

Dynamic Additive and Multiplicative Effects (DAME) Model with Application to the United Nations Voting Behaviors

Bomin Kim¹, Xiaoyue Niu¹, David Hunter¹, and Xun Cao²

¹Department of Statistics, The Pennsylvania State University

²Department of Political Science, The Pennsylvania State University

Abstract

In this paper, we introduce the dynamic version of a Gaussian additive and multiplicative effects (AME) model—a statistical regression model for symmetric discrete-time networks that are correlated over time. Our model extends the additive and multiplicative latent factor network model of Hoff (2009) and 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 United Nations 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 changing number of nodes over time. Overall, the DAME model shows significantly better fit to the dataset compared to the 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 or collaboration networks to disease transmission. Because a network naturally evolves over time, there has been 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 side, 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 important 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) to model symmetric discrete-time networks, with emphasis on the latent structures—unmeasured attributes of nodes for tie formation in networks and their evolution over time.

Since Hoff et al. (2002) developed a class of models where the probability of a relation between actors depends on the positions of individuals in an unobserved “social space,” there have been several versions of the latent projection model in which the probability of a tie between nodes i and j is determined by the two nodes’ latent vectors using $f(v_i, v_j) = (v_i' v_j) |v_j|^{-1}$, gradually providing additional explicability and flexibility. Hoff (2005) first introduced the symmetric multiplicative interaction effect $(v_i' v_j)$ 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, 2005). After modifying the parameterization of multiplicative effects into the form of an eigendecomposition $u_i^T \Lambda u_j$, Hoff (2008) claimed that this “latent eigenmodel” is able to represent

a wide array of patterns in the data, due to the fact that the eigenmodel provides an unrestricted low-rank approximation to the symmetric relational data. Hoff (2009) extended the framework to model asymmetric social networks using the singular value decomposition $u_i^T D v_j$. Finally, Hoff et al. (2014) and Minhas et al. (2016a) combined the additive and multiplicative effects to model the second-order (or reciprocity) and third-order dependencies and established the AME—additive and multiplicative effects—regression model for dyadic response data y_{ij} , in which

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

where $\alpha(\mathbf{u}_i, \mathbf{v}_j) = \mathbf{u}_i^T D \mathbf{v}_j$ and $g(\theta_{ij})$ is a link function, 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 latent space models in the last decade (Kim et al., 2017). For instance, latent distance models have been widely used for dynamic networks 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 introduced the concept of dynamic latent factors and this idea was expanded by Ward et al. (2013) to analyze bilateral trade using the generalized bi-linear mixed effect model (Hoff, 2005). This model allowed time-varying edge covariate parameters as well as latent factors, however, it did not involve any dynamic information in the dataset. Most recently, He and Hoff (2017) introduced a coevolution model for the analysis of longitudinal network and nodal attribute data, including latent nodal attributes. Their model—multiplicative coevolution regression (MCR) model—provides a large benefit of allowing the possibility that nodes may change their nodal (or latent) attributes \mathbf{X}_t depending on their past relation as well as the evolution of the network \mathbf{Y}_t , but their contagion of the network is limited to first-order autoregressive model, or AR(1). 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 network serving as an example of the general model. Still, the latent variable component is not emphasized in those papers.

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), with the model formulation

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

for $i < j$, with $x_{ih}(\cdot) \sim \text{GP}(0, \tau_h^{-1} c_X)$ and $\mu(\cdot) \sim \text{GP}(0, c_\mu)$, where $x_i(t) = [x_{i1}(t), \dots, x_{iH}(t)]'$ for $i = 1, \dots, V$ are the latent vectors of node i , together with the baseline $\mu(t)$. The variance multipliers τ_h^{-1} for $h = 1, \dots, H$ are shrinkage parameters with a gamma prior, and c_μ 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 strong advantage of learning the number of latent dimensions H in the model (Bhattacharya and Dunson, 2011). 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 $x_i(t)' x_j(t)$ to represent multiplicative latent factor effects. However, the form $u_i(t)^T D(t) u_j(t)$ used in the AME model allows the flexibility of negative eigenvalues $d_r(t)$. According to Hoff (2008), the parameterization of $u_i(t)^T D(t) u_j(t)$ can represent both positive or negative transitivity in varying degrees; on the other hand, the parameterization of $v_i(t)^T v_j(t)$ 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.

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). In addition, 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 as in Durante and Dunson (2013, 2014b), our method does not require any initial guess about correlations and further enables efficient estimation of the fixed and random effect parameters. Second, in order to increase flexibility and accuracy of the model, DAME allows the number of nodes to change over time by allowing a special case of missing values, which are referred to as “structural zeros” in this paper. In what follows, we introduce the DAME model and describe how we take advantage of temporal correlation, derive the sampling equations for hierarchical Bayesian inference, present simulation studies to show the advantage of DAME over alternative approaches, and apply DAME model to the United Nations voting network.

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 $\mathbf{Y} = \{Y^1, \dots, Y^T\}$, where the entry y_{ij}^t denotes any relational data corresponding to the node pair (i, j) at timepoint t , using the observed covariates $\mathbf{X} = \{X^1, \dots, X^T\}$ where $X^t = \{X_1^t, \dots, X_P^t\}$. 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, β_p^t is the corresponding unknown coefficient, and z_{ij}^t is the unobserved random effect. Following the latent factor model of Minhas et al. (2016a), we also model z_{ij}^t to account for potential higher-order dependencies in the node pair (i, j) with the additive and multiplicative form

$$z_{ij}^t = \theta_i^t + \theta_j^t + u_i^{t'} D^t u_j^t + \epsilon_{ij}^t, \quad (4)$$

where θ_i^t and θ_j^t are the node-specific additive random effects, $u_i^{t'} D^t u_j^t$ is the multiplicative random effect with D^t denoting the $R \times R$ diagonal matrix of eigenvalues (d_1^t, \dots, d_R^t) and $u_i^t = [u_{i1}^t, \dots, u_{iR}^t]'$ denoting the R -length vector of latent coordinates of node i , and ϵ_{ij}^t is the pure noise.

Following the prior specifications in Bhattacharya and Dunson (2011) and Durante and Dunson (2013), we assume independent Gaussian process (GP) priors for the parameters $\{\beta_p\}_{p=1}^P$, $\{\theta_i\}_{i=1}^N$, and $\{d_r\}_{r=1}^R$, where β_p , θ_i , and d_r vectors are all T -dimensional and have $T \times T$ covariance matrices with GP parameters c^β, c^θ , and c^d , respectively. On the other hand, we assume independent Normal prior for $\{\{u_{ir}^t\}_{i=1}^N\}_{r=1}^R\}_{t=1}^T$ and ϵ_{ij}^t . These priors depend on shrinkage parameters $\{\tau_p^\beta\}_{p=1}^P$, τ^θ , $\{\tau_r^d\}_{r=1}^R$, $\{\{\tau_{rt}^u\}_{r=1}^R\}_{t=1}^T$, and σ_e^2 from inverse-Gamma (IG) priors. Specifically, we let

1. $\beta_p \sim \text{GP}(0, \tau_p^\beta c_p^\beta)$ for $p = 1, \dots, P$, where $\tau_p^\beta \sim \text{IG}(a_\beta, b_\beta)$ and $c_p^\beta(t, t') = f(\kappa_p^\beta, |t - t'|)$;
2. $\theta_i \sim \text{GP}(0, \tau^\theta c^\theta)$ for $i = 1, \dots, N$, where $\tau^\theta \sim \text{IG}(a_\theta, b_\theta)$ and $c^\theta(t, t') = f(\kappa^\theta, |t - t'|)$;
3. $d_r \sim \text{GP}(0, \tau_r^d c_r^d)$ for $r = 1, \dots, R$, where $\tau_r^d \sim \text{IG}(a_d, b_d)$ and $c_r^d(t, t') = f(\kappa_r^d, |t - t'|)$;
4. $u_{ir}^t \sim \text{N}(0, \tau_{rt}^u)$ for $i = 1, \dots, N; r = 1, \dots, R$, and $t = 1, \dots, T$, where $\tau_{rt}^u \sim \text{IG}(a_u, b_u)$;
5. $\epsilon_{ij}^t \sim \text{N}(0, \sigma_e^2)$, where $\sigma_e^2 \sim \text{IG}(a_\sigma, b_\sigma)$.

For the inverse-Gamma hyperparameters, we specify values of a (shape) and b (scale) which may vary across β, θ, d , and u .

To effectively model the temporal evolutions, we parameterize the covariance matrices using the standard Exponential covariance function:

$$f(\kappa, d) = \exp\left(-\frac{d}{\kappa}\right),$$

where d is the Euclidean distance between two timepoints and the prior for κ is given as $\kappa \sim \text{half-Cauchy}(0, 5)$ —a relatively weakly-informative prior. When the timepoints are evenly spaced, the covariance matrix is a symmetric Toeplitz matrix, a matrix in which each descending diagonal from left to right is constant. To aid computational efficiency, we compute the inverse of Toeplitz matrices using Trench and Durbin methods (Golub and Van Loan, 2012).

Some papers (Bhattacharya and Dunson, 2011; Durante and Dunson, 2013, 2014b) use the squared exponential covariance function $f(\kappa, d) = \exp(-d^2/\kappa)$, which is appropriate for modeling very smooth functions. In addition, the parameter κ that characterizes length-scale of the process is fixed as an input variable in these works. However, smoothness is not necessarily required for discrete-time network modeling, and in practice, prior knowledge on how much the networks are correlated over time is unavailable. Choosing appropriate value of κ can be a challenge and thus the DAME model addresses this issue by estimating κ .

2.2 Posterior Computation

We take a Bayesian approach to infer the parameters of the DAME. 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, while we use Metropolis-Hastings to sample the pair of variance and length GP parameters (τ, κ) . This section outlines the steps and sampling equations we use for our MCMC updates of the DAME, where the proofs of each step can be found in the supplementary material.

To begin with, let E^t denote the $N \times N$ matrix of random noises, where the $(i, j)^{th}$ entry is defined as $E_{ij}^t = y_{ij}^t - \left(\sum_{p=1}^P \beta_p^t X_{ijp}^t + \theta_i^t + \theta_j^t + u_i^{t'} D^t u_j^t\right)$. Given that the distribution of the observed network $\mathbf{Y} = \{Y^1, \dots, Y^T\}$ conditional on all the parameters can be written as the product of Normal probability density function (pdf)

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

we sequentially update each parameter from its full conditional distribution in the following sampling steps, in which IG and MVN stand for inverse-Gamma and multivariate Normal:

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 $p \in \{1, \dots, P\}$ in random order, sample β_p as follows:
 - (a) Sample $(\tau_p^\beta, \kappa_p^\beta)$ using a bivariate Metropolis-Hastings algorithm (refer to Equation (7) in supplementary material)
 - (b) Sample $\beta_p \sim \text{MVN}_T(\tilde{\mu}_{\beta_p}, \tilde{\Sigma}_{\beta_p})$ with

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

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

3. Sample θ_i as follows:

- (a) Sample $(\tau^\theta, \kappa^\theta)$ using bivariate Metropolis-Hastings algorithm (refer to Equation (8) in supplementary material)
- (b) For each $i \in \{1, \dots, N\}$ in random order, sample $\theta_i \sim \text{MVN}_T(\tilde{\mu}_{\theta_i}, \tilde{\Sigma}_{\theta_i})$ with

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

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

4. For each $r \in \{1, \dots, R\}$ in random order, sample d_r as follows:

- (a) Sample (τ_r^d, κ_r^d) using univariate Metropolis-Hastings algorithm (refer to Equation (9) in supplementary material)
- (b) Sample $d_r \sim \text{MVN}_T(\tilde{\mu}_{d_r}, \tilde{\Sigma}_{d_r})$ with

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

where $E_{ij[-r]}^t = E_{ij}^t + u_{ir}^t u_{jr}^t$.

5. For each $t \in \{1, \dots, T\}$ and $i \in \{1, \dots, N\}$ in random order, sample u_i^t as follows:

- (a) For each $r \in \{1, \dots, R\}$, sample $\tau_{rt}^u \sim \text{IG}(\frac{N}{2} + a_u, \frac{1}{2} \sum_{i=1}^N (u_{ir}^t)^2 + b_u)$ and construct $R \times R$ covariance matrix $\tau_t^u = \text{diag}(\tau_{1t}^u, \dots, \tau_{Rt}^u)$
- (b) Sample $u_i^t \sim \text{MVN}_R(\tilde{\mu}_{u_i^t}, \tilde{\Sigma}_{u_i^t})$ with

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

where $E_{ij[-u]}^t = E_{ij}^t + u_i^t D^t u_j^t$.

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

2.3 Varying Number of Nodes

In many longitudinal networks, new nodes can join the network or existing nodes can disappear at any timepoint, so allowing for varying number of nodes is a critical feature of dynamic network models. For example, continuous-time network models which are built upon survival analysis (Butts, 2008; Vu et al., 2011) naturally allow the number of nodes to vary over time, and the stochastic actor oriented model of Snijders et al. (2010) adds data specification options to allow ‘joiners’ and ‘leavers’, by specifying structural zeros for all elements of the variables from actors (or nodes) who are absent at a given timepoint. Nevertheless, our adaptation of this idea to the current setting is novel in the literature as far as we know.

Our idea is similar to the one used in Snijders et al. (2010): We code a node that has not yet joined or that has dropped out as a ‘structural zero’. More precisely, we define the $N \times T$ matrix of availability A as an input to the model, where the (n, t) th element is then defined as

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

Here, N is the number of actors who are part of the network at any time $1 \leq t \leq T$. While missing values corresponding to nodes n and times t for which $A_{nt} = 1$ are updated along with parameter values at each MCMC iteration, when $A_{nt} = 0$ the corresponding additive effect estimates θ_n^t and latent position estimates u_n^t are specified as ‘not available (NA)’. Furthermore, other parameters including the fixed effects $\{\beta_p\}_{p=1}^P$ are estimated excluding the entries of structural zeros.

3 Simulation Study

We provide a simulation study to evaluate the performance of our proposed model on the 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 case of fitting network that is highly correlated across time, and 2) prove that the eigenvalue formulation of the multiplicative random effects (i.e. $u'Du$) has the benefit of revealing various type of transitivity effects.

3.1 Estimating Strong Correlations

We generated a set of relational data \mathbf{Y} for $N = 20$ and $T = 10$ according to the generative process in Section 2.1, with $P = 1$, $R = 2$, all inverse-Gamma hyperparameters $(a, b) = (2, 1)$, and $(\kappa^\beta, \kappa^\theta, \kappa^d) = (10, 10, 10)$ such that the resulting dynamic network is highly correlated across time. For example, the highest correlation of the parameters with $d = 1$ is $f(\kappa = 10, d = 1) = 0.905$, and the lowest correlation of the parameters with $d = 9$ is $f(\kappa = 10, d = 1) = 0.407$. We ran 6,000 MCMC iterations which proved to be enough for ensure full convergence, and then discarded the first 1,000 samples with thinning = 10.

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

$$dc_l = \rho(\text{degrees}^{1, \dots, (T-l)}, \text{degrees}^{(1+l), \dots, T}), \quad (6)$$

where $\rho(\cdot)$ is the Spearman correlation between two vectors, and $\text{degrees}^{1, \dots, (T-l)}$ and $\text{degrees}^{(1+l), \dots, T}$ indicate the vector of all degrees from timepoint $t = 1$ to $t = (T - l)$ and $t = (1 + l)$ to $t = T$, respectively, which result in both to be $N \times (T - l)$ -length vector.

Figure 1 compares the lagged degree correlations (dc) from our model and the “independence model” —same general framework but without estimating Gaussian process covariance parameters (by fixing all $\kappa = 0$), which are calucated from the degrees constructed from their respective posterior estimates. This comparison highlights the excellent performance of the DAME, in correctly estimating the true temporal correlation across the timepoints. These can be noticed by comparing with the true degree correlation statistics, where the posterior estimates from the DAME model perfectly recovered all the degree correlations from lag 1 to lag 3. On the contrary, the independence model always exhibit 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— when not estimating the correlation structure.

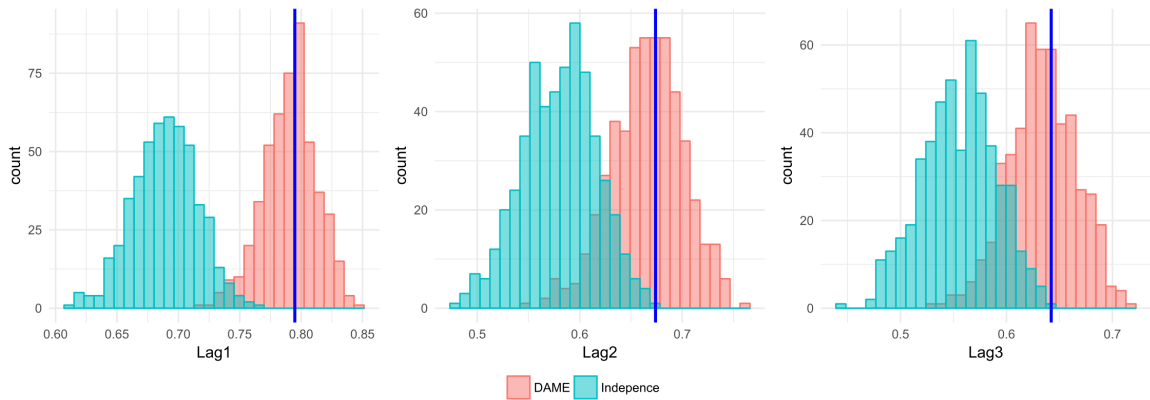


Figure 1: Histogram of posterior lagged degree correlations (from lag 1 to lag 3): the DAME (red) and independence model (green), with the observed dc statistic (blue lines).

3.2 Capturing Transitivity

As introduced in Section 1, there exists two types of transivities, positive and negative transitivity, and the parameterization of without D term (i.e. $u'u$) is not able to capture negative transitivity. To test if the DAME can explain both positive negative transivities, we conducted another set of simulation study. Similar to Section 3.1, we generated \mathbf{Y} with $N = 20$, $T = 10$, $P = 1$, $R = 2$, and $\forall(a, b) = (2, 1)$. Considering that our new goal is not to estimate temporal correlations but to represent transitivity effect, this time we modify two settings: 1) set $(\kappa^\beta, \kappa^\theta, \kappa^d) = (0, 0, 0)$ such that all paramters and the resulting dynamic network is independent over time, and 2) fix $d_r^t = \pm 2$ for $r = 1, 2$ and $t = 1, \dots, 10$ such that the generated network exhibits positive, mixed, or negative transitive feature, respectively. Again, we ran 6,000 MCMC iterations and discarded the first 1,000 with thinning = 10, however, we fixed κ 's at the true values and did not estimate the covariance paramters (i.e. independence model) such that the difference in performance only comes from the multiplicative effects formulations— $u'Du$ and $u'u$. If we estimated κ 's for this comparison, the difference in results may not only arise from the formulation of multiplicative effects, but also possibly come from lack of correlation structure in $u'u$ because our modeling framework do not assign any correlation on u 's.

Figure 2 shows a graphical comparison between the two formulations of the multiplicative random effects with respect to the first, second, and third moments of degree (i.e. $\text{degree}(\hat{Y})$, $\text{degree}(\hat{Y}^2)$ and $\text{degree}(\hat{Y}^3)$, respectively), for a randomly chosen node. For the case of positive transitivity ($\forall d_r^t = +2$), our model and its alternative (without D term) do not show significant differences; both formulations achieve great performance in replicating the first, second, and third moments of node 2's degree. On the contrary, when we fit the network with mixed ($\forall d_{r=1}^t = -2$ and $\forall d_{r=2}^t = +2$) or negative transitivity ($\forall d_r^t = -2$), the two formulations reveal noticeable differences. While the DAME can still recover the true degree in all three moments, the model without D term shows inaccuracy in simulating a network that is close to the true data. Not only the $u'u$ model introduces bias, but also it shows significantly lower precision than our model. In addition, the evidence of $u'u$ model's failure to estimate the transitivity statistics becomes larger as the network tends toward stronger negative transitivity and as we move to higher moments degrees. These findings strongly support our argument that $u'Du$ should be preferred over $u'u$, as a standard form of multiplicative random effects.

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 became the standard data sources to measure the state preferences. The study of state preferences in international politics is one of the most important topics for students of international relations (Wendt, 1994). With regard to policy implications, for instance, states are the key actors at the global stage, knowing their preferences towards each other and tastes on difference issues, helps us predict future foreign policies and state behaviors. Unfortunately, the existing studies ignored three important features of the dataset. First, votes are highly correlated across timepoints, because they are the reflections from history. Bailey et al. (2017) proposed a dynamic ordinal spatial IRT (item response theory) model and allowed for better inter-temporal comparisons, however, their model limited the temporal dependence to be lag 1 (i.e. Markovian assumption—votes at time $(t+1)$ is only dependent on votes at time t). Second, although the researchers have viewed “voting” as dyadic behavior and 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 has never been analyzed using the dynamic network models. Third, third-order dependence has not been modeled in the literature despite the fact that voting decisions are not limited to dyadic calculations—country A's decision to vote along the line with country B might well be influenced by a country C's decision. In order to overcome the limitations and provide the new insight at the same time, we apply the DAME to voting behavior in the United Nations General Assembly roll-call votes, from 1983 to 2014.

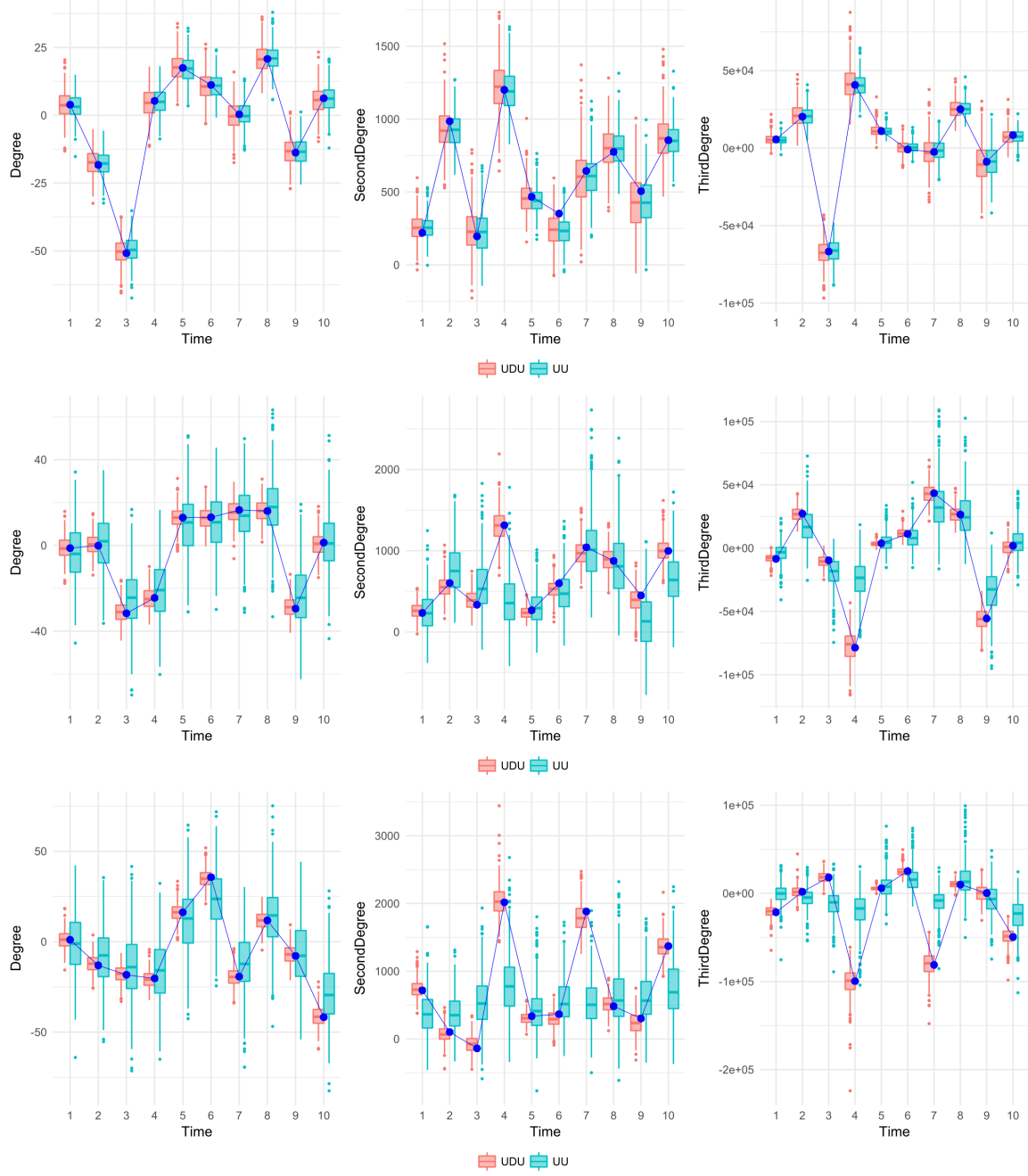


Figure 2: Posterior distribution of the first (left), second (center), and third (right) moment degrees using positive transitive network (upper), mixed transitive network (middle), and negative transitive network (lower): the DAME (red) and $u'u$ model (green), with the observed statistics (blue dots).

Here are the data processing procedures used in the analysis. First, the countries included in this analysis were determined by taking intersection of countries available for each variables (e.g. votes, polity score, GDP) and then dropping countries with missings (NA’s) over 10 years, while the remaining missings were imputed using previous year’s data. Full list of countries and their abbreviations are listed in Table A in Appendix A. Specifically, the voting data was obtained from Voeten et al. (2016), where we use the subset of the votes called ‘important votes’, the votes identified as important by U.S. State Department report Voting Practices in the United Nations¹. We did not include all votes from the original data in that non-important votes may not strongly reflect state preference of the countries. Throughout 1983 to 2014, the number of important votes on average is 12 per year, ranges from 6 to 28. We constructed the response $\mathbf{Y} = \{Y^1, \dots, Y^{32}\}$, where each Y^t is 97×97 matrix, using the variable ‘agree3unimportant’ in the original dataset, which is a voting similarity index (0–1) computed using 3 category vote data ($Y = \text{“yes”}$ or approval for an issue; $A = \text{abstain}$; $N = \text{“no”}$ or disapproval for an issue)². Note that abstention is counted as half-agreement with a yes or no vote (Voeten et al., 2016), while two abstention is treated as full-agreement. For basic summary of United Nations voting data including the average number of common votes and the average proportion of agreement or voting similarity index per year, see Table 3 in Appendix B.

As an exploratory data analysis, Table 1 illustrates the lagged degree correlation (defined in Equation 6) of the observed dataset to measure how strong United Nations voting data is correlated over time. There exists strong positive correlation in how countries vote in the United Nations General Assembly over time, and as the distance between two timepoints become larger the correlation gets weaker. This provides solid evidence that our model with Gaussian process assumption is one of the appropriate approaches to account for strong temporal dependence in the 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 of United Nations Voting Data for Lag 1 to 10.

Next, to dynamically model the 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 every $p = 1, \dots, 5$, we set the explanatory variable X_{ijp}^t as below.

1. Intercept $_{ij}^t$: constant 1,
2. log(distance) $_{ij}^t$: log of the geographic distance between country i and country j ,
3. Alliance $_{ij}^t$: 1 if country i and country j have alliance, and 0 otherwise,
4. Polity difference $_{ij}^t$: absolute difference in polity IV number between country i and country j ,
5. Lower trade-to-GDP ratio $_{ij}^t$: index of economic dependence using bilateral trade weighted by each country’s gross domestic product (GDP), as defined in Gartzke (2000)

$$\min\left(\frac{\text{Trade}_{ijt}}{\text{GDP}_{it}}, \frac{\text{Trade}_{ijt}}{\text{GDP}_{jt}}\right),$$

6. Common language $_{ij}^t$: indicator of whether country i and country j share the same language,

¹For 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’.

²Voting 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)

where all covariates are symmetric $X_{ijp}^t = X_{jip}^t$. Note that intercept is included to account for the baseline degree of agreement at each timepoint. Two variables, $\log(\text{distance})$ and common language are time-invariant covariates, although their coefficients may vary over time. Correlations between the five covariates (except the intercept) are attached in Appendix C.

Moreover, we specified the matrix of availability A introduced in Section 2.3 to reflect six countries' nonparticipation in the United Nations General Assembly:

1. North Korea (PRK) have structural zeros from $t = 1$ to $t = 8$:
 - North Korea did not vote until North Korea and South Korea have simultaneously admitted to United Nations in 1991.
2. South Korea (ROK) have structural zeros from $t = 1$ to $t = 8$:
 - South Korea did not vote until North Korea and South Korea have simultaneously admitted to United Nations in 1991.
3. Russia (RUS) has structural zeros from $t = 1$ to $t = 9$:
 - 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 zeros from $t = 13$ to $t = 21$:
 - The country did not participate the UNGA roll-call votes during 1995–2003. Under the rule of Saddam, Iraq was under severe sanctions from the international community including the United Nations since 1990.

Therefore, any missing values corresponding the country and missing period are treated as structural zeros. As in Section 2.3, other non-structural missing values are treated as random missing (and thus imputed).

4.2 Reduced-Rank Structure of Voting Network

In this section, we demonstrate the special low-rank structure of the voting network, which can be generalized into any other similarly constructed agreement networks. First, for simplicity, we assume static network by fixing $T = 1$ and consider the $N \times N$ adjacency matrix V^1 representing the network consists of single vote. Since each entry of V^1 denotes a voting similarity index (0-1) computed using 3 categories ("yes", "abstain", and "no"), the $(i, j)^{th}$ element V_{ij}^1 can only have three possible values,

$$V_{ij}^1 = \begin{cases} 1, & \text{agreement} \\ 0.5, & \text{half-agreement (one abstain)} \\ 0, & \text{disagreement} \end{cases}$$

which implies perfect 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 , and 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 V^1 to be 3, thus we can use low rank factorization of the matrix V^1 using $R = 3$. If we were to apply the DAME model to this type of single vote network (without additive effects and explanatory variables), the maximum size of dimension for the latent factors we could fit is $R = 3$ and the estimated latent factors can be viewed as the distinct constructs behind the vote.

Although it is worth understanding the constraint behind single votes, this is not a practical issue in modeling United Nations voting network, where we aggregate multiple votes per year (minimum number of important votes per year is 6). When we add M number of matrices representing each vote, the aggregated matrix $V^1 + V^2 + \dots + V^M$ gets closer to the full rank thus we no longer need to concern about the maximum dimension of the latent factors in fitting the DAME model. Moreover, including the observed covariates \mathbf{X} and additive random effects θ 's also relax the reduced rank structure, since the multiplicative latent effect $u'Du$ is modeled after we subtract those effects from the response matrix or array (See Section 2). Accounting for large variability in the observed covariates and additive random effects, the residuals no longer maintain the perfect transitivity structure.

4.3 Model Validation

We perform a model validation to check if the DAME is an accurate and credible model for the data and to understand the role of different effects at the same time. We fitted four different dynamic AME models, one with additive and multiplicative effects (DAME), one with only multiplicative effects (ME), one with only additive effects (AE), and the last without any random effects (NO), where all four models included 6 fixed effects in Section 4.1. We constructed 95% credible intervals of the degree statistics constructed from the model-simulated response $\hat{\mathbf{Y}}$ from the four separate models.

Figure 3 shows the comparison of posterior degree distributions of the country “Israel”. First of all, there is huge benefit of correcting the bias when we add additive effects, compared to the model with no random effects. Next, when we compare between AE and ME, it is noticeable that the multiplicative effects significantly reduced the interval of estimates. This happens since the additive effects have very small variances while the multiplicative effects have larger variance estimates, which significantly lowered the error variance estimates $\hat{\sigma}_e^2$ that plays a big role in the precision of simulated data. Our model with both additive and multiplicative effects (DAME) seems to perform similar to the ME model in this plot, however, further analysis showed that the DAME still outperform the ME model in terms of both accuracy and precision. Overall, not only the DAME model showed the most accurate estimates over time, but it also had the narrowest 95% credible intervals among the four. These findings emphasize the importance of including the additive and multiplicative term which enable the model to capture some features not explicable by fixed effects and maximize the model performance. More results and interpretation using the full DAME model are demonstrated in Section 4.4, and the posterior predictive plots checking the overall degree distributions (aggregate all nodes and timepoints) can be found in Appendix D.

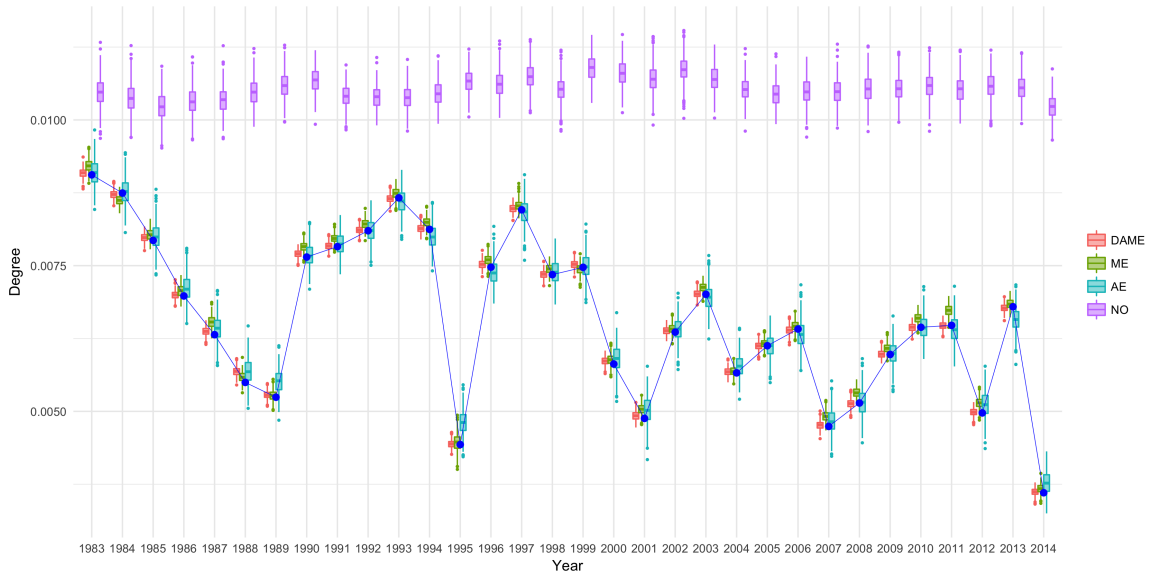


Figure 3: 95% credible intervals for posterior degree of Israel (ISR), comparing four models: the DAME (red), only multiplicative effects (green), only additive effects (blue), and no random effects (blue), with the blue dots represent the country’s observed degree statistics.

4.4 Parameter Estimation and Interpretation

In this section, we apply the model with the dimension of latent factors chosen as $R = 2$, based on some preliminary experiments where increasing the dimension of latent factors did not significantly improve the model fitting. For example, when $R = 3$, the last value of estimated eigenvalue d_3 is very close to zero for the most of the timepoints $t = 1, \dots, 32$. Thus, we present the results from the same settings as Section 4.3, which are 30,000 Gibbs iterations with a burn-in of 5,000, where

every 50th sample was taken as a thinning. All inverse-Gamma hyperparameters $(a, b) = (2, 1)$, and the covariance parameters (κ, τ) were estimated, following Section 2.2. Figure 4 shows the posterior mean estimates of the fixed effect coefficients $\{\beta_p\}_{p=1}^6$ with their corresponding 95% credible intervals. [Interpretation of fixed effect plot](#)

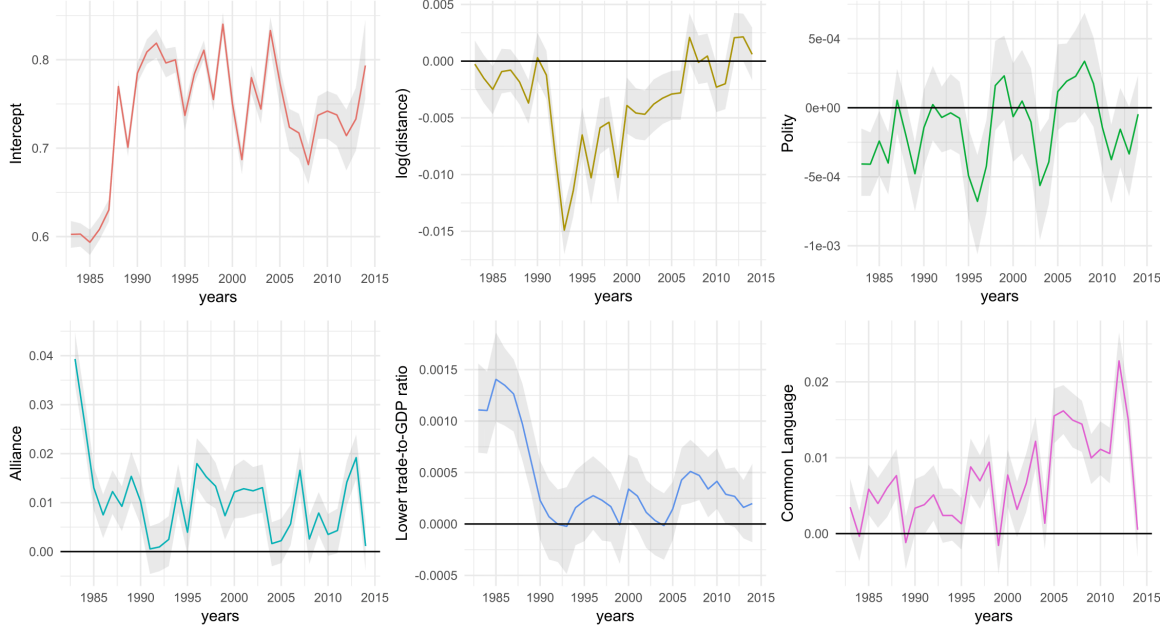


Figure 4: Posterior mean for the fixed effect coefficients $\{\beta_p\}_{p=1}^6$ (colored line): Intercept, log(distance), Alliance, Polity difference, Lower trade-to-GDP ratio, and Common Language, and their corresponding 95% credible intervals (grey areas).

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 results from 21 countries, where the countries are chosen based on the most active countries during the ten year period of 2004 – 2014 (Hoff, 2015). Here, the action types include negative material actions, positive material actions, negative verbal actions and positive verbal actions. The 21 countries are marked with * in Appendix A. Figure 5 is the posterior mean estimates of each node or country's additive random effects which capture the node-specific and within-dyad dependence. [Interpretation of additive random effects](#)

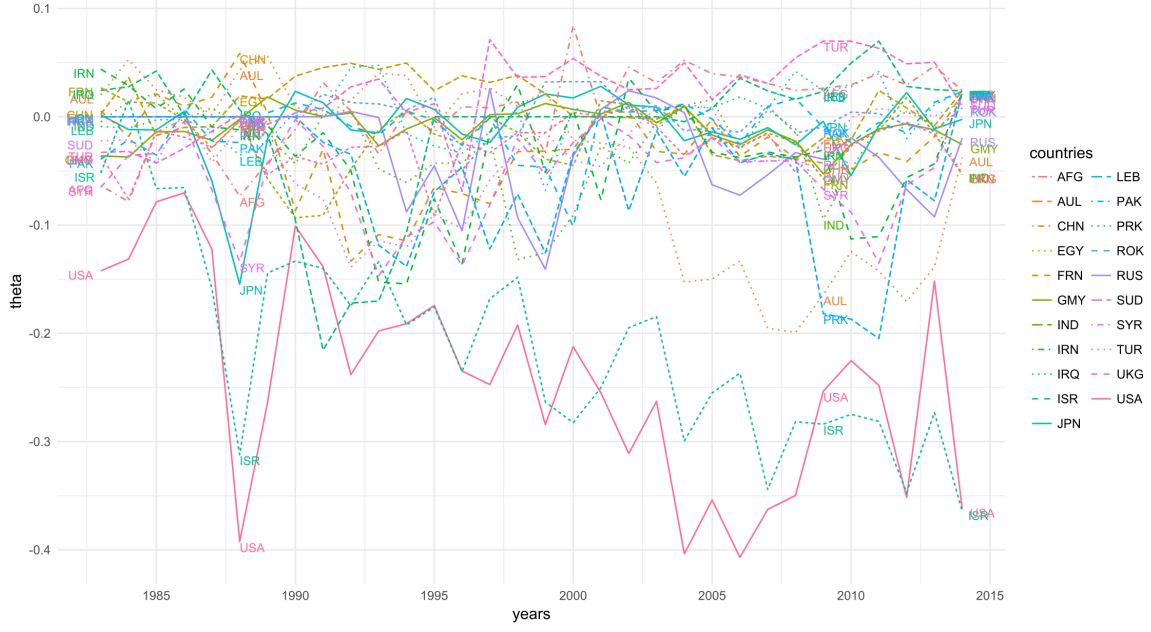


Figure 5: Posterior mean estimates for the additive random effect estimates θ .

The estimated latent factor coordinates in Figure 6, which can be viewed as the latent positions of the nodes, demonstrate the remaining higher-order network dependencies in United Nations voting behaviors. To determine the posterior estimates of u without identifiability issue, we applied eigen-decomposition on every posterior sample of the multiplicative effect matrix $u'Du$, and let D be the diagonal matrix of R -number of eigenvalues and u be the corresponding eigenvectors. For Figure 6, we applied Procrustes transformation on each posterior estimate of u and multiplied \sqrt{d} , and then obtained the 95% confidence regions. **Interpretations — What should be said about this multiplicative effect plots?**

As illustrated, the model estimated additive and multiplicative latent effects and the movement of latent positions reveal that United Nations voting network reflects interesting and meaningful foreign policy positions and alliance of various countries, even after controlling for other covariates that are considered critical in the international relation studies.

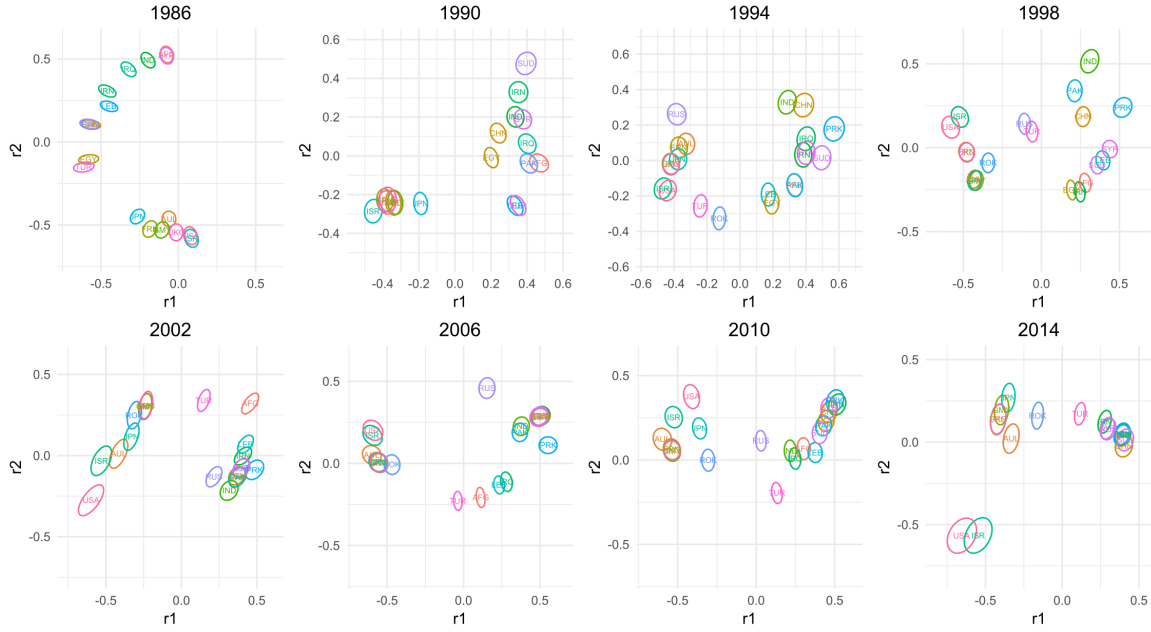


Figure 6: Posterior mean estimates for the multiplicative random effects and their corresponding 95% credible regions for the 21 selected countries for every 4 years between 1983 and 2014.

5 Discussion

Using the correlation associated with network at each time point makes better use of dynamic networks than modeling them as a separate network snapshots. Our algorithm eliminates the need to make an arbitrary user-defined covariance structure for determining the prior on the parameters, making it easier to understand the temporal dependence from the data which is highly informative and also lead to more accurate inference. As a time-varying coefficient extension of the additive and multiplicative effects (AME) model, the dynamic additive and multiplicative network effects (DAME) model can flexibly learn time-varying network structures, while inferring the effects of latent variables. 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 from the latent positions over time.

We have demonstrated such effectiveness of our model by modeling the United Nations Voting networks. Although we illustrated the entire framework in the context of symmetric or undirected networks, our model can be easily extended to allow directed networks, following the specification of 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 extremely large number of nodes or time-points, our model has a broad range of applicability, suggesting promising computational methods that can accommodate huge networks.

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.
- Golub, G. H. and Van Loan, C. F. (2012). *Matrix computations*, volume 3. JHU 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*.

- Marshall, M. G., Jaggers, 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.
- 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: List of 97 countries included in the analysis

B Summary of UN Voting Network

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 3: Summary of United Nations Voting Data: Average number of common votes (upper) and average proportion of agreement or voting similarity index (lower) per year

C Correlation between Independent Variables

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 4: Summary of United Nations Voting Data: Average number of common votes (upper) and average proportion of agreement or voting similarity index (lower) per year

D Posterior Predictive Checks on Degree Statistics

