durante, D. and Dunson, D. B. (2014b). Bayesian dynamic financial networks with time-varying predictors. *Statistics & Probability Letters*, 93:19–26.
- Friel, N., Rastelli, R., Wyse, J., and Raftery, A. E. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences*, 113(24):6629–6634.
- Gartzke, E. (1998). Kant we all just get along? opportunity, willingness, and the origins of the democratic peace. *American Journal of Political Science*, pages 1–27.
- Gartzke, E. (2000). Preferences and the democratic peace. *International Studies Quarterly*, 44(2):191–212.
- Gibler, D. M. (2008). *International military alliances, 1648-2008*. CQ Press.
- Hanneke, S., Fu, W., Xing, E. P., et al. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.
- He, Y. and Hoff, P. D. (2017). Multiplicative coevolution regression models for longitudinal networks and nodal attributes. *arXiv preprint arXiv:1712.02497*.
- Hoff, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems*, pages 657–664.
- Hoff, P., Fosdick, B., Volfovsky, A., and Stovel, K. (2014). amen: Additive and multiplicative effects modeling of networks and relational data. *R package version 0.999*. URL: <http://CRAN.R-project.org/package=amen>.
- Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469):286–295.
- Hoff, P. D. (2009). Multiplicative latent factor models for description and prediction of social networks. *Computational and mathematical organization theory*, 15(4):261–272.
- Hoff, P. D. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, 55(1):530–543.
- Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *The Annals of Applied Statistics*, 9(3):1169.
- Hoff, P. D. et al. (2011). Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.

- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.
- Kim, B., Lee, K., Xue, L., and Niu, X. (2017). A review of dynamic network models with latent variables. *arXiv preprint arXiv:1711.10421*.
- Leibenstein, H. (1966). Shaping the world economy: suggestions for an international economic policy.
- Marshall, M. G., Jagers, K., and Gurr, T. R. (2014). Polity IV annual time-series, 1800–2013. *Center for International Development and Conflict Management at the University of Maryland College Park*.
- Mattes, M., Leeds, B. A., and Carroll, R. (2015). Leadership turnover and foreign policy change: Societal interests, domestic institutions, and voting in the United Nations. *International Studies Quarterly*, 59(2):280–290.
- Minhas, S., Hoff, P. D., and Ward, M. D. (2016a). Inferential approaches for network analyses: Amen for latent factor models. *arXiv preprint arXiv:1611.00460*.
- Minhas, S., Hoff, P. D., and Ward, M. D. (2016b). A new approach to analyzing coevolving longitudinal networks in international relations. *Journal of Peace Research*, 53(3):491–505.
- Rasmussen, C. E. (2004). Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer.
- Rodrigue, J.-P., Comtois, C., and Slack, B. (2009). *The geography of transport systems*. Routledge.
- Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter*, 7(2):31–40.
- Sarkar, P., Siddiqi, S. M., and Gordon, G. J. (2007). A latent space approach to dynamic embedding of co-occurrence data. In *AISTATS*, pages 420–427.
- Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.
- Sewell, D. K. and Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116.
- Signorino, C. S. and Ritter, J. M. (1999). Tau-b or not tau-b: measuring the similarity of foreign policy positions. *International Studies Quarterly*, 43(1):115–144.
- Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60.
- Voeten, E. (2000). Clashes in the assembly. *International Organization*, 54(2):185–215.
- Voeten, E. (2004). Resisting the lonely superpower: Responses of states in the United Nations to US dominance. *Journal of Politics*, 66(3):729–754.
- Voeten, E., Strezhnev, A., and Bailey, M. (2016). United Nations general assembly voting data.
- Vu, D. Q., Hunter, D., Smyth, P., and Asuncion, A. U. (2011). Continuous-time regression models for longitudinal networks. In *Advances in Neural Information Processing Systems*, pages 2492–2500.
- Ward, M. D., Ahlquist, J. S., Rozenas, A., et al. (2013). Gravity’s rainbow: A dynamic latent space model for the world trade network. *Network Science*, 1(01):95–118.
- Ward, M. D. and Hoff, P. D. (2007). Persistent patterns of international commerce. *Journal of Peace Research*, 44(2):157–175.
- Wendt, A. (1994). Collective identity formation and the international state. *American Political Science Review*, 88(2):384–396.
- Xu, K. S. and Hero III, A. O. (2013). Dynamic stochastic blockmodels: Statistical models for time-evolving networks. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pages 201–210. Springer.



# Appendices

## A List of Countries in Voting Network

Abbreviation	Country or area name	Abbreviation	Country or area name
AFG*	Afghanistan	KUW	Kuwait
ALB	Albania	LEB*	Lebanon
ALG	Algeria	LIB	Libya
ANG	Angola	MAA	Mauritania
ARG	Argentina	MEX	Mexico
AUL*	Australia	MLI	Mali
BAH	Bahrain	MOR	Morocco
BEN	Benin	MZM	Mozambique
BFO	Burkina Faso	NEW	New Zealand
BNG	Bangladesh	NIC	Nicaragua
BOL	Bolivia	NIG	Nigeria
BRA	Brazil	NIR	Niger
BUI	Burundi	NOR	Norway
BUL	Bulgaria	NTH	Netherlands
CAN	Canada	OMA	Oman
CAO	Cameroon	PAK*	Pakistan
CEN	Central African Republic	PAN	Panama
CHL	Chile	PAR	Paraguay
CHN*	China	PER	Peru
COL	Colombia	PHI	Philippines
CON	Congo	POL	Poland
COS	Costa Rica	POR	Portugal
DEN	Denmark	PRK*	North Korea
DOM	Dominican Republic	QAT	Qatar
ECU	Ecuador	ROK*	South Korea
EGY*	Egypt	RUS*	Russia
FIN	Finland	RWA	Rwanda
FRN*	France	SAL	El Salvador
GAB	Gabon	SAU	Saudi Arabia
GAM	Gambia	SEN	Senegal
GMY*	Germany	SIE	Sierra Leone
GHA	Ghana	SPN	Spain
GRC	Greece	SUD*	Sudan
GUA	Guatemala	SUR	Suriname
GUI	Guinea	SYR*	Syrian Arab Republic
GUY	Guyana	TAZ	Tanzania
HAI	Haiti	TOG	Togo
HON	Honduras	TRI	Trinidad and Tobago
HUN	Hungary	TUN	Tunisia
IND*	India	TUR*	Turkey
INS	Indonesia	UAE	United Arab Emirates
IRN*	Iran (Islamic Republic of)	UGA	Uganda
IRQ*	Iraq	UKG*	United Kingdom
ISR*	Israel	URU	Uruguay
ITA	Italy	USA*	United States of America
JAM	Jambia	VEN	Venezuela
JOR	Jordan	ZAM	Zambia
JPN*	Japan	ZIM	Zimbabwe
KEN	Kenya		

Table 2: Full list of the 97 countries, where the 21 most active countries during the ten year period of 2004 – 2014 (Hoff, 2015) are marked with \*. 17

## B Summary of the UN Voting Network—Non-important Votes

Year	1983	1984	1985	1986	1987	1988	1989	1990	1991
Joint votes	126.709	128.302	134.563	139.387	133.998	123.612	104.472	78.943	61.753
Agreement	0.847	0.852	0.843	0.854	0.877	0.872	0.885	0.878	0.864
Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Joint votes	58.263	52.001	56.187	62.019	61.477	58.104	50.969	55.071	52.731
Agreement	0.842	0.835	0.847	0.835	0.844	0.831	0.851	0.837	0.839
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Joint votes	50.362	58.282	63.418	61.860	59.651	73.235	64.857	62.353	57.893
Agreement	0.815	0.823	0.830	0.814	0.835	0.831	0.819	0.829	0.804
Year	2010	2011	2012	2013	2014				
Joint votes	57.336	54.404	60.216	56.016	66.197				
Agreement	0.822	0.798	0.823	0.802	0.814				

Table 3: Summary of the United Nations voting data for non-important votes: Average number of common votes (*upper*) and average voting similarity index (*lower*) per year.

## C Summary of the UN Voting Network—Important Votes

Year	1983	1984	1985	1986	1987	1988	1989	1990	1991
Joint votes	7.384	7.441	8.129	9.201	8.283	5.121	12.605	7.361	8.559
Agreement	0.696	0.724	0.725	0.748	0.725	0.773	0.798	0.832	0.836
Year	1992	1993	1994	1995	1996	1997	1998	1999	2000
Joint votes	13.458	11.013	13.447	24.932	9.655	10.026	8.427	10.776	8.970
Agreement	0.806	0.773	0.785	0.827	0.765	0.805	0.776	0.829	0.768
Year	2001	2002	2003	2004	2005	2006	2007	2008	2009
Joint votes	9.191	12.826	12.142	8.432	8.855	10.956	10.681	10.845	10.946
Agreement	0.729	0.816	0.755	0.835	0.783	0.733	0.739	0.700	0.750
Year	2010	2011	2012	2013	2014				
Joint votes	12.289	9.067	7.575	10.171	12.048				
Agreement	0.744	0.751	0.733	0.754	0.847				

Table 4: Summary of the United Nations voting data for important votes: Average number of common votes (*upper*) and average voting similarity index (*lower*) per year.

## D Correlation between Observed Covariates

correlation	log(distance)	polity	alliance	Trade/GDP	language
log(distance)	1.000	0.136	-0.508	-0.355	-0.286
polity	0.136	1.000	-0.264	-0.095	-0.142
alliance	-0.508	-0.264	1.000	0.275	0.417
Trade/GDP	-0.355	-0.095	0.275	1.000	0.100
language	-0.286	-0.142	0.417	0.100	1.000

Table 5: Pearson correlation coefficients between the observed dyadic covariates.

## E Posterior Predictive Checks on Degree Statistics

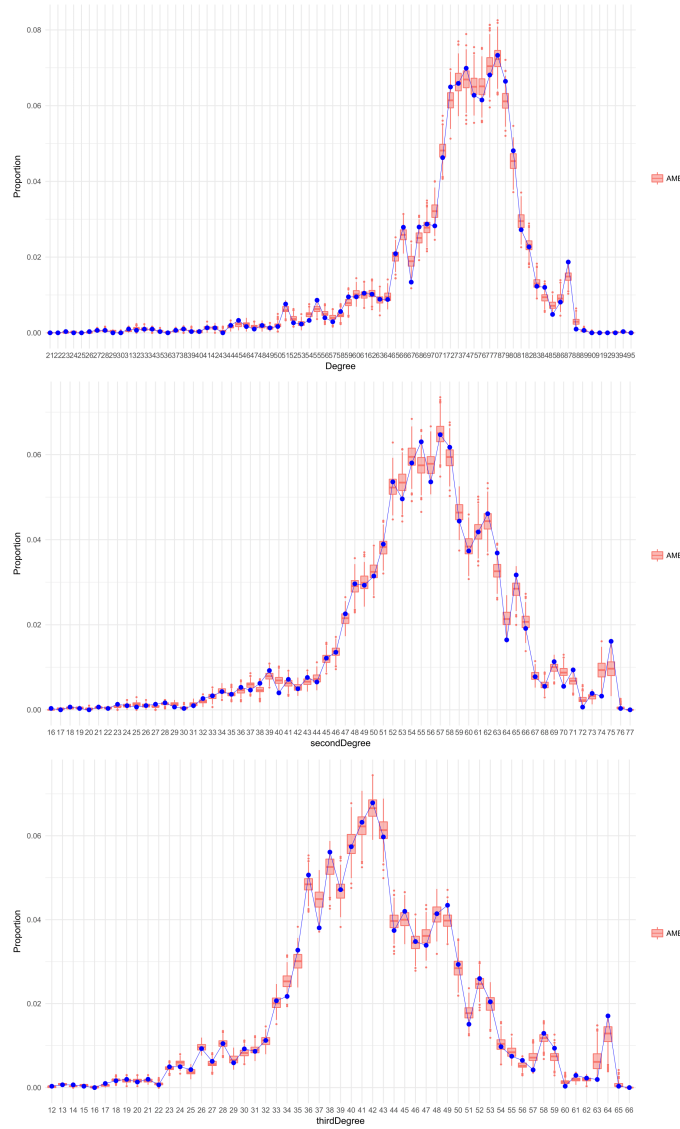


Figure 7: Posterior predictive plots of the overall degree distributions aggregating all nodes and timepoints: the first (*upper*), second (*middle*), and third moments (*lower*) shown with the dots representing observed statistics.

## F Reduced-Rank Structure of Voting Network

The particular construction of the voting network endows it with a particular low-rank structure, which is shared by other similarly constructed “agreement” networks. For simplicity, we assume a static network by fixing  $T = 1$  and consider the  $N \times N$  adjacency matrix  $\mathbf{V}^1$  representing the network defined by a single vote. Since each entry of  $\mathbf{V}^1$  summarizes a voting similarity index computed using 3 categories (“yes”, “abstain”, and “no”), the  $(i, j)^{th}$  element  $V_{ij}^1$  can only have three possible values, 1 for agreement, 0 for disagreement, or 0.5 for half-agreement (one abstention). This implies transitivity of the agreement network—i.e., if there is an agreement between  $i$  and  $j$ , and also between  $j$  and  $h$ , then there must be an agreement between  $i$  to  $h$ . Therefore, any two edges sharing one node, such as  $V_{ij}^1$  and  $V_{jh}^1$ , automatically determines the third edge between the unshared nodes  $V_{ih}^1$ . This constraint makes the maximum rank of  $\mathbf{V}^1$  to be at most 3. In other words, if we were to apply the DAME model to this type of single vote network (without any additive effects and explanatory variables), the maximum rank of dimension for the multiplicative effects we could fit for low rank factorization is  $R = 3$ . Moreover, each dimension  $r$  in the estimated  $\mathbf{u}$  and  $\mathbf{D}$  can be viewed as the distinct constructs behind the vote.

Despite the constraint satisfied by a single vote matrix, reduced rank does not appear to affect the modeling of the United Nations voting network, where we aggregate multiple votes per year (minimum number of important votes per year is 6). Each aggregated matrix  $\mathbf{V} = \mathbf{V}^1 + \mathbf{V}^2 + \dots + \mathbf{V}^M$  has large or full rank, so we ignore the rank of the multiplicative effects in fitting the DAME model. Furthermore, including the observed covariates  $\mathbf{X}$  and additive random effects  $\boldsymbol{\theta}$ ’s also mitigate the reduced rank structure, since the multiplicative effect  $\mathbf{u}'\mathbf{D}\mathbf{u}$  is modeled after we subtract those effects from the response.