Dynamic Stochastic Blockmodel Regression for Social Networks: Application to International Militarized Conflicts*

Santiago Olivella† Tyler Pratt‡ Kosuke Imai§

First Draft: July 12, 2018
This Draft: March 2, 2020

Abstract

A primary goal of social science research is to understand how latent group memberships predict the dynamic process of social network evolution. In the modeling of international conflicts scholars hypothesize that membership in geopolitical coalitions shapes the decision to engage in militarized conflict. Such theories explain the ways in which nodal and dyadic characteristics affect the evolution of relational ties over time via their effects on group memberships. To aid the empirical testing of these arguments, we develop a dynamic model of social networks by combining a hidden Markov model with a mixed-membership stochastic blockmodel that identifies latent groups underlying the network structure. Unlike existing models, we incorporate covariates that predict node membership in latent groups as well as the direct formation of edges between dyads. While prior substantive research often assumes the decision to engage in militarized conflict is independent across states and static over time, we demonstrate that conflict patterns are driven by states' evolving membership in geopolitical blocs. Changes in monadic covariates like democracy shift states between coalitions, generating heterogeneous effects on conflict over time and across states.

Keywords: clustering, hidden Markov model, latent variable modeling, mixed-membership stochastic blockmodel, network modeling

*The methods described in this paper can be implemented via the open-source statistical software, NetMix, available at `https://CRAN.R-project.org/package=NetMix`.

†Assistant Professor of Political Science, UNC-Chapel Hill. Email: `olivella@unc.edu`

‡Assistant Professor of Political Science, Yale University. Email: `tyler.pratt@yale.edu`

§Professor, Department of Government and Department of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge 02138. Email: `imai@harvard.edu`, URL: `https://imai.fas.harvard.edu/`

1 Introduction

Social scientists often posit theories about the effects of latent groups of actors on relational outcomes of interest over time. For example, international relations scholars have examined the so-called "democratic peace" hypothesis, which states that blocs of actors — defined by their democratic institutions — rarely engage in wars amongst themselves (e.g., Oneal and Russett, 1999). These theories often define latent groups of actors that underlie the structures of social networks, and stipulate how the formation and evolution of these groups give rise to various actions and behaviors (Lorrain and White, 1971).

To aid the empirical testing of these theories, we develop a dynamic model of social networks that extends the mixed-membership stochastic blockmodel (MMSBM; Airoldi et al., 2008). The MMSBM is a popular generalization of the stochastic blockmodel (SBM; Wang and Wong, 1987), which is a factor analytic model for network data characterized by latent groups of nodes (Hoff, 2009). Unlike the SBM, the MMSBM allows nodes to instantiate a variety of group memberships in their interactions with other nodes. Our proposed dynamic mixed-membership stochastic blockmodel (dynMMSBM) enables the memberships of latent groups to evolve over time, while simultaneously incorporating both dyadic and nodal attributes that affect the formation of groups and ties.

Our approach frees applied researchers from the need to resort to a commonly used two-step procedure to evaluate theories, whereby memberships are first estimated, and then regressed on covariates of interest (e.g., Wasserman and Faust, 1994). Furthermore, the proposed model allows for the prediction of group membership and future network ties of previously unobserved nodes. To facilitate the application of our proposed model, we develop a fast Bayesian inference algorithm by relying on a variational approximation to the collapsed posterior (Teh et al., 2007). We offer an open-source software R package, NetMix (available on CRAN) that implements the proposed methodology.

We use the dynMMSBM to conduct the dynamic analysis of international conflicts

among states over the last two centuries. Political scientists have long sought to explain the causes of interstate conflict and predict its outbreak. A prominent literature on the "democratic peace," for example, explores whether democratic countries constitute a uniquely peaceful community of states. A significant body of evidence attests to the low rate of conflict among democratic dyads (e.g., Maoz and Russett, 1993; Oneal and Russett, 1999). Others argue that the relationship is spurious, driven by impermanent geopolitical coalitions that generated common interests among democracies (e.g., Farber and Gowa, 1997; Gowa, 2011). Analysts of the democratic peace typically want to account for these underlying coalitions, and in particular ask whether democratic political systems encourage states to enter the same geopolitical blocs — a question our model is designed to address.

When analyzing conflict data, the most common methodological approach is to assume the conditional independence of state dyad-year observations given some covariates within the generalized linear model framework (e.g., Gleditsch and Hegre, 1997; Mansfield and Snyder, 2002; Gartzke, 2007; Dafoe et al., 2013). Recent analyses, however, have turned to network models to relax this conditional independence assumption. Maoz et al. (2006), for instance, use a measure of structural equivalence among dyads as a covariate in the logistic regression. Hoff and Ward (2004) employ random effects to explicitly model network dependence in dyadic data. Ward et al. (2007) apply the latent space model developed by Hoff et al. (2002) to international conflict. Cranmer and Desmarais (2011) propose and apply a longitudinal extension of the exponential random graph model (ERGM) to conflict data. We build on this emerging body of scholarship that seeks to model complex dependencies in the conflict network.

Methodologically, our work extends the growing literature on dynamic modeling of social networks that exhibit some degree of stochastic equivalence. In addition to the SBM, a variety of models are generally available to accommodate such networks. For instance, the latent position cluster model (Handcock et al., 2007) and the recently developed ego-

ERGM (Salter-Townshend and Brendan Murphy, 2015) incorporate equivalence classes into the latent distance and the ERGM models, respectively. Although the more flexible SBM (and all SBM-based models, such as ours) can capture disassortative relationships that these other models have a harder time accommodating, they all share the highly restrictive assumption that nodes play a single role in all their interactions.

Models like the overlapping/multiple-membership SBM (Latouche et al., 2011; Kim and Leskovec, 2013) or the mixed-membership SBM (MMSBM; Airoldi et al., 2008) fully address this issue by allowing nodes to belong to multiple equivalence classes. Typically, however, these models are limited by the fact that they impose the independence of group memberships over time and across nodes, as well as the independence of dyads conditional on the equivalence structure. This makes it difficult to accommodate networks that display both stochastic equivalence and some degree of heterogeneity across nodes (e.g. networks that have very skewed degree distributions).

Subsequent work therefore focuses on relaxing some of these independence assumptions. For instance, Sweet et al. (2014) incorporates dyadic covariates into the MMSBM, thus allowing for connectivity patterns that are not exclusively the result of the equivalence structure. And White and Murphy (2016) incorporates node-specific attributes as predictors of the mixed-membership vectors, thus eliminating the assumption that all nodes in an equivalence class are exchangeable. Recent work by Yan et al. (2019) has shown that likelihood-based estimators of these covariate effect parameters have desirable asymptotic properties, lending further confidence in the validity of these extensions. The proposed dynMMSBM derives from these developments, allowing for dyadic covariates at the edge-formation stage and for nodal predictors of the mixed-membership vectors.

Even more attention has been devoted to relaxing the assumption of independence of networks observed over time, resulting in important advances to apply the MMSBM in dynamic network settings (e.g. Xing et al., 2010; Ho and Xing, 2015; Fan et al., 2015). As

most social networks have a temporal dimension, being able to model the dynamic evolution of relational outcomes is of paramount importance to applied researchers. However, while these models offer flexible approaches to accounting for temporal dynamics, they often rely on continuous state space approaches like the Kalman filter, making it difficult to periodize a network's historical evolution.

Since researchers typically periodize history into distinct "epochs" to make sense of a phenomenon's evolution, more discrete approaches to network dynamics would be better suited to the typical needs of social scientists. Accordingly, the dynMMSBM relies on a hidden Markov process to capture the evolution of equivalence class-based network formation. Furthermore, by assuming that the blockmodel itself (i.e. the matrix of edge propensities across and within latent classes) remains constant over time—so that only memberships into classes are allowed to evolve—we avoid the issues of identification raised by Matias and Miele (2017) that affect some of the earlier dynamic MMSBM specifications.

To the best of our knowledge, our model is the first to tackle both the need to incorporate dyadic and nodal attributes and the need to account for temporal dynamics simultaneously, in an effort to develop an effective model that can be readily employed in applied research.

2 The Interstate Conflict Network

The study of interstate conflict is of great interest to international relations scholars and policy makers. The ability to predict violent political clashes has attracted a large and growing literature on conflict forecasting (e.g., Schrodt, 1991; Beck et al., 2000; Ward et al., 2013; Hegre et al., 2017). In addition, scholars have sought to understand how specific political institutions, processes, and power asymmetries affect war and peace among states (e.g., Barbieri, 1996; Oneal and Tir, 2006; Hegre, 2008).

Empirical studies of interstate conflict are most commonly conducted at the level of the state dyad-year. In these analyses, dyad-year observations are typically assumed to be independent conditional on the covariates included in the model (e.g., Maoz and Russett, 1993;

Farber and Gowa, 1997; Goldsmith, 2007; Gowa, 2011; Dafoe et al., 2013). However, there are reasons to believe conflict patterns violate this assumption. For centuries, states have managed conflict through formal and informal coalitions. Alliances, for example, affect the probability of conflict both among allied states and between allies and non-allies. Many militarized conflicts (most notably, the World Wars) are *multilateral* in nature: states do not decide to engage in conflict as a series of disconnected dyads, but are drawn into war or maintain peace as a result of their membership in preexisting groups.

To address this cross-sectional and temporal dependency, we propose a network model of interstate conflict that acknowledges the tendency of states to sort into geopolitical coalitions. In the dynMMSBM, nodal attributes influence the formation of unobserved groups, and the effect of these attributes can vary over time, depending on which "epoch" of history the interaction takes place in.

We examine the onset of militarized disputes among 216 states in the years 1816–2008, based on the Militarized Interstate Dispute (MID) dataset version 4.1 (Ghosn and Bennett, 2003). A MID occurs when one state engages in a government-sanctioned "threat, display or use of military force" against "the government, official representatives, official forces, property, or territory of another state" (Jones et al., 1996, 168). Ties in the network are formed when a new MID occurs between two states; subsequent years of the same dispute are coded as 0. The onset of a MID is a relatively rare event, occurring in approximately 0.4% of the 805,243 state dyad-year observations in our sample.

In defining the structural components of the dynMMSBM, we begin with the standard specification used in the democratic peace literature. This research agenda is among the most prominent theoretical debates in the study of interstate conflict, and it explores whether democracies engage each other in conflict at lower rates than other regime types. While recent network applications have re-examined the democratic peace debate (e.g., Hoff and Ward, 2004; Ward et al., 2007; Cranmer and Desmarais, 2011), the dynMMSBM

offers some distinct advantages.

First, the structure of the dynMMSBM can accommodate many theoretical mechanisms of the democratic peace theory about why democracies form a distinct community of states that have achieved a "separate peace" among themselves. This behavior may arise from the norms of compromise prevalent in democratic societies (Maoz and Russett, 1993), the ability of democratic states to credibly signal their intentions (Fearon, 1994), or the process by which democracies select into conflicts (Bueno de Mesquita et al., 2004).

The dynMMSBM could reveal such a community by identifying a latent group that exhibits low rates of intra-group conflict and that democratic states are more likely to join. Other hypotheses of the democratic peace — for example, the possibility of a similar "dictatorial peace" among autocratic states (Peceny et al., 2002) and interactions between democracy and power asymmetries (Bueno de Mesquita et al., 2004) — are also easily accommodated by the blockmodel structure. Each latent group is associated with its own set of nodal covariates and has unique rates of intra- and inter-group conflict.

Second, the direct inclusion of nodal variables obviates the need to restructure monadic covariates to fit a dyadic dataset, which has exacerbated a debate in the democratic peace literature regarding the appropriate dyadic specification of democracy (see Dafoe et al., 2013). Finally, the dynamic implementation provides flexibility for the effect of democracy to vary over time, as hypothesized by Farber and Gowa (1997).

3 The Proposed Methodology

In this section, we describe the proposed methodology. We first define the model and then derive a fast estimation algorithm based on a variational approximation.

3.1 The Dynamic Mixed-Membership Stochastic Blockmodel

Suppose that we observe a social network as graph $G_t = (V_t, E_t)$ for each time period $t \in \{1, \ldots, T\}$, where V_t and E_t represent a set of nodes and a set of directed edges,

respectively (undirected networks are discussed later in this section). We allow node set V_t to vary over time, since nodes may enter and exit the network at different points in time. In our application, some countries are born into or disappear from the international system during the study period. We use N_t to denote the number of nodes in V_t, i.e., $N_t = |V_t|$.

For each ordered pair of nodes pt and qt in V_t, we define an outcome variable $Y_{pqt} = 1$ if there is an edge from pt to qt in G_t, i.e., $(pt, qt) \in E_t$, and $Y_{pqt} = 0$ otherwise. We form an $N_t \times N_t$ sociomatrix $\mathbf{Y}_t$ with typical element Y_{pqt}. In addition, we observe a J_x-dimensional vector of time-varying covariates for each node pt, denoted by $\mathbf{x}_{pt}$, as well as a J_d-dimensional vector of time-varying covariates for each dyad (pt, qt), denoted by $\mathbf{d}_{pqt}$.

Like the SBM, the MMSBM makes the relational outcomes Y_{pqt} a function of K latent groups to which nodes belong. The distinctive feature of the MMBSM, however, is that while a node can belong to one latent group when interacting with the other node of a given dyad, different groups may be instantiated by the same node in other dyadic relationships. Group membership also depends on whether a node is playing a role of sender or receiver. We take this mixed membership framework one step further by allowing a node to belong to different latent groups across time periods, even when it is interacting with the same node in the same role (either as a sender or a receiver).

Formally, we define a K-dimensional indicator vector $\mathbf{z}_{pt\rightarrow qt}$ ($\mathbf{w}_{qt\leftarrow pt}$) whose kth element $z_{pt\rightarrow qt,k}$ ($w_{qt\leftarrow pt,k}$) is equal to one if node pt (qt) instantiates group k when interacting with node qt (pt) as a sender (receiver). As in the standard MMSBM, the group-by-group propensities of edge formation can be collected in a $K \times K$ matrix $\mathbf{B}$ — the so called *blockmodel*, or "role-compatibility matrix." For the purpose of identification, and unlike other approaches incorporating a time dimension in the MMSBM (e.g. Xing et al., 2010), we do not allow the blockmodel to vary over time (see Matias and Miele, 2017, for a discussion of problems associated with allowing the blockmodel to vary over time). That is, only node memberships into groups are allowed to differ across time periods.

We now describe the proposed model, dynMMSBM. We begin by modeling the edge indicator Y_{pqt} with a generalized linear model whose linear predictors consist of dyadic time-varying covariates $\mathbf{d}_{pqt}$, as well as a fixed effect specific to the interaction between two groups to which node pt and qt belong. This group interaction fixed effect is represented by the corresponding element of the blockmodel $\mathbf{B}$. This part of the model is given by

$$\begin{aligned} Y_{pqt} &\sim \text{Bernoulli}\left(\prod_{g=1,h=1}^{K} \theta_{pqtgh}^{z_{pt\to qt,g}\times w_{qt\leftarrow pt,h}}\right), \\ \theta_{pqtgh} &= g^{-1}\left(B_{gh} + \mathbf{d}_{pqt}^{\top}\boldsymbol{\gamma}\right), \end{aligned}$$

where B_{gh} is the (g,h)th element of the blockmodel, $\boldsymbol{\gamma}$ is a J_d-dimensional vector of coefficients, and $g(\cdot)$ is a link function. In our application, and given our sampling model for Y_{pqt}, we use the logistic link function. By including a set of dyadic predictors $\mathbf{d}_{pqt}$, the dynMMSBM allows tie formation probabilities to be different even for pairs of nodes that have instantiated the same latent groups at a given point in time.

In this model, each node has a time-specific probability of instantiating a group in any given interaction. We model these *mixed-membership* probability vectors, denoted $\boldsymbol{\pi}_{pt}$, as a time-specific mixture of M separate Dirichlet distributions with common concentration parameter ξ. We let the parameter vector of each Dirichlet distribution depend on the set of time-varying nodal covariates $\mathbf{x}_{pt}$. This enables the prediction of group memberships for nodes even when they are not part of the training set. Finally, we model the dynamic network evolution by defining a first-order hidden Markov model for the M mixture components of the the mixed-membership vectors. Thus, the coefficients of the dyadic covariates in the mean of the corresponding Dirichlet distributions are allowed to be different, depending on which hidden state the corresponding time period is in. Formally, we have,

$$\begin{aligned} \mathbf{z}_{pt\to qt} &\sim \text{Multinom}(1, \boldsymbol{\pi}_{pt}), \\ \mathbf{w}_{qt\leftarrow pt} &\sim \text{Multinom}(1, \boldsymbol{\pi}_{qt}), \\ \boldsymbol{\pi}_{pt} &\sim \prod_{m=1}^{M} \left[\text{Dirichlet}\left(\exp(\mathbf{x}_{pt}^{\top}\boldsymbol{\beta}_m)\right)\right]^{s_{tm}}, \end{aligned}$$

$$\begin{aligned}
\mathbf{s}_t &\sim \text{Multinom}(\mathbf{A}^\top \mathbf{s}_{t-1}), \\
\mathbf{s}_1 &\sim \text{Multinom}(\boldsymbol{\lambda}, 1),
\end{aligned}$$

where $s_{tm} = 1$ when time period t is in hidden state m ($s_{tm} = 0$ otherwise), $\boldsymbol{\beta}_m$ is a $J_x \times K$ matrix of state-specific coefficients, $\mathbf{A}$ is an $M \times M$, row-normalized matrix of transition probabilities, and $\boldsymbol{\lambda}$ is the M-dimensional vector of prior probabilities over initial states.

To complete the model, we specify the following prior distributions,

$$\begin{aligned}
\mathbf{A}_m^\top &\sim \text{Dirichlet}(\mathbf{1}, \eta), \\
B_{gh} &\sim \text{N}(\mu_{gh}, \sigma_{gh}^2), \\
\boldsymbol{\beta}_{km} &\sim \text{N}(\boldsymbol{\mu}_\beta, \sigma_\beta^2 \mathbf{I}), \\
\boldsymbol{\gamma} &\sim \text{N}(\boldsymbol{\mu}_\gamma, \sigma_\gamma^2 \mathbf{I}),
\end{aligned}$$

where $\mathbf{A}_m$ is the mth row of $\mathbf{A}$, η is the hyperprior concentration parameter of a symmetric Dirichlet distribution, and μ_{gh} and σ_{gh} are hyperprior location and scale parameters for the intensity of affinity between corresponding groups. We choose the values of location and scale hyperprior parameters, $\boldsymbol{\mu}_\beta$, $\boldsymbol{\mu}_\gamma$, σ_β^2 and σ_γ^2, to help regularize the model fit.

Thus, according to the model, the full joint distribution of data $\mathbf{Y}$, latent variables $\mathbf{Z}$, $\mathbf{W}$, $\boldsymbol{\Pi}$, $\mathbf{S}$, and parameters $\{\mathbf{B}, \mathbf{A}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ is given by,

$$\begin{aligned}
&P(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\Pi}, \mathbf{S}, \mathbf{B}, \mathbf{A}, \boldsymbol{\beta} \mid \mathbf{X}, \mathbf{D}) \\
&= \prod_{m=1}^{M} \left[P(s_{1m}) \prod_{t=2}^{T} P(s_{t,m} \mid \mathbf{s}_{t-1}, \mathbf{A}) \right] \prod_{t=1}^{T} P(\mathbf{Y}_t, \mathbf{Z}_t, \mathbf{W}_t, \boldsymbol{\Pi}_t \mid \mathbf{s}_t, \mathbf{B}, \boldsymbol{\beta}_m, \boldsymbol{\gamma}, \mathbf{X}_t, \mathbf{D}_t) \\
&\qquad \times P(\boldsymbol{\gamma}) \prod_{g,h=1}^{K} P(B_{gh}) \prod_{m=1}^{M} \left[P(\mathbf{A}_m) \prod_{k=1}^{K} P(\boldsymbol{\beta}_{km}) \right],
\end{aligned}$$

where

$$\begin{aligned}
&P(\mathbf{Y}_t, \mathbf{Z}_t, \mathbf{W}_t, \boldsymbol{\Pi}_t \mid \mathbf{S}_t, \mathbf{B}, \boldsymbol{\beta}_m, \boldsymbol{\gamma}, \mathbf{X}_t, \mathbf{D}_t) \\
&= \prod_{pt,qt \in V_t} \Big[P(Y_{pqt} \mid \mathbf{z}_{pt \to qt}, \mathbf{w}_{qt \leftarrow pt}, \mathbf{B}, \boldsymbol{\gamma}, \mathbf{D}_t) P(\mathbf{z}_{pt \to qt} \mid \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{qt \leftarrow pt} \mid \boldsymbol{\pi}_{qt}) \Big]
\end{aligned}$$

$$\times \prod_{pt \in V_t} \prod_{m=1}^{M} P(\boldsymbol{\pi}_{pt} \mid \mathbf{s}_t, \boldsymbol{\beta}_m, \mathbf{x}_{pt})^{s_{tm}}.$$

This framework can be adapted to handle undirected networks with only minor revisions. In such cases, both the outcome matrix $\mathbf{Y}_t$ and the blockmodel $\mathbf{B}$ will be symmetric, as the distinction between a sender a receiver role becomes unnecessary. Accordingly, and to avoid redundancies, products over pairs of nodes $pt, qt \in V_t$ are now taken over pairs such that $qt > pt$ at any given time. Otherwise, the model definition remains identical.

3.2 Marginalization

As discussed in Section 3.3, we define a factorized approximation to the posterior distribution of our model's parameters in order to drastically reduce the computation time. A typical approximating distribution would factorize over all parameters. In the true posterior, however, latent variables $\mathbf{z}_{pt \to qt}$ ($\mathbf{w}_{qt \leftarrow pt}$) and the mixed-membership parameters $\boldsymbol{\pi}_{pt}$ ($\boldsymbol{\pi}_{qt}$) are usually strongly dependent (Teh et al., 2007). Similarly, the Markov state indicators $\mathbf{s}_t$ and parameters in the transition kernel $\mathbf{A}$ are normally strongly correlated in the true posterior. Therefore, we marginalize out the latent mixed-membership vectors and the Markov transition probabilities and then approximate the marginalized posterior.

We first focus on the portion of the joint density that involves $\boldsymbol{\Pi}$. Define,

$$\alpha_{ptmk} \;=\; \exp(\mathbf{x}_{pt}^{\top} \boldsymbol{\beta}_{km}), \tag{1}$$

as the kth element of a K-dimensional vector that serves as the parameter of the Dirichlet distribution from which mixed memberships are drawn, and let $\xi_{ptm} = \sum_{k=1}^{K} \alpha_{ptmk}$. We then marginalize $\boldsymbol{\Pi}$ as follows,

$$\begin{aligned}
&\int \prod_{t=1}^{T} \prod_{pt \in V_t} \left[\prod_{m} P(\boldsymbol{\pi}_{pt} \mid \boldsymbol{\alpha}_{ptm})^{s_{tm}} \right] \prod_{qt \in V_t} P(\mathbf{z}_{pt \to qt} | \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{pt \leftarrow qt} \mid \boldsymbol{\pi}_{pt}) \, \mathrm{d}\boldsymbol{\Pi}_{pt} \\
&= \prod_{t=1}^{T} \prod_{pt \in V_t} \prod_{m=1}^{M} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ptmk} + C_{ptk})}{\Gamma(\alpha_{ptmk})} \right]^{s_{tm}},
\end{aligned} \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function, and $C_{ptk} = \sum_{qt \in V_t} (z_{pt \to qt,k} + w_{pt \leftarrow qt,k})$ represents the number of times node pt instantiates group k across its interactions with all other nodes

qt present at time t, whether as a sender or as a receiver. Note that we replace $\sum_{k=1}^{K} C_{ptk}$ with $2N_t$ because all nodes must instantiate exactly one group when interacting with other nodes at any given t—once as a sender and once more as a receiver. In the undirected case, this term reduces to N_t (see Appendix A.1).

Furthermore, the transition probabilities have independent Dirichlet priors that are conjugate to the multinomial distribution over states at any given time. Thus, we can adopt a similar strategy when marginalizing over the rows of $\mathbf{A}$. We focus on the portion of the joint distribution that involves $\mathbf{A}$, and marginalize $\mathbf{A}$ as follows,

$$\int \prod_{t=2}^{T} P(\mathbf{s}_t \mid \mathbf{s}_{t-1}, \mathbf{A}) \prod_{m=1}^{M} P(\mathbf{A}_m)\, \mathrm{d}\mathbf{A} \;=\; \prod_{m=1}^{M} \frac{\Gamma(M\gamma)}{\Gamma(M\gamma + U_m)} \prod_{n=1}^{M} \frac{\Gamma(\gamma + U_{mn})}{\Gamma(\gamma)},$$

where $U_{mn} = \sum_{t=2}^{T} s_{tn} s_{t-1,m}$ is the number of times the Markov chain transitions from state m to state n, and $U_m = \sum_{t=2}^{T} \sum_{n} s_{tn} s_{t-1,m}$ is the total number of times the Markov chain transitions from m (potentially to stay at m).

Putting it all together, the marginalized posterior distribution is given by,

$$\begin{aligned} & P(\mathbf{Z}, \mathbf{W}, \mathbf{S}, \mathbf{B}, \boldsymbol{\beta} \mid \mathbf{X}, \mathbf{D}, \mathbf{Y}) \\ \propto & \prod_{m=1}^{M} \left[\frac{\Gamma(M\eta)}{\Gamma(M\eta + U_m)} \prod_{n=1}^{M} \frac{\Gamma(\eta + U_{mn})}{\Gamma(\eta)} \right] P(\mathbf{s}_1) \prod_{t=2}^{T} \prod_{m=1}^{M} \prod_{pt \in V_t} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ptmk} + C_{ptk})}{\Gamma(\alpha_{ptmk})} \right]^{s_{tm}} \\ & \times \prod_{t=1}^{T} \prod_{pt,qt \in V_t} \prod_{g,h=1}^{K} \left(\theta_{pqtgh}^{Y_{pqt}} (1 - \theta_{pqtgh})^{1 - Y_{pqt}} \right)^{z_{pt \to qt, g} w_{qt \leftarrow pt, h}} P(\mathbf{B}) P(\boldsymbol{\beta}) P(\boldsymbol{\gamma}). \end{aligned} \tag{3}$$

3.3 Estimation via Variational Approximation

For posterior inference, we rely on a mean-field variational approximation to the marginalized posterior distribution (Jordan et al., 1999; Teh et al., 2007). We first define a factorized distribution of the latent variables $\mathbf{Z}$, $\mathbf{W}$, and $\mathbf{S}$ as,

$$Q(\mathbf{S}, \mathbf{Z}, \mathbf{W} \mid \mathbf{K}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) \;=\; \prod_{t=1}^{T} Q_1(\mathbf{s}_t \mid \boldsymbol{\kappa}_t) \prod_{pt,qt \in V_t} Q_2(\mathbf{z}_{pt \to qt} \mid \boldsymbol{\phi}_{pt \to qt}) Q_2(\mathbf{w}_{qt \leftarrow pt} \mid \boldsymbol{\psi}_{qt \leftarrow pt}),$$

where $\boldsymbol{\kappa}_t$, $\boldsymbol{\phi}_{pt \to qt}$, and $\boldsymbol{\psi}_{qt \leftarrow pt}$ are variational parameters.

We use this factorized distribution to bound the log posterior from below. We then iterate between finding an optimal $\widetilde{Q}$ (the E-step) and optimizing the corresponding lower-bound with respect to parameters $\mathbf{B}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ (the M-step). Below, we provide our variational EM algorithm. Appendix A.2 contains its complete derivation.

The variational update for the parameters in the distribution of $\mathbf{Z}$ is given by,

$$\hat{\phi}_{pt\to qt,k} \propto \prod_{m=1}^{M} \left[\exp\left[\mathbb{E}_{\widetilde{Q}_2}[\log(\alpha_{ptmk} + C'_{ptk})]\right]\right]^{\kappa_{tm}} \prod_{g=1}^{K} \left(\theta_{pqtkg}^{Y_{pqt}}(1-\theta_{pqtkg})^{1-Y_{pqt}}\right)^{\psi_{qt\leftarrow pt,g}},$$

where $C'_{ptk} = C_{ptk} - z_{pt\to qt,k}$ and the expectation is taken over the variational distribution of $\mathbf{Z}$. This corresponds to the (unnormalized) probability vector in a multinomial distribution. By symmetry, the update for $\psi_{qt\leftarrow pt,k}$ is similarly defined, and the two updates associated with a dyad can be computed in parallel to speed up computation. Also, in order to avoid costly computation of the Poisson-Binomial probability mass function (which is required given the nature of our C'_{ptk} term), we approximate the expectations in these updates by using a zeroth-order Taylor series expansion, so that $\mathbb{E}_{\widetilde{Q}_2}[\log(\alpha_{ptmk} + C'_{ptk})] \approx \log\left(\alpha_{ptmk} + \mathbb{E}_{\widetilde{Q}_2}\left[C'_{ptk}\right]\right)$ (Asuncion et al., 2009).

For $t = 2, \ldots, T-1$, the variational updates for the parameters of $\mathbf{S}$ are given by,

$$\begin{aligned}
\hat{\kappa}_{tm} \propto\ & \exp\left[-\mathbb{E}_{\widetilde{Q}_1}[\log(M\eta + U'_m)]\right] \exp\left[\kappa_{t+1,m}\kappa_{t-1,m}\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{mm} + 1)]\right] \\
& \times \exp\left[(\kappa_{t-1,m} - \kappa_{t-1,m}\kappa_{t+1,m} + \kappa_{t+1,m})\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{mm})]\right] \\
& \times \prod_{n\neq m} \exp\left[\kappa_{t+1,n}\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{mn})]\right] \prod_{n\neq m} \exp\left[\kappa_{t-1,n}\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{nm})]\right] \\
& \times \prod_{pt\in V_t} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \prod_{k=1}^{K} \frac{\mathbb{E}_{\widetilde{Q}_1}[\Gamma(\alpha_{ptmk} + C_{ptk})]}{\Gamma(\alpha_{ptmk})}\right],
\end{aligned}$$

where $U'_m = U_m - s_{t,m}$ and $U'_{mn} = U_{mn} - s_{tm}s_{t+1,n}$. This definition of the term U'_{mn} is valid whenever $m \neq n$ and $t \neq T$ (for other cases, see Appendix A.2). Once again, this corresponds to the (unnormalized) probability vector in a multinomial distribution.

To obtain the estimates of the regression parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and the blockmodel $\mathbf{B}$, we find optimal values with respect to the approximate lower bound, defined as the log expectation of equation (3) over the variational distribution minus the entropy of the variational

parameters. The resulting product-of-multinomials form of $\widetilde{Q}$ (which relied only on a factorizing assumption) allows us to compute the necessary expectations. To find optimal values, we use an iterative quasi-Newton algorithm, and provide the gradients required for this step in Appendix A.2.

Finally, we compute standard errors for all regression parameters by first sampling from the approximate posteriors of the latent variables, and then obtaining expected values of the log-posterior Hessian evaluated at the approximate MAP estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\mathbf{B}$.

3.4 Implementation Details

Like other mixed-membership models, there are important practical considerations when fitting the dynMMSBM. First, as with any EM-type algorithms, finding good starting values is essential. In particular, the quality of starting values for the sufficient statistics in the C_{ptk} term of equation (2) proved to be highly consequential. In our experience, two approaches worked similarly well: an initial clustering based on a spectral decomposition of the network's adjacency matrix (Jin et al., 2018), and taking a few samples from the posterior of the simpler mixed-membership stochastic blockmodel (without covariates) of Airoldi et al. (2008). We apply these strategies separately to each time-stamped network, and resolve the ensuing label-switching problem by re-aligning the (assumed constant) blockmodels using a graph matching algorithm (Lyzinski et al., 2014).[1] Other parameters and latent variables can be easily initialized using (generalized) linear models (in the case of regression coefficients) or k-means clustering (in the case of sufficient statistics for the $U_{\cdot}$ terms in equation (3)).

Finally, although it is typical to use absolute change in the lower bound as a criterion for evaluating convergence of variational Bayes models, the lack of a meaningful scale for this quantity makes choice of a tolerance level difficult. As an alternative, we monitor whether

[1] For applications with a small number of latent blocks, an exact solution based on enumeration of all group label permutations is also feasible.

the percent change in the lower bound falls below a user-supplied tolerance level. In our applications, we used a tolerance of 10^{-3} for this purpose.

3.5 A Simulation Study

Using synthetic dynamic networks, we evaluate the estimation accuracy with respect to the mixed-membership vectors and the blockmodel matrices under three scenarios: easy, realistic, and hard learning problems. We also examine the quality of regression coefficient estimates, and the ability of the model to recover the parameters associated with the underlying HMM. Finally, we compare the results of fitting a fully specified dynMMSBM and fitting a separate MMSBM (without covariates) to each time period, showing the substantial gains in error reduction resulting from the use of our proposed model. Given space limitations, the details and results of the simulation study are given in Appendix A.3.

4 Empirical Analysis

We apply the dynMMSBM to the interstate conflict network data described in Section 2. We show that the proposed model uncovers the essential geopolitical coalitions that drive conflict patterns and generates novel insights into the heterogeneous effect of key covariates, like democracy. Finally, we demonstrate that the dynMMSBM outperforms the standard logistic regression model in forecasting future conflicts.

4.1 The Setup

We model conflict as an undirected network in which ties arise from states' evolving membership in six latent groups. While the substantive results presented below are not conditional on the number of latent groups, we found that six provided sufficient flexibility to model different types of evolving coalitions that can be qualitatively interpreted. Six latent groups also performed the best when evaluating out-of-sample prediction compared to other numbers of groups (see Appendix A.4).

We analyze the onset of militarized disputes among 216 states in the years 1816–2008. We include two node-level covariates $\mathbf{x}_{pt}$ — the degree of democracy in a state's domestic government and the state's military capability — that are hypothesized to influence membership in the latent groups (Maoz and Russett, 1993; Hegre, 2008). We measure levels of democracy using the variable `POLITY`, from the Polity IV dataset (Marshall et al., 2017). States are assigned a polity score each year ranging from -10 to 10, with higher values representing more democratic political institutions. The mean polity score in our sample is -0.43. Roughly six percent of state years are assigned the minimum score of -10, and 16% receive the maximum of 10. Moreover, to measure the military capability of states (`MILITARY CAPABILITY`), we use version 5.0 of the composite index (CINC scores, Singer et al., 1972), and take the log to account for its skewed distribution. The association between these covariates and the latent group memberships is assumed to depend on two hidden Markov states.

In addition, we include four dyadic variables $\mathbf{d}_{pqt}$ that are expected to predict conflicts beyond the effects of the equivalence classes induced by the blockmodel. These include a dichotomous indicator for a formal alliance (`ALLIANCE`) between states in a given year; data on alliances comes from version 4.1 of the COW Formal Alliances dataset (Gibler, 2009). We also include geographic distance (`DISTANCE`) and the presence of a contiguous border (`BORDER`) between states (Stinnett et al., 2002). A count of common memberships in international organizations (`IO CO-MEMBERSHIPS`) addresses the possibility that interaction in these organizations decreases conflict (Oneal and Russett, 1999). Following the literature, a count of years since the last militarized dispute between each dyad and a cubic spline control for further temporal trends (Beck et al., 1998). Finally, to account for the missing values of some predictors, we rely on a missing-indicator approach, adding dummy variables that indicate which observations have missing values in the corresponding variable, and replacing all missing values with zero.

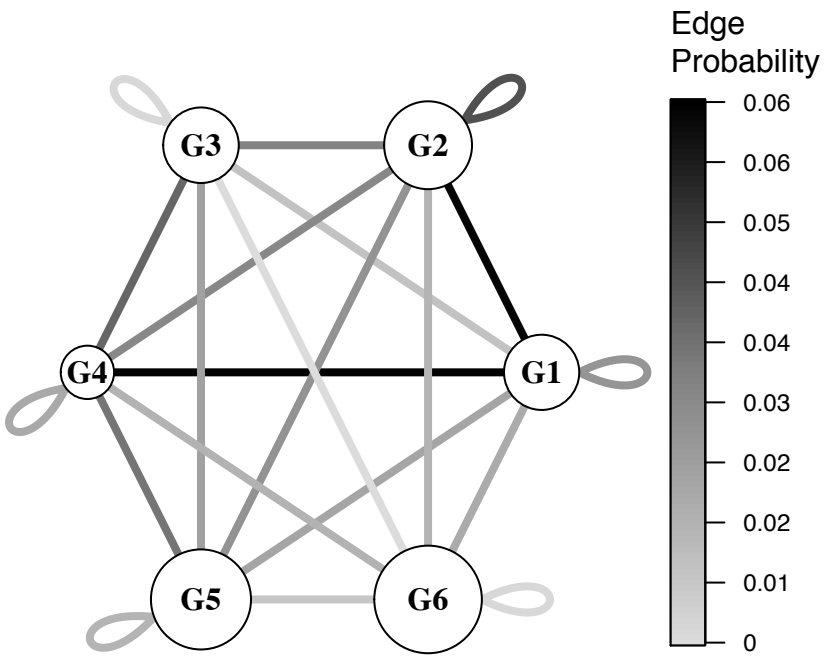

Figure 1: **Estimated blockmodel in the conflict network**. The figure displays a representation of the estimated blockmodel in the form of a network graph. The size of the nodes (circles) in the figure reflect the frequency with which countries instantiate membership in each group. Weighted edges (lines) represent the probability of conflict between groups, with darker edges indicating a higher propensity to start a conflict.

The model is fitted using our open-source software package NetMix. Estimation took one hour and sixteen minutes on a laptop computer with a 1.6Ghz CPU, converging after 25 EM iterations. Note that the estimation time drops to approximately 50 minutes without the optional Hessian computation, which calculates standard errors for the blockmodel, monadic, and dyadic coefficients.

4.2 Memberships in the Latent Groups

The dynMMSBM allows us to characterize membership in each latent group as well as the expected relationships between them. Figure 1, which depicts each latent group as a node on a graph, summarizes which groups tend to be instantiated most often. The figure also shows the estimated rates of conflict between groups, using darker-shaded edges to indicate a higher propensity of conflict onset. Table 3 of Appendix A.4 presents the exact blockmodel estimates used to create this figure.

The size of the nodes (circles) in the figure reflects the frequency with which states

instantiate membership in each group. Group 6 is the most populous group, representing 27% of state-year observations in the sample. Group 5 is the second largest (24%), followed by Groups 2 (19%), 3 (13%), 1 (13%), and 4 (3%).

Group 1 has elevated rates of inter-group conflict with groups 2 and 4, as evidenced by the darker shade of the edge between them in the figure. When a state from Group 1 interacts with a state from Group 2, there is a 6.2% chance that a militarized dispute will occur between them. Probing the mixed-membership vectors of individual states reveals that these two groups capture geopolitical divisions between blocs of powerful states. The United States, United Kingdom, Germany, and their Western European allies often instantiate Group 1, while China, Russia, and other Eastern bloc states tend to instantiate membership in Group 2.

Other groups also reveal important structure in the international system. Group 3 includes many states that maintained a foreign policy of neutrality throughout much of the 20th century (e.g., Ireland, Sweden, Switzerland, and Costa Rica). Despite their formally neutral stance, these states maintained close diplomatic relations with the Western allies that populate Group 1. According to the blockmodel, these states have low rate of conflict with Group 1 (1.2%) and are less bellicose overall. Group 4 includes many states that were caught in the crossfire of the intense geopolitical conflict between the Western and Eastern coalitions represented by Groups 1 and 2. Afghanistan, Cambodia, and Yugoslavia are among the states with high membership in Group 4 that were sites of proxy conflicts during the Cold War period. Group 5 is composed of many autocratic states in the Middle East and Africa, while Group 6 features small or geographically remote states.

A closer evaluation of estimated memberships during the Cold War period lends further credence to the validity of the model. As noted earlier, the Cold War period was defined by a geopolitical rivalry between an Eastern bloc, led by the Soviet Union, and a Western bloc, led by the United States and its NATO allies. To see if the dynMMSBM can recover

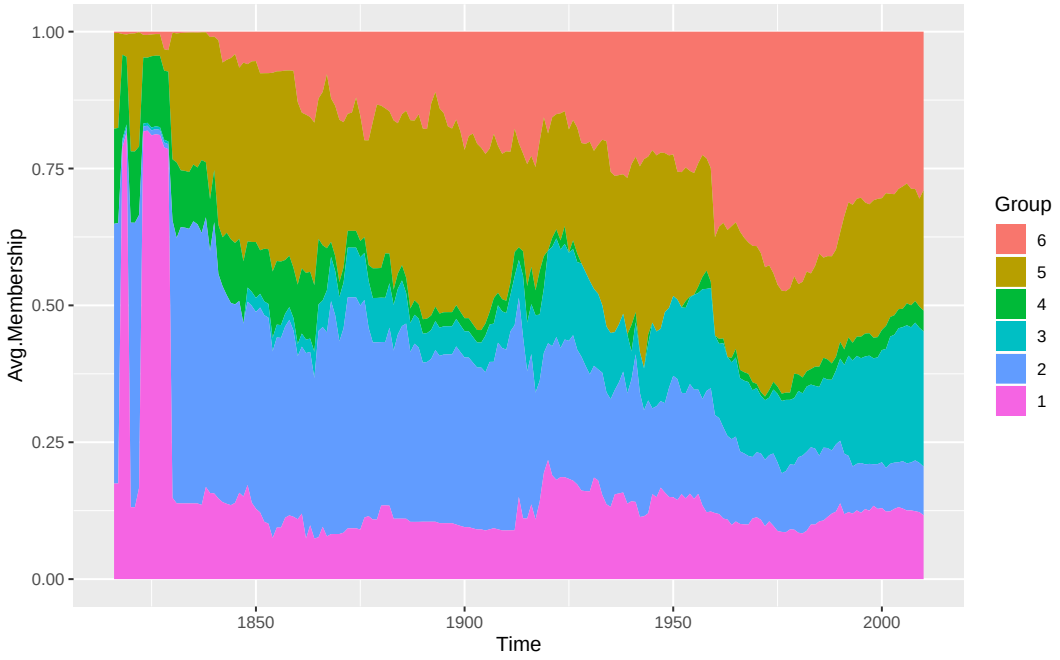

Figure 2: **Membership in Latent Groups over Time**. The figure shows the average proportion of membership in six latent groups for each year from 1816–2010.

the underlying geopolitical structure of the Cold War era, we identify the 15 states with the highest average membership probability in each latent group during the period of 1950–1990. We do this by estimating $\frac{1}{T}\sum_{t=1950}^{1990}\pi_{ptg}$ for every state given each latent group g. The states with the highest membership in each group are listed in Table 4 of Appendix A.4.

The distribution of states across the groups is consistent with presence of competing geopolitical coalitions during the Cold War. Group 1 contains the major NATO allies, such as the United States, United Kingdom, Germany, Italy and Canada. Non-NATO members that sided with the NATO, including Japan and Australia, also instantiate Group 1 at high rates. Group 2 consists of the Soviet Union and its allies in the Eastern bloc (Russia/Soviet Union, China, East Germany, Czechoslovakia, and Romania). The estimated blockmodel indicates the competing coalitions experience abnormally high rates of conflict.

4.3 The Dynamics of Membership

The dynMMSBM further allows us to examine how latent group membership changes over time. Figure 2 displays the evolution of group membership over time. Latent groups expand

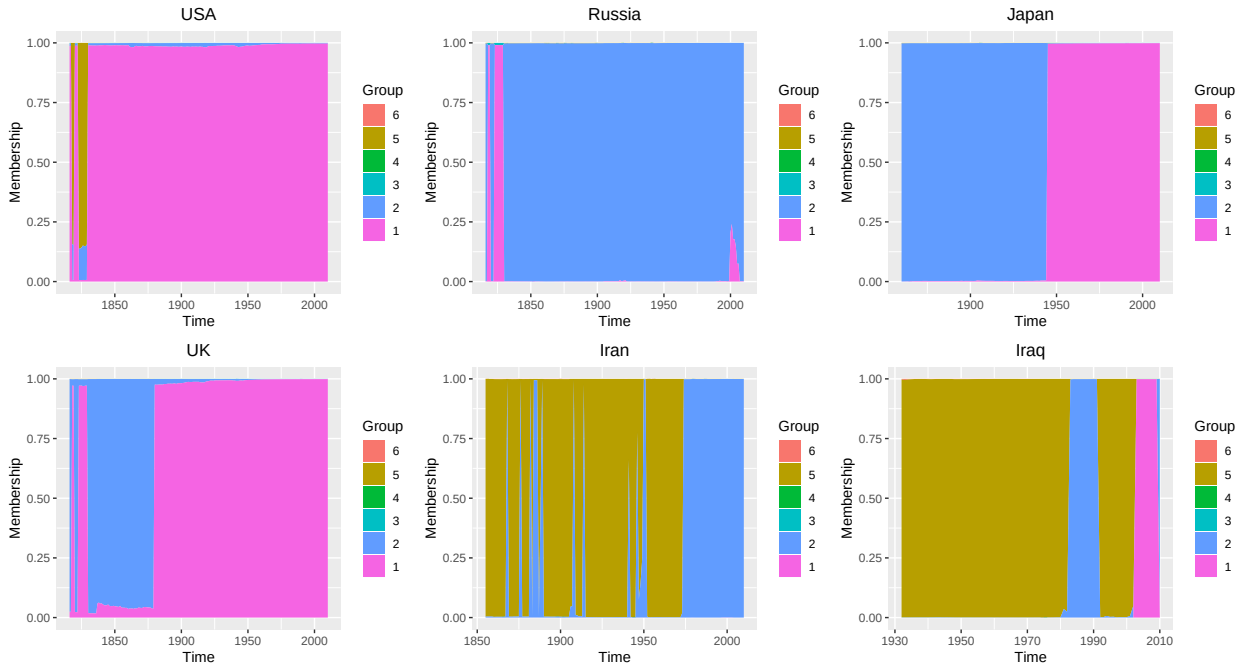

Figure 3: **Average Node Membership over Time, Select States**. The figure shows, for six states, the average rate of membership in four latent groups in each year the state is present in the network.

and contract as states move in and out of geopolitical coalitions. Groups 1 and 2 are more prominent in earlier time periods, when the state system was composed of fewer states and great powers frequently resorted to war to settle political conflicts. In the post-World War II era, decolonization and the attendant increase in independent states generates increased membership in Group 6, a cluster that represents small states that tend to avoid military conflict. The most peaceful clusters, Group 3 and 6, increase in membership after World War II, when norms against military aggression began to consolidate.

Figure 3 displays latent group membership by year for a select group of states. There is significant variation across states and within some states over time. The United States and United Kingdom have high membership in Group 1, as discussed above. US membership in Group 1 is stable over the period of the study, while the UK consolidates its membership in this coalition after transitioning to a democratic political system. Russia's membership is overwhelmingly dominated by Group 2. At the end of the Cold War, the implosion of the Soviet system shifts Russia slightly to Group 1 before reverting fully back to Group 2.

Predictor	Dyadic	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
INTERCEPT		0.402	0.672	−4.758	−4.430	−3.462	−25.802
		(0.062)	(0.058)	(0.065)	(0.068)	(0.068)	(0.051)
POLITY		0.157	−0.054	0.136	−0.022	−0.087	−0.510
		(0.004)	(0.002)	(0.003)	(0.004)	(0.002)	(0.006)
MILITARY		0.641	0.487	−0.328	−0.147	−0.292	−3.314
CAPABILITY		(0.012)	(0.010)	(0.010)	(0.010)	(0.008)	(0.030)
BORDERS	2.500						
	(0.404)						
DISTANCE	0.000						
	(0.402)						
ALLIANCE	0.283						
	(0.406)						
IO CO-MEMBERS	0.000						
	(0.402)						
PEACE YRS	−0.149						
	(0.402)						

N nodes: 216; N dyads: $805,243$; N time periods: 192
Lower bound at convergence: $-155,058.1$

Table 1: **Estimated Coefficients and their Standard Errors**. The table shows the estimated coefficients (and standard errors) of the two monadic predictors for each of six latent groups, as well as those of the dyadic predictors for edge formation. We present the results from the first Markov state, which accounts for almost the entire time period. The estimated coefficients for cubic splines and indicators for variable missingness are not shown.

Japan, Iran, and Iraq further demonstrate how political shocks like revolution and foreign intervention are reflected in the evolution of latent membership. Japan experiences a sharp shift from Group 2 to Group 1 upon its loss in World War II and subsequent occupation by American forces. Iranian membership transitions quickly from Group 5 to Group 2 around the time of the Iranian Revolution in 1979, consistent with the anti-American posture of the Khomeini government. In Iraq, the sudden shift to Group 1 in 2003 reflects the invasion by the US and allied countries and the installation of a friendly government.

4.4 Covariate Effects

The dynMMSBM also enables the examination of covariate relations that can help characterize the nature of each estimated latent group. The upper panel of Table 1 displays coefficient estimates for the monadic covariates `POLITY` and `MILITARY CAPABILITY`. The estimates represent the effect of each covariate on the log-odds of membership in each latent group. We display the coefficients only for Markov state 1, since almost the entire time period under study, i.e., 95.4%, is estimated to derive from this state.

Democratic regimes (i.e., those with high POLITY scores) are most likely to instantiate membership in Groups 1 and 3. This is consistent with the interpretation of Group 1 as the Western alliance of liberal democracies during the Cold War, and Group 3 as Western-leaning neutral states. Autocratic regimes sort into Group 6 at the highest rate. Greater military capability is positively associated with membership in Groups 1 and 2 and negatively associated with membership in the other clusters.

In addition to obtaining estimates for the coefficients in our model, we can also predict how the probability of edge formation changes as a function of the node's monadic covariates. In the generative process of the model, group memberships are instantiated for each dyad in each time period. As a result, states in the conflict network are assigned to a latent group each time they interact with another state in a given year. Because the probability of edge formation depends on the group membership of both nodes in a dyad, a change in one node's monadic predictor will yield heterogeneous effects across dyads, nodes, and time.

For example, consider the change in predicted conflict propensity when each node's POLITY score is increased by one standard deviation (6.78). During this exercise, we allow POLITY scores to increase only up to the maximum value (10) so that a POLITY score never exceeds this value. The overall average effect of this change on the probability of edge formation, averaging all dyadic interactions and time periods, is positive but negligible in size: 0.0001. Thus, increasing the degree of democracy in a country results in a minor increase in conflict, given the underlying geopolitical coalitions throughout the time period.

There is, however, a significant amount of heterogeneity in this effect across states and over time. Figure 4 shows, for each state, the difference in expected probability of interstate conflict due to an increase of one standard deviation in POLITY score. Some states (such as India, Russia, and the United States) are predicted to be more peaceful, on average, if they become more democratic. Many others, however, are estimated to be more conflict-prone (e.g., Syria, Iraq, Afghanistan, Cuba, and Saudi Arabia), An increase in polity shifts

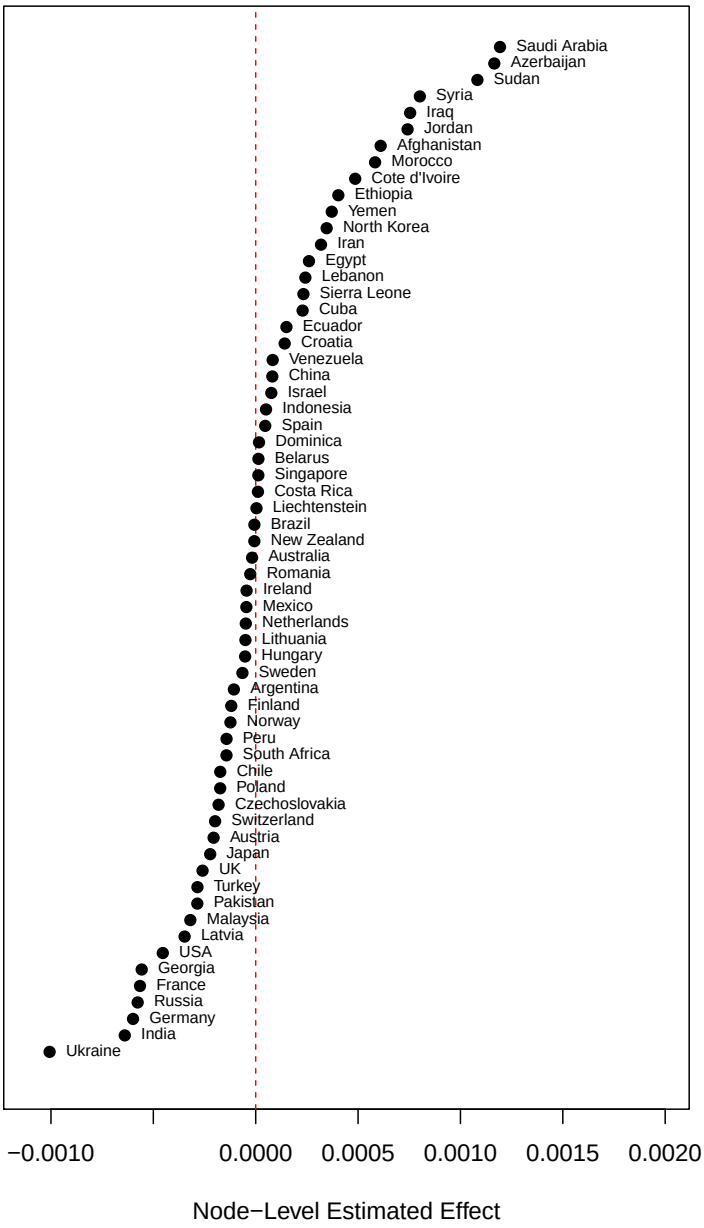

Figure 4: **Estimated Effects of Covariate Shift in Polity over Time, Select States**. The figure shows the estimated change in the probability of interstate conflict if a state's POLITY score is increased by one standard deviation (6.78) from its observed value.

these states into different latent groups that are more conflictual, on average.

The effect of democracy varies due to the latent group structure of the model. In general, shifts in monadic predictors will generate effects that are non-linear and contingent upon the existing group membership of the node in question and the other nodes in the network. Figure 5 looks within states to gauge the effect of the shift in POLITY over

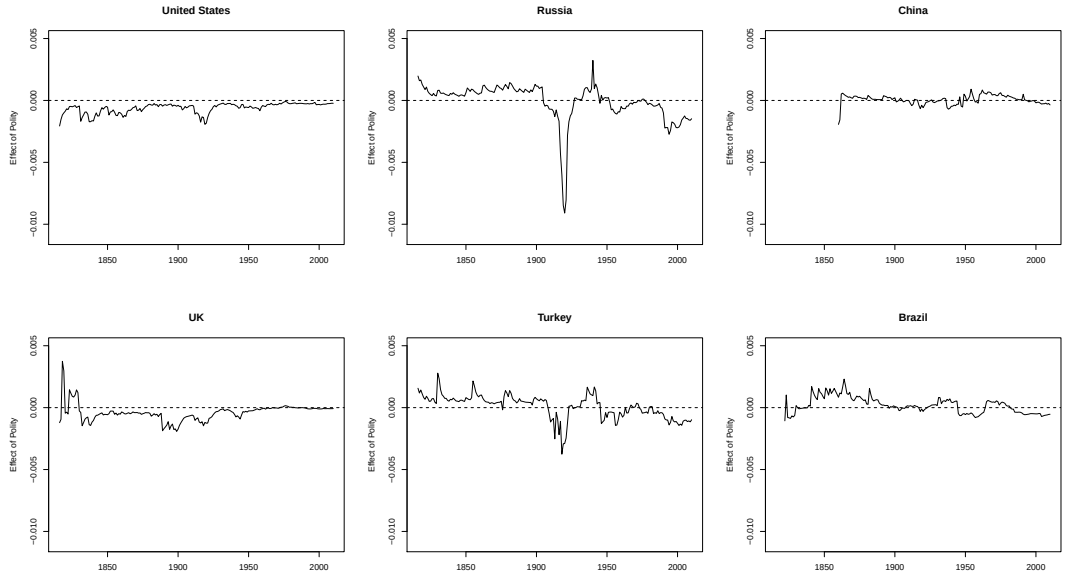

Figure 5: **Effect of Shift in Polity over Time, Select States**. The figure shows the estimated change in the probability of interstate conflict over time if a state's POLITY score is increased by one standard deviation (6.78) from its observed value.

Figure 6: **Estimated Aggregate Effect of Shift in Polity over Time**. The figure shows the estimated average change in the probability of interstate conflict when states' POLITY scores are increased by one standard deviation (6.78).

time, revealing additional heterogeneity. Finally, Figure 6 displays the average effect for each year in the time period. An increase in democracy induces more conflict, on average, throughout most of the sample. The effect is only consistently negative during first World

War and the interwar period, hitting its nadir in 1918 (-0.002). The impact of polity is attenuated in recent years, when the estimated effect of increasing polity approaches zero.

Finally, dyadic predictors operate outside the latent group membership structure, directly influencing the probability of conflict among states. In a sense, they serve as controls for alternative networks defined on the same node set. The dyadic coefficient estimates appear in the bottom panel of Table 1. Consistent with existing work, sharing a border significantly increases the likelihood of conflict. Greater geographic distance between states, formal alliances, and joint membership in international organizations have no statistically discernible effect on conflict propensity, once the blockmodel structure is accounted for.

4.5 Comparison with Logistic Regression

Finally, we compare the forecasting performance of the dynMMSBM to that of the standard logistic regression model prevalent in the democratic peace literature. We fit this regression model to the same interstate conflict data organized in the dyad-year format using the identical set of predictors. The only difference is that, in keeping with the convention in the literature, we transform the monadic variables (`POLITY` and `MILITARY CAPABILITY`) to a dyadic structure. We follow the conventional approach to specifying `POLITY` by including two separate variables measuring the democracy level of the less democratic country and that of the more democratic country in a dyad (e.g., Dafoe et al., 2013). The `MILITARY CAPABILITY` variable is restructured as the ratio of the more powerful state's military capability to the less powerful state's military capability.

We then conduct an out-of-sample validation exercise on the years 2009-2010, which were excluded from our initial sample. We follow Goldstone et al. (2010) in using a 2-year window for out-of-sample validation. We use the parameters of the dynMMSBM and logit models to predict the onset of conflict for dyad-years in the 2009-2010 period. Because the models include peace years and cubic splines as predictors, we impute these variables based on estimated probabilities of conflict in the out-of-sample set. To impute, we first forecast

conflict in the year 2009 and then sample from the predicted probabilities of conflict to update the peace years variable for each dyad. For the dynMMSBM, we let the network evolve according to the estimated Markov transition probabilities.

We evaluate the predictive accuracy of both models by comparing their predictions with the observed pattern of conflict in 2009-2010. Figure 12 in the Appendix A.4 displays the receiver operating curves (ROCs) for each model. The dynMMSBM slightly outperforms the logit model in out-of-sample prediction. The area under the ROC curve is 0.975 for the dynMMSBM and 0.963 for the logistic regression model.

5 Conclusion

We have introduced the dynMMSBM, a generalization of the mixed-membership stochastic blockmodel that incorporates dyadic and nodal attributes, and accounts for episodic temporal evolution of networks using a hidden-Markov process. The proposed model enables researchers to evaluate dynamic theories about the role of individual characteristics on the generation of relational outcomes when abstract groups of actors are the driving force behind tie formations. The dynMMSBM also helps identify periods in time when a network exhibits distinctive patterns of interactions among actors.

Using a network defined by almost 200 years of militarized interstate disputes in the international system, our model uncovers previously understudied spatial and temporal heterogeneity in the so called "democratic peace," whereby regime type is expected to affect the likelihood that any two countries engage in militarized actions against each other. Our model also uncovers the evolving nature of unobserved geopolitical coalitions, with memberships that conform to theoretical expectations — with liberal democracies aligned in one bloc, and more authoritarian regimes aligned in another.

This paper provides applied researchers with a model that can accommodate a variety of theorized relationships for dynamic network outcomes that display some form of stochastic equivalence. We make available the open-source R software package NetMix

that implements the proposed methodology. In the future, we plan to further extend the model's applicability to a variety of outcome variable types. Similarly, and given their prevalence in social scientific research, we plan to extend the model to accommodate bipartite or affiliation networks.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008), "Mixed membership stochastic blockmodels," *Journal of Machine Learning Research*, 9, 1981–2014.

Asuncion, A., Welling, M., Smyth, P., and Teh, Y. W. (2009), "On smoothing and inference for topic models," in *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, AUAI Press, pp. 27–34.

Barbieri, K. (1996), "Economic interdependence: A Path to Peace or a Source of Interstate Conflict?" *Journal of Peace Research*, 33, 29–49.

Beck, N., Katz, J. N., and Tucker, R. (1998), "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable," *American Journal of Political Science*, 42, 1260–1288.

Beck, N., King, G., and Zeng, L. (2000), "Improving Quantitative Studies of International Conflict: A Conjecture," *American Political Science Review*, 94, 21–35.

Bueno de Mesquita, B., Morrow, J. D., Siverson, R. M., and Smith, A. (2004), "Testing Novel Implications from the Selectorate Theory of War," *World Politics*, 56, 363–388.

Cranmer, S. J. and Desmarais, B. A. (2011), "Inferential Network Analysis with Exponential Random Graph Models," *Political analysis*, 19, 66–86.

Dafoe, A., Oneal, J. R., and Russett, B. (2013), "The Democratic Peace: Weighing the Evidence and Cautious Inference," *International Studies Quarterly*, 57, 201–214.

Fan, X., Cao, L., and Da Xu, R. Y. (2015), "Dynamic infinite mixed-membership stochastic blockmodel," *IEEE transactions on neural networks and learning systems*, 26, 2072–2085.

Farber, H. S. and Gowa, J. (1997), "Common interests or common polities? Reinterpreting the democratic peace," *The Journal of Politics*, 59, 393–417.

Fearon, J. D. (1994), "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review*, 88, 577–592.

Gartzke, E. (2007), "The capitalist peace," *American journal of political science*, 51, 166–191.

Ghosn, F. and Bennett, S. (2003), "Codebook for the dyadic militarized interstate incident data, version 3.0," *Online: http://correlatesofwar.org*, 24.

Gibler, D. M. (2009), *International Military Alliances, 1648-2008*, CQ Press.

Gleditsch, N. P. and Hegre, H. (1997), "Peace and democracy: Three levels of analysis," *Journal of Conflict Resolution*, 41, 283–310.

Goldsmith, B. E. (2007), "A liberal peace in Asia?" *Journal of Peace Research*, 44, 5–27.

Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., Ulfelder, J., and Woodward, M. (2010), "A Global Model for Forecasting Political Instability," *American Journal of Political Science*, 54, 190–208.

Gowa, J. (2011), *Ballots and bullets: The elusive democratic peace*, Princeton University Press.

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007), "Model-based clustering for social networks," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170, 301–354.

Hegre, H. (2008), "Gravitating Toward War: Preponderance May Pacify, but Power Kills," *Journal of Conflict Resolution*, 52, 566–589.

Hegre, H., Metternich, N. W., Nygaard, H. M., and Wucherpfennig, J. (2017), "Introduction: Forecasting in Peace Research," *Journal of Peace Research*, 54, 5–18.

Ho, Q. and Xing, E. P. (2015), *Handbook of Mixed Membership Models and Their Applications*, CRC Press, chap. Analyzing time-evolving networks using an evolving cluster mixed membership blockmodel, pp. 489–525.

Hoff, P. D. (2009), "Multiplicative latent factor models for description and prediction of social networks," *Computational and mathematical organization theory*, 15, 261.

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, 1090–1098.

Hoff, P. D. and Ward, M. D. (2004), "Modeling Dependencies in International Relations Networks," *Political Analysis*, 12, 160–175.

Jin, J., Ke, Z. T., and Luo, S. (2018), "SCORE+ for Network Community Detection," *CoRR*, abs/1811.05927.

Jones, D. M., Bremer, S. A., and Singer, J. D. (1996), "Militarized interstate disputes, 1816–1992: Rationale, coding rules, and empirical patterns," *Conflict Management and Peace Science*, 15, 163–213.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999), "An introduction to variational methods for graphical models," *Machine learning*, 37, 183–233.

Kim, M. and Leskovec, J. (2013), "Nonparametric multi-group membership model for dynamic networks," in *Advances in neural information processing systems*, pp. 1385–1393.

Latouche, P., Birmelé, E., Ambroise, C., et al. (2011), "Overlapping stochastic block models with application to the french political blogosphere," *The Annals of Applied Statistics*, 5, 309–336.

Lorrain, F. and White, H. C. (1971), "Structural equivalence of individuals in social networks," *The Journal of mathematical sociology*, 1, 49–80.

Lyzinski, V., Fishkind, D. E., and Priebe, C. E. (2014), "Seeded graph matching for correlated Erdös-Rényi graphs." *Journal of Machine Learning Research*, 15, 3513–3540.

Mansfield, E. D. and Snyder, J. (2002), "Incomplete democratization and the outbreak of military disputes," *International Studies Quarterly*, 46, 529–549.

Maoz, Z., Kuperman, R. D., Terris, L., and Talmud, I. (2006), "Structural equivalence and international conflict: A social networks analysis," *Journal of Conflict Resolution*, 50, 664–689.

Maoz, Z. and Russett, B. (1993), "Normative and structural causes of democratic peace, 1946–1986," *American Political Science Review*, 87, 624–638.

Marshall, M., Gurr, T. R., and Jaggers, K. (2017), "Polity IV Project, Political Regime Characteristics and Transitions, 1800-2016." *Polity IV Project-Dataset Users' Manual*.

Matias, C. and Miele, V. (2017), "Statistical clustering of temporal networks through a dynamic stochastic block model," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 1119–1141.

Oneal, J. R. and Russett, B. (1999), "The Kantian peace: The pacific benefits of democracy, interdependence, and international organizations, 1885–1992," *World politics*, 52, 1–37.

Oneal, J. R. and Tir, J. (2006), "Does the Diversionary Use of Force Threaten the Democratic Peace? Assessing the Effect of Economic Growth on Interstate Conflict, 1921–2001," *International Studies Quarterly*, 50, 755–779.

Peceny, M., Beer, C. C., and Sanchez-Terry, S. (2002), "Dictatorial Peace?" *American Political Science Review*, 96, 15–26.

Salter-Townshend, M. and Brendan Murphy, T. (2015), "Role analysis in networks using mixtures of exponential random graph models," *Journal of Computational and Graphical Statistics*, 24, 520–538.

Schrodt, P. A. (1991), "Prediction of Interstate Conflict Outcomes Using a Neural Network," *Social Science Computer Review*, 9, 359–380.

Singer, J. D., Bremer, S., and Stuckey, J. (1972), "Capability Distribution, Uncertainty, and Major Power War, 1820-1965," *Peace, war, and numbers*, 19, 48.

Stinnett, D. M., Tir, J., Diehl, P. F., Schafer, P., and Gochman, C. (2002), "The Correlates of War (COW) Project Direct Contiguity Data, version 3.0," *Conflict Management and Peace Science*, 19, 59–67.

Sweet, T., Thomas, A., and Junker, B. (2014), *Handbook of mixed membership models and their applications*, CRC Press, chap. Hierarchical mixed membership stochastic blockmodels for nultiple networks and experimental interventions, pp. 463–488.

Teh, Y. W., Newman, D., and Welling, M. (2007), "A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation," in *Advances in Neural Information Processing Systems*, pp. 1353–1360.

Wang, Y. J. and Wong, G. Y. (1987), "Stochastic blockmodels for directed graphs," *Journal of the American Statistical Association*, 82, 8–19.

Ward, M. D., Metternich, N. W., Dorff, C. L., Gallop, M., Hollenbach, F. M., Schultz, A., and Weschle, S. (2013), "Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction," *International Studies Review*, 15, 473–490.

Ward, M. D., Siverson, R. M., and Cao, X. (2007), "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace," *American Journal of Political Science*, 51, 583–601.

Wasserman, S. and Faust, K. (1994), *Social network analysis: Methods and applications*, vol. 8, Cambridge university press.

White, A. and Murphy, T. B. (2016), "Mixed-membership of experts stochastic blockmodel," *Network Science*, 4, 48–80.

Xing, E. P., Fu, W., and Song, L. (2010), "A state-space mixed membership blockmodel for dynamic network tomography," *The Annals of Applied Statistics*, 4, 535–566.

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. (2019), "Statistical inference in a directed network model with covariates," *Journal of the American Statistical Association*, 114, 857–868.

A Appendix

A.1 Marginalizing the membership vectors and the transition probabilities

In this appendix, we show how to marginalize $\mathbf{\Pi}$.

$$
\begin{aligned}
&\int \cdots \int \prod_{t=1}^{T} \prod_{p \in V_t} \left[\prod_{m=1}^{M} P(\boldsymbol{\pi}_{pt} \mid \boldsymbol{\alpha}_{ptm})^{s_{tm}} \right] \prod_{q \in V_t} P(\mathbf{z}_{p \to q,t} \mid \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{p \leftarrow q,t} | \boldsymbol{\pi}_{pt}) \, \mathrm{d}\boldsymbol{\pi}_1 \ldots \mathrm{d}\boldsymbol{\pi}_{N_t} \\
&= \prod_{t=1}^{T} \prod_{p \in V_t} \int \prod_{m=1}^{M} [P(\boldsymbol{\pi}_{pt} \mid \boldsymbol{\alpha}_{ptm})]^{s_{tm}} \prod_{q \in V_t} P(\mathbf{z}_{p \to q,t} \mid \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{p \leftarrow q,t} | \boldsymbol{\pi}_{pt}) \, \mathrm{d}\boldsymbol{\pi}_{pt} \\
&= \prod_{t=1}^{T} \prod_{p \in V_t} \int \prod_{m=1}^{M} \left[\frac{\Gamma(\xi_{ptm})}{\prod_{k=1}^{K} \Gamma(\alpha_{ptmk})} \prod_{k=1}^{K} \pi_{ptk}^{\alpha_{ptmk}-1} \right]^{s_{tm}} \prod_{q \in V_t} \prod_{k=1}^{K} \pi_{ptk}^{z_{p \to q,t,k}} \pi_{ptk}^{w_{p \leftarrow q,t,k}} \, \mathrm{d}\boldsymbol{\pi}_{pt} \\
&= \prod_{t=1}^{T} \prod_{p \in V_t} \prod_{m=1}^{M} \left[\frac{\Gamma(\xi_{ptm})}{\prod_{k=1}^{K} \Gamma(\alpha_{ptmk})} \right]^{s_{tm}} \\
&\quad \times \int \prod_{k=1}^{K} \pi_{ptk}^{\sum_{m=1}^{M} s_{tm} \alpha_{ptmk} - 1} \prod_{q \in V_t} \prod_{k=1}^{K} \pi_{ptk}^{z_{p \to q,t,k}} \pi_{ptk}^{w_{p \leftarrow q,t,k}} \, \mathrm{d}\boldsymbol{\pi}_{pt}
\end{aligned}
$$

As they share a common base, we can simplify the products and define $C_{ptk} = \sum_{q \in V_t} (z_{p \to q,t,k} + w_{p \leftarrow q,t,k})$ to show that the above equation is equivalent to,

$$
\prod_{t=1}^{T} \prod_{p \in V_t} \prod_{m=1}^{M} \left[\frac{\Gamma(\xi_{ptm})}{\prod_{k=1}^{K} \Gamma(\alpha_{ptmk})} \right]^{s_{tm}} \int \prod_{k=1}^{K} \pi_{ptk}^{\sum_{m=1}^{M} s_{tm} \alpha_{ptmk} + C_{ptk} - 1} \, \mathrm{d}\boldsymbol{\pi}_{pt}
$$

The integrand can be recognized as the kernel of a Dirichlet distribution. As the integral is over the entire support of this Dirichlet, we can easily compute it as the inverse of the corresponding normalizing constant,

$$
\prod_{t=1}^{T} \prod_{p \in V_t} \prod_{m=1}^{M} \left[\frac{\Gamma(\xi_{ptm})}{\prod_{k=1}^{K} \Gamma(\alpha_{ptmk})} \right]^{s_{tm}} \frac{\prod_k \Gamma(\sum_{m=1}^{M} s_{tm} \alpha_{ptmk} + C_{ptk})}{\Gamma(\sum_{m=1}^{M} s_{tm} \xi_{ptm} + 2N_t))}
$$

where the sum of C_{ptk} over groups k is equal to twice the number of nodes (as nodes must instantiate at least one group in each of interactions, once as a sender and once again as a receiver) in directed networks. A simple reorganization of factors (along with the fact that $s_{t,m}$ is an indicator vector, whereby $\sum_m s_{tm} x = \prod_m x^{s_{tm}}$) yields equation (2) in Section 3.2.

A.2 Details of the Variational EM Algorithm

A.2.1 E-step

E-step 1: Z and W

To obtain the updates of the $\boldsymbol{\phi}_{p\to q,t}$ variational parameters, we begin by restricting equation (3) to the terms that depend only on $\mathbf{z}_{p\to q,t}$ (for specific p and q nodes in V_t) and taking the logarithm of the resulting expression,

$$\begin{aligned}
&\log P(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{S}, \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{X}, \mathbf{D}) \\
&= z_{p\to q,t,k} \sum_{g=1}^{K} w_{q\leftarrow p,t,g} \left\{ Y_{pqt} \log(\theta_{pqtkh}) + (1 - Y_{pqt}) \log(1 - \theta_{pqtkh}) \right\} \\
&\quad + \sum_{m=1}^{M} s_{tm} \log \Gamma(\alpha_{ptmk} + C_{ptg}) + \text{const.}
\end{aligned}$$

Now, note that $C_{ptk} = C'_{ptk} + z_{p\to q,t,g}$ and that, for $x \in \{0, 1\}$, $\Gamma(y + x) = y^x \Gamma(y)$. Since the $z_{p\to q,t,k} \in \{0, 1\}$, we can re-express $\log \Gamma(\alpha_{ptmk} + C_{ptk}) = z_{p\to q,t,k} \log(\alpha_{ptmk} + C'_{ptk}) + \log \Gamma(\alpha_{ptmk} + C'_{ptk})$ and thus simplify the expression to,

$$\begin{aligned}
&z_{p\to q,t,k} \sum_{g=1}^{K} w_{q\leftarrow p,t,g} \left\{ Y_{pqt} \log(\theta_{pqtkg}) + (1 - Y_{pqt}) \log(1 - \theta_{pqtkg}) \right\} \\
&\quad + z_{p\to q,t,k} \sum_{m=1}^{M} s_{tm} \log \left(\alpha_{ptmk} + C'_{ptk} \right) + \text{const.}
\end{aligned}$$

We proceed by taking the expectation under the variational distribution $\widetilde{Q}$:

$$\begin{aligned}
&\mathbb{E}_{\widetilde{Q}} \{ \log P(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{s}, \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{D}, \mathbf{X}) \} \\
&= z_{p\to q,t,g} \sum_{g=1}^{K} \mathbb{E}_{\widetilde{Q}_2}(w_{q\leftarrow p,t,g}) \big(Y_{pqt} \log(\theta_{pqtkg}) + (1 - Y_{pqt}) \log(1 - \theta_{pqtkg}) \big) \\
&\quad + z_{p\to q,t,g} \sum_{m=1}^{M} \mathbb{E}_{\widetilde{Q}_1}(s_{tm}) \; \mathbb{E}_{\widetilde{Q}_2} \left\{ \log \left(\alpha_{ptmk} + C'_{ptk} \right) \right\} + \text{const.}
\end{aligned}$$

The exponential of this expression corresponds to the (unnormalized) parameter vector of a multinomial distribution $\widetilde{Q}_2(\mathbf{z}_{p\to q,t} \mid \boldsymbol{\phi}_{p\to q,t})$. The update for $\mathbf{w}_{q\leftarrow p,t}$ is similarly derived.

E-step 2: S

Isolating terms in Equation 3 that are not constant with respect to s_{tm} for a specific $t \neq 1$ and m, and rolling all other terms into a const., we have

$$P(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{s}, \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{D}, \mathbf{X}) = \Gamma(M\eta + U_m)^{-1} \prod_{m=1}^{M} \prod_{n=1}^{M} \Gamma(\eta + U_{mn}) \prod_{p \in V_t} \left[\prod_{k=1}^{K} \frac{\Gamma(\alpha_{ptmk} + C_{ptk})}{\Gamma(\alpha_{ptmk})} \right]^{s_{tm}} + \text{const.}$$

To isolate terms that depend on s_{tm} for specific $t > 1$, m and $n \neq m$, define the following useful quantities:

$$\begin{aligned} U'_m &= U_m - s_{tm} \\ U'_{mm} &= U_{mm} - s_{t-1,m}s_{tm} - s_{tm}s_{t+1,m} \\ U'_{nm} &= U_{nm} - s_{t-1,m}s_{tm} \\ U'_{mn} &= U_{mn} - s_{tm}s_{t+1,n} \end{aligned}$$

Focusing on the terms involving U_m and U_{mn}, and working on a typical case in which $1 < t < T$, we can isolate parts that do not depend on s_{tm} by again recalling that, for $x \in \{0, 1\}$, $\Gamma(y + x) = y^x \Gamma(y)$:

$$\begin{aligned} &\Gamma\left(M\eta + s_{tm} + U'_m\right)^{-1} \Gamma(\eta + s_{t+1,m}s_{tm} + s_{t-1,m}s_{tm} + U'_{mm}) \\ &\quad \times \prod_{n \neq m}^{M} \Gamma(\eta + s_{t+1,n}s_{tm} + U'_{mn})\Gamma(\eta + s_{tm}s_{t-1,n} + U'_{nm}) \\ &= (M\eta + U'_m)^{-s_{tm}} \Gamma(M\eta + U'_m)^{-1} \left\{ (\eta + U'_{mm} + 1)^{s_{t+1,m}s_{t-1,m}} (\eta + U'_{mm})^{s_{t-1,m} - s_{t-1,m}s_{t+1,m} + s_{t+1,m}} \right\}^{s_{tm}} \\ &\quad \times \Gamma(\eta + U'_{mm}) \prod_{n \neq m}^{M} (\eta + U'_{mn})^{s_{t+1,n}s_{tm}} \Gamma(\eta + U'_{mn}) \prod_{n \neq m}^{M} (\eta + U'_{nm})^{s_{tm}s_{t-1,n}} \Gamma(\eta + U'_{nm}) \end{aligned}$$

at which point all $\Gamma(\cdot)$ terms are constant with respect to s_{tm} and can be rolled into the normalizing constant so that

$$\begin{aligned} &P(\mathbf{Y}, \mathbf{Z}, \mathbf{S}, \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{D}, \mathbf{X}) \\ &= (M\eta + U'_m)^{-s_{tm}} \left\{ (\eta + U'_{mm} + 1)^{s_{t+1,m}s_{t-1,m}} (\eta + U'_{mm})^{s_{t-1,m} - s_{t-1,m}s_{t+1,m} + s_{t+1,m}} \right\}^{s_{tm}} \end{aligned}$$

$$\times \prod_{n \neq m}^{M} (\eta + U'_{mn})^{s_{t+1,n} s_{tm}} (\eta + U'_{nm})^{s_{tm} s_{t-1,n}}$$
$$\times \prod_{p \in V_t} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ptmk} + C_{ptk})}{\Gamma(\alpha_{ptmk})} \right]^{s_{tm}} + \text{const.}$$

Taking the logarithm and expectations under the variational distribution $\widetilde{Q}$ with respect to all variables other than s_{tm}, we have,

$$\begin{aligned}
\log \hat{\kappa}_{tm} = & -s_{tm}\ \mathbb{E}_{\tilde{Q}_1}[\log(M\eta + U'_m)] + s_{tm}\kappa_{t+1,m}\kappa_{t-1,m}\ \mathbb{E}_{\tilde{Q}_1}[\log(\eta + U'_{mm} + 1)] \\
& + s_{tm}(\kappa_{t-1,m} - \kappa_{t-1,m}\kappa_{t+1,m} + \kappa_{t+1,m})\ \mathbb{E}_{\tilde{Q}_1}[\log(\eta + U'_{mm})] \\
& + s_{tm} \sum_{n \neq m}^{M} \kappa_{t+1,n}\ \mathbb{E}_{\tilde{Q}_1}[\log(\eta + U'_{mn})] \\
& + s_{tm} \sum_{n \neq m}^{M} \kappa_{t-1,n}\ \mathbb{E}_{\tilde{Q}_1}[\log(\eta + U'_{nm})] + s_{tm} \sum_{p \in V_t} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \right] \\
& + s_{tm} \sum_{p \in V_t} \sum_{k=1}^{K} \mathbb{E}_{\tilde{Q}} \left[\log \left[\frac{\Gamma(\alpha_{ptmk} + C_{ptk})}{\Gamma(\alpha_{ptmk})} \right] \right] + \text{const.}
\end{aligned}$$

This corresponds to a multinomial distribution $\widetilde{Q}_1(\mathbf{s}_t|\boldsymbol{\kappa}_{tm})$, such that the mth element of its parameter vector is

$$\begin{aligned}
\hat{\kappa}_{tm}\ \propto\ & \exp\left[-\mathbb{E}_{\widetilde{Q}_1}[\log(M\eta + U'_m)] \right] \exp\left[\kappa_{t+1,m}\kappa_{t-1,m}\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{mm} + 1)] \right] \\
& \times \exp\left[(\kappa_{t-1,m} - \kappa_{t-1,m}\kappa_{t+1,m} + \kappa_{t+1,m})\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{mm})] \right] \\
& \times \prod_{n \neq m} \exp\left[\kappa_{t+1,n}\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{mn})] \right] \exp\left[\kappa_{t-1,n}\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{nm})] \right] \\
& \times \prod_{p \in V_t} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \prod_{k=1}^{K} \frac{\mathbb{E}_{\widetilde{Q}_1}[\Gamma(\alpha_{ptmk} + C_{ptk})]}{\Gamma(\alpha_{ptmk})} \right]
\end{aligned}$$

which must be normalized. When $t = T$, the term simplifies to

$$\begin{aligned}
\hat{\kappa}_{Tm} \propto & \exp\left[-\mathbb{E}_{\tilde{Q}_1}[\log(M\eta + U'_m)] \right] \prod_{n=1}^{M} \exp\left[\kappa_{T-1,m}\ \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{nm})] \right] \\
& \times \prod_{p \in V_T} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \prod_{k=1}^{K} \frac{\mathbb{E}_{\widetilde{Q}_1}[\Gamma(\alpha_{pTmk} + C_{pTk})]}{\Gamma(\alpha_{pTmk})} \right]
\end{aligned}$$

As before, the expectations can be approximated using a zero-order Taylor expansion.

A.2.2 M-step

Lower Bound

We first provide the expression for the lower bound,

$$
\begin{aligned}
\mathcal{L}(\tilde{Q}) \;=\; & \mathbb{E}_{\widetilde{Q}}[\log P(\mathbf{Y},\mathbf{Z},\mathbf{W},\mathbf{s},\mathbf{B},\boldsymbol{\beta}\mid\mathbf{X})] - \mathbb{E}_{\widetilde{Q}}[\log \widetilde{Q}(\mathbf{s},\mathbf{Z},\mathbf{W}\mid\mathbf{K},\boldsymbol{\Phi},\boldsymbol{\Psi})] \\
=\; & \log(P(s_1)) + \log\Gamma(M\eta) - \sum_{m=1}^{M}\mathbb{E}_{\widetilde{Q}}[\log\Gamma(M\eta+U_m)] \\
& + \sum_{m=1}^{M}\sum_{n=1}^{M}\mathbb{E}_{\widetilde{Q}}[\log\Gamma(\eta+U_{mn})] - \log\Gamma(\eta) \\
& + \sum_{t=1}^{T}\sum_{m=1}^{M}\kappa_{tm}\sum_{p\in V_t}\Big[\log\Gamma\left(\xi_{ptm}\right) - \log\Gamma\left(\xi_{ptm}+2N_t\right)\Big] \\
& + \sum_{t=1}^{T}\sum_{m=1}^{M}\kappa_{tm}\sum_{p\in V_t}\sum_{k=1}^{K}\Big[\mathbb{E}[\log\Gamma(\alpha_{ptmk}+C_{ptk})] - \log\Gamma(\alpha_{ptmk})\Big] \\
& + \sum_{t=1}^{T}\sum_{(p,q)\in E_t}\sum_{g,h=1}^{K}\phi_{p\to q,t,g}\psi_{q\leftarrow p,t,h}\left\{Y_{pqt}\log\theta_{pqtgh} + (1-Y_{pqt})\log(1-\theta_{pqtgh})\right\} \\
& - \sum_{g,h=1}^{K}\frac{(B_{gh}-\mu_{gh})^2}{2\sigma_{gh}^2} - \sum_{j=1}^{J_d}\frac{(\gamma_j-\mu_\gamma)^2}{2\sigma_\gamma^2} - \sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{j=1}^{J_x}\frac{(\beta_{mkj}-\mu_\beta)^2}{2\sigma_\beta^2} \\
& - \sum_{t=1}^{T}\sum_{m=1}^{M}\kappa_{tm}\log\kappa_{tm} - \sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{(p,q)\in E_t}\sum_{k=1}^{K}\{\phi_{p\to q,t,k}\log\phi_{p\to q,t,k} - \psi_{q\leftarrow p,t,h}\log(\psi_{q\leftarrow p,t,k})\}
\end{aligned}
$$

M-step 1: update for $\mathbf{B}$

Restricting the lower bound to terms that contain B_{gh}, we obtain

$$
\begin{aligned}
\mathcal{L}(\widetilde{Q}) \;=\; & \sum_{t=1}^{T}\sum_{p,q\in E_t}\sum_{g,h=1}^{K}\phi_{p\to q,t,g}\psi_{q\leftarrow p,t,h}\{Y_{pqt}\log\theta_{pqtgh} + (1-Y_{pqt})\log(1-\theta_{pqtgh})\} \\
& - \sum_{g,h=1}^{K}\frac{(B_{gh}-\mu_{gh})^2}{2\sigma_{gh}^2} + \text{const.}
\end{aligned}
$$

We optimize this lower bound with respect to $\mathbf{B}_{gh}$ using a gradient-based numerical optimization method. The corresponding gradient is given by,

$$
\frac{\partial\mathcal{L}_{B_{gh}}}{\partial B_{gh}} \;=\; \sum_{t=1}^{T}\sum_{p,q\in E_t}\phi_{p\to q,t,g}\psi_{q\leftarrow p,t,h}\left(Y_{pqt}-\theta_{pqtgh}\right) - \frac{B_{gh}-\mu_{B_{gh}}}{\sigma_{B_{gh}}^2}
$$

M-step 2: update for $\boldsymbol{\gamma}$

Restricting the lower bound to terms that contain $\boldsymbol{\gamma}$, and recalling that $\theta_{pqtgh} = [1 + \exp(-B_{gh} - \mathbf{d}_{pqt}\boldsymbol{\gamma})]^{-1}$, we have

$$\begin{aligned}\mathcal{L}(\tilde{Q}) \;=\; & \sum_{t=1}^{T} \sum_{p,q \in E_t} \sum_{g,h=1}^{K} \phi_{p \to q,t,g} \psi_{q \leftarrow p,t,h} \left\{ Y_{pqt} \log \theta_{pqtgh} + (1 - Y_{pqt}) \log(1 - \theta_{pqtgh}) \right\} \\ & - \sum_{j}^{J_d} \frac{(\gamma_j - \mu_\gamma)^2}{2\sigma_\gamma^2} + \text{const.}\end{aligned}$$

To optimize this expression with respect to γ_j (the jth element of the $\boldsymbol{\gamma}$ vector), we again use a numerical optimization algorithm based on the following gradient,

$$\frac{\partial \mathcal{L}(\widetilde{Q})}{\gamma_j} \;=\; \sum_{t=1}^{T} \sum_{p,q \in E_t} \sum_{g,h=1}^{K} \phi_{p \to q,t,g} \psi_{q \leftarrow p,t,h} d_{pqtj} \left(Y_{pqt} - \theta_{pqtgh} \right) - \frac{\gamma_j - \mu_\gamma}{\sigma_\gamma^2}$$

M-step 3: update for $\boldsymbol{\beta}_m$

Let $\alpha_{ptmk} = \exp\left(\mathbf{x}_{pt}^\top \boldsymbol{\beta}_{km}\right)$ and $\xi_{ptm} = \sum_{k=1}^{K} \alpha_{ptmk}$. To find the optimal value of $\boldsymbol{\beta}_{km}$, we roll all terms not involving the coefficient vector into a constant:

$$\begin{aligned}\mathcal{L}(\widetilde{Q}) \;=\; & \sum_{t=1}^{T} \sum_{m=1}^{M} \kappa_{tm} \sum_{p \in V_t} \left[\log \Gamma(\xi_{ptm}) - \log \Gamma(\xi_{ptm} + 2N_t) \right] \\ & + \sum_{t=1}^{T} \sum_{m=1}^{M} \kappa_{tm} \sum_{p \in V_t} \sum_{k=1}^{K} \Big[\mathbb{E}_{\widetilde{Q}_2} [\log \Gamma(\alpha_{ptmk} + C_{ptk})] - \log \Gamma(\alpha_{ptmk}) \Big] \\ & - \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{j=1}^{J_x} \frac{(\beta_{mkj} - \mu_\beta)^2}{2\sigma_\beta^2} + \text{const.}\end{aligned}$$

No closed form solution exists for an optimum w.r.t. β_{mkj}, but a gradient-based algorithm can be implemented to maximize the above expression. The corresponding gradient with respect to each element of $\boldsymbol{\beta}_{mk}$ is given by,

$$\begin{aligned}\frac{\partial \mathcal{L}(\tilde{Q})}{\partial \beta_{mkj}} \;=\; & \sum_{t=1}^{T} \kappa_{tm} \sum_{p \in V_t} \alpha_{ptmk} x_{ptj} \Big(\mathbb{E}_{\widetilde{Q}_2} [\breve{\psi}(\alpha_{ptmk} + C_{ptk}) - \breve{\psi}(\alpha_{ptmk})] \\ & + \Big[\breve{\psi}(\xi_{ptm}) - \breve{\psi}(\xi_{ptm} + 2N_t) \Big] \Big) \\ & - \frac{\beta_{mkj} - \mu_\beta}{\sigma_\beta^2}\end{aligned}$$

where $\breve{\psi}(\cdot)$ is the digamma function. Once again, we can approximate expectations of non-linear functions of random variables using a zeroth-order Taylor series expansion. As is the case of the multinomial logit model, we set $\boldsymbol{\beta}_{1,m} \equiv 0 \; \forall m$, making group 1 a reference for identification purposes.

A.3 A Simulation Study

Our synthetic networks are composed of 100 nodes observed over $t \in \{1, \ldots, 9\}$ time periods, and are constructed as follows:

1. For each node pt and dyad pqt at time $t > 1$, generate a single monadic and dyadic predictor using a random walk, so that $x_{pt} = x_{p,t-1} + \epsilon_{xt}$, $d_{pqt} = d_{pq,t-1} + \epsilon_{dt}$, with $x_{p1} \sim N(0, 2)$, $d_{pq,1} \sim N(0, 2)$, and $\epsilon_{xt} \sim N(0, 1)$, $\epsilon_{dt} \sim N(0, 1)$.

2. For each node at time t, sample a 2-dimensional mixed-membership vector from a 2-component mixture of Dirichlet distributions, so that

$$\boldsymbol{\pi}_{pt} \sim \prod_{m=1}^{2} \left[\text{Dirichlet}\left(\exp\left(\mathbf{x}_{pt}^{\top}\boldsymbol{\beta}_m\right)\right)\right]^{s_{tm}}$$

   where $\mathbf{x}_{pt} = [1 \; x_{pt}]^{\top}$, and s_{tm} indicates a state $m \in \{1, 2\}$ of the hidden Markov process, such that $s_{t1} = 1$ for $t \in \{1, \ldots, 5\}$, $s_{t2} = 1$ for $t \in \{6, \ldots, 9\}$, and $s_{tm} = 0$ otherwise (i.e. there is a changepoint in the underlying left-to-right HMM between time-points 5 and 6).

3. For each node pt and qt in directed dyad pqt, sample a pair of group memberships

$$z_{pt \to qt} \sim \text{Categorical}\left(\boldsymbol{\pi}_{pt}\right) \text{ and } w_{qt \leftarrow pt} \sim \text{Categorical}(\boldsymbol{\pi}_{qt})$$

4. Finally, and for the same dyad, sample an edge

$$Y_{pqt} \sim \text{Bernoulli}\left(\text{logit}^{-1}\left(B_{z_{pt \to qt}, w_{qt \leftarrow pt}} + \gamma_1 d_{pqt}\right)\right)$$

   where $\gamma_1 = 0.25$.

To explore the conditions under which the model performs best, as well as those under which learning the model's various parameters can be particularly challenging, we refine this data-generating process by defining three sets of values for $\boldsymbol{B}$ and $\boldsymbol{\beta}$ designed to generate easy, realistic, and hard learning scenarios. They differ in the extent to which memberships are truly mixed (with more clearly defined memberships being easier to learn), and with respect to the extent to which the blockmodels generate distinct equivalence classes of nodes (with more clearly defined block structures being easier to learn). Accordingly, each scenario's DGP is completed using the parameters in presented in Table 2.

	Easy	**Realistic**	**Hard**
$g^{-1}(\mathbf{B}) =$	$\begin{bmatrix} 0.85 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$	$\begin{bmatrix} 0.65 & 0.35 \\ 0.20 & 0.75 \end{bmatrix}$	$\begin{bmatrix} 0.45 & 0.10 \\ 0.20 & 0.55 \end{bmatrix}$
$\boldsymbol{\beta}_1 =$	$\begin{bmatrix} -4.5 & -4.5 \\ 0.0 & 0.0 \end{bmatrix}$	$\begin{bmatrix} 0.05 & 0.75 \\ -0.75 & -1.0 \end{bmatrix}$	$\begin{bmatrix} 1.5 & 1.5 \\ -0.75 & -1.0 \end{bmatrix}$
$\boldsymbol{\beta}_2 =$	$\begin{bmatrix} -4.5 & -4.5 \\ 0.0 & 0.0 \end{bmatrix}$	$\begin{bmatrix} -0.05 & 0.55 \\ -0.75 & 0.75 \end{bmatrix}$	$\begin{bmatrix} 2.5 & 0.5 \\ -0.75 & 0.75 \end{bmatrix}$

Table 2: **Parameters in three different dynamic network DGPs**. The three columns correspond to three types of networks, varying in terms of inferential complexity. In turn, the rows contain the corresponding values of the blockmodel $\mathbf{B}$ and the regression coefficient vectors $\boldsymbol{\beta}$, one for each state of the HMM.

Generating a single, 9-period network under each of these scenarios results in the mixed memberships depicted in Figure 7, which shows the density of membership into the first of two groups across all nodes and time periods. While the 'easy' scenario has very clearly defined memberships of most nodes into one of the underlying groups, the 'hard' scenario has a substantial number of nodes whose membership is decidedly more mixed. The more 'realistic' scenario has a non-negligible number of nodes whose membership is mixed, and

a distinct group imbalance in favor of the second group.

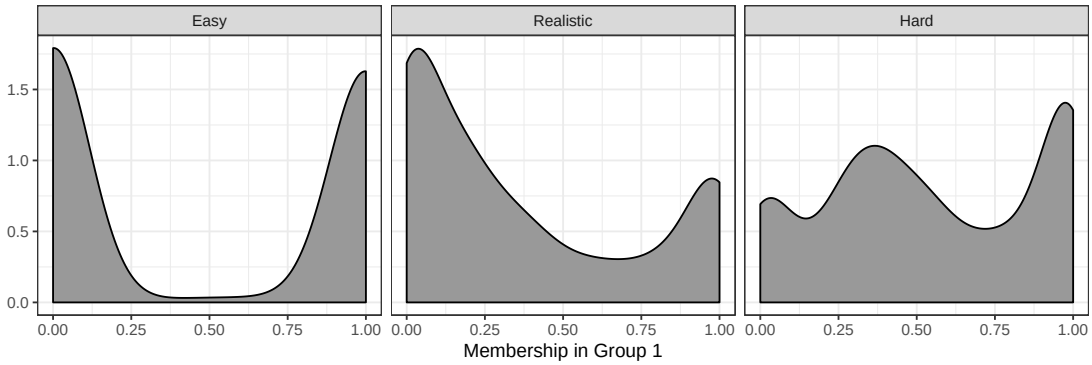

Figure 7: **Simulated mixed-memberships in synthetic networks**. The plots depict the mixed-membership vectors of nodes in three simulated networks, each with 100 nodes observed over 9 time periods. It shows the memberships of nodes in networks generated under an 'easy' DGP (i.e. one where memberships are not mixed, and in which the block structure is clear), 'hard' (i.e. one where memberships are extremely mixed, and no block structure is apparent in the network) and 'realistic' (i.e. where some nodes display a mixture of group memberships, and a block structure is somewhat apparent in the network) on the left, right, and central panels, respectively.

A.3.1 Accuracy of estimation: mixed-memberships and blockmodels

Overall, and as expected, the accuracy with which dynMMSBM can retrieve the true mixed-membership vectors depends on the problem's complexity. The top panel of Figure 8 shows the estimated mixed-membership values against their known, true values, evidencing a decrease in estimation accuracy as we move from an easy to a hard inferential task. Despite the clear deterioration, dynMMSBM is still able to produce good quality estimates even under hard inferential situations, with estimates that have a 0.82 correlation with their true values.

The model is also able to accurately estimate the blockmodel structure, as the bottom row of Figure 8 reveals. For each cell of the blockmodels, the true probability of an edge between members of any two groups is shown in white letters, while the cell itself is colored in accordance to the corresponding estimated values. Once again, and although the quality of these estimates (predictably) decreases as the inferential complexity of the scenario increases, the estimation error remains low.

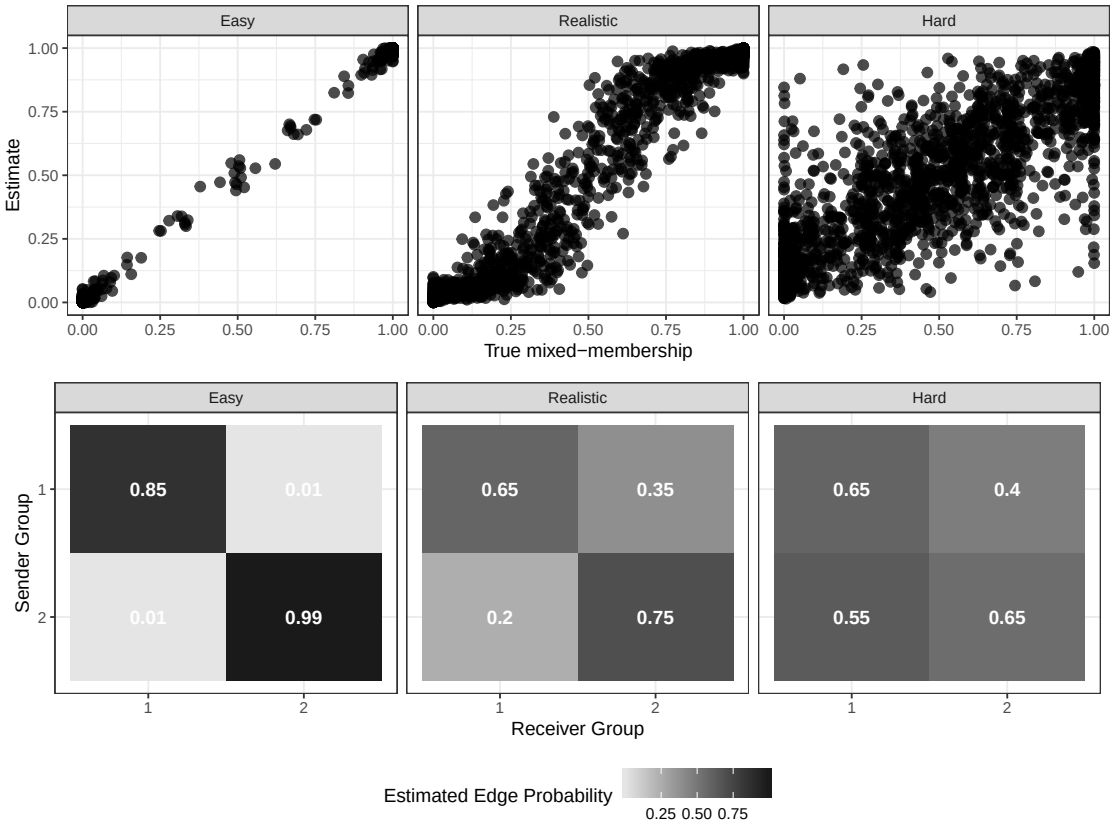

Figure 8: **Estimation accuracy**. For each DGP scenario, the figure shows the estimated mixed-membership vectors (top row) and the estimated blockmodels (bottom row) against their known values (indicated by the white numbers in each cell of the blockmodel for the bottom row). Overall, accuracy of retrieval both sets of parameters depends on the complexity of the learning problem, although recovery is generally very good, even under 'hard' inferential conditions.

A.3.2 Estimation accuracy: regression coefficients

The two most distinctive features of the proposed model are its ability to incorporate predictors of the mixed-membership vectors and to account for network dynamics. We evaluate the accuracy with which our proposed estimation strategy recovers known parameter values. To do so, we simulate 100 replicates of the 9-period network described above, generated under our more 'realistic' DGP and holding all design matrices constant across replicates. After generating all 50 networks, we use our model to obtain estimates of the effect of the monadic predictor on block memberships, as well as of the marginal probability that the hidden Markov process is in either of the two states for each time period.

Figure 9 shows, for each time period, the distribution of estimated effect sizes of the

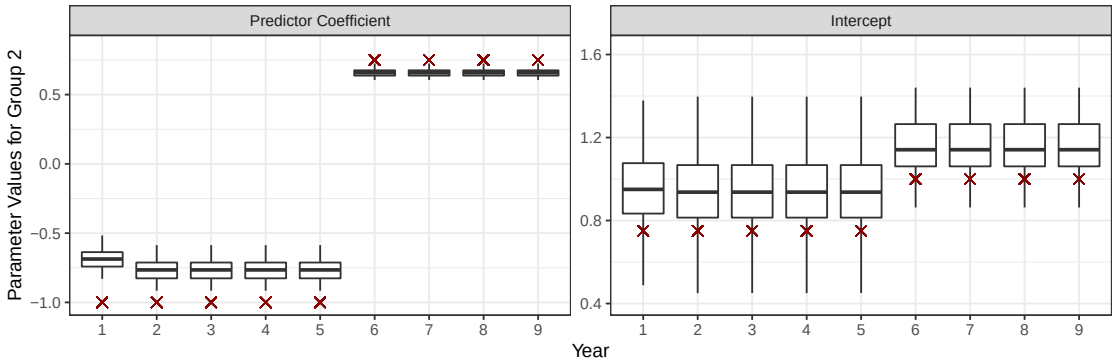

Figure 9: **Estimated parameters of block membership regression**. The figure shows, for each time period, the HMM-weighted effect of a continuous predictor on the probability of instantiating latent group 2 (left panel), and the HMM-weighted intercept of the corresponding regression line (right panel), estimated on 100 networks generated according to our realistic DGP. In each instance, the red "x" indicates the true parameter value for that time period, given a known HMM state.

monadic predictor and intercepts for the regression of membership into the second latent group (as boxplots), along with the true parameter values (shown as a red "x"). We obtain estimates for each time period by computing the weighted average of estimated parameters in the two hidden Markov states, using estimated marginal probabilities over states in each time period as our weights. The model is typically able to identify the underlying Markov state that generated the networks, which in turn translates into correctly estimated (albeit regularized) effects of the monadic covariate on membership probabilities. Quality of recovery for regression parameters associated with a given block depends heavily on the extent to which that block is commonly instantiated in the network. And although changes in intercepts across time periods are also correctly recovered, the intercepts themselves tend to be overestimated. This phenomenon, which we found to be common in all our simulations, is likely the result of the difficulty in pinning down the precision of the latent membership vectors. Despite these issues, the mean of the memberships is correctly recovered (as shown earlier in Figure 8).

A.3.3 Comparison to alternative modeling approaches

Finally, and to further evaluate the benefits of modeling the dynamic nature of the network, we estimate a separate MMSBM model to the networks in each time period, and compare

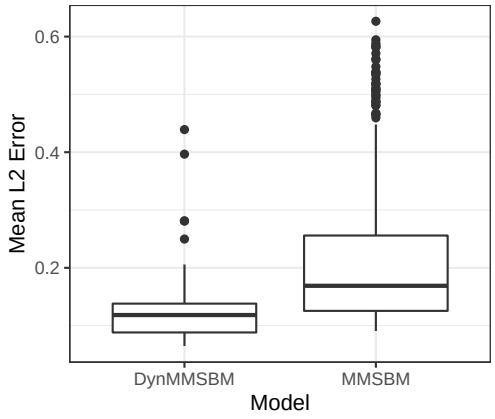

Figure 10: **Error for estimated mixed membership vectors**. The figure shows average L_2 distances between estimated and true mixed-membership vectors for all nodes in each of 50 replicated dynamic networks. On the left, estimates are generated using dynMMSBM. On the right, estimates are generated using the canonical MMSBM, fit separately to the nine time periods in each simulated network.

their estimated mixed-memberships to those of a single dynMMSBM estimated on the full set of networks. In both cases, we omit all covariates, but estimate the $\boldsymbol{\alpha}_{ptm}$ parameters associated with the mixed-membership vectors. After estimating both sets of models on each of the 50 replications of the "realistic" networks, we compute the average L_2 error in estimated mixed-memberships across nodes. The results are presented in Figure 10.

In general, dynMMSBM performs consistently better than the MMSBM estimated on each time period, and the latter shows much more variability in terms of accuracy. A major challenge for the per-year approach consists of realigning the estimated group labels, which (under the assumptions of our model) should be done by realigning the cells of the blockmodel, as all other parameters (such as the mixed-memberships themselves) are subject to change overtime. Being estimated using just a fraction of the data, however, the blockmodels obtained in the per-year approach prove too noisy to be useful in the realignment exercise, thus contributing to the variable accuracy of the non-dynamic approach. In contrast, dynMMSBM is able to recover the underlying blockmodel much more accurately, thus contributing to the correct estimation of the latent memberships across simulations.

A.4 Additional Empirical Results

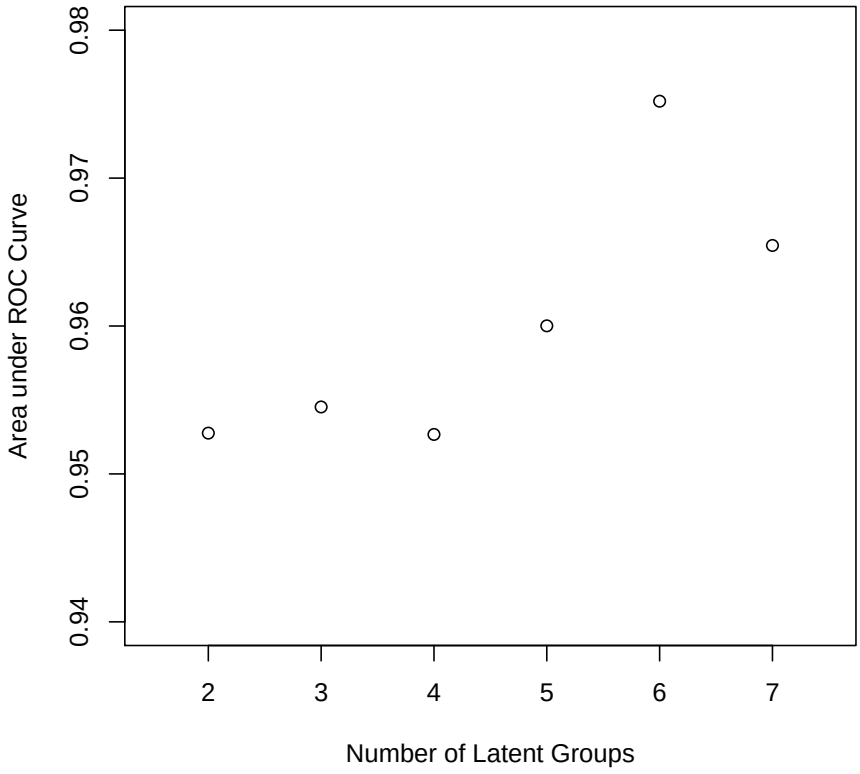

Figure 11: **Out of Sample Prediction, Conflict Models**. The figure plots the area under the ROC curve for models with 2-7 Latent Groups. Each model is fit on data from 1816-2008 and used to predict conflict in the period 2009-2010.

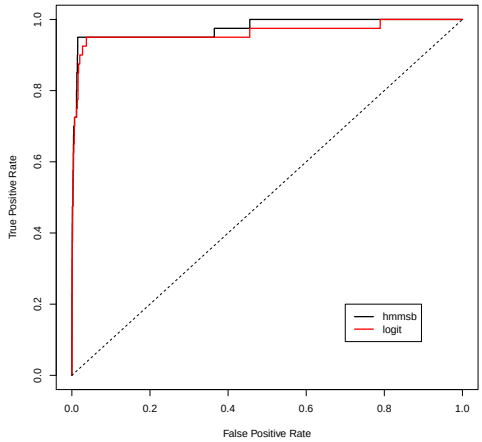

Figure 12: **ROC Curve: Logit, Dynamic Mixed-membership SBM Models**. To perform the forecast, we exclude the final two years (2009-2010) from the dataset and estimate each model on the preceding years (1816-2008). Then we predict the missing years based solely on the covariate data.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Group 1	0.027	0.062	0.012	0.063	0.022	0.020
Group 2	0.062	0.045	0.033	0.031	0.027	0.017
Group 3	0.012	0.033	0.005	0.040	0.023	0.004
Group 4	0.063	0.031	0.040	0.021	0.036	0.018
Group 5	0.022	0.027	0.023	0.036	0.017	0.011
Group 6	0.020	0.017	0.004	0.018	0.011	0.005

Table 3: **Group-Level Edge Formation Probabilities**. The table displays the probability of interstate conflict between nodes that instantiate membership in each of six latent groups. The diagonal shows rates of intra-group conflict and off-diagonal shows rates of conflict between groups.

Group 1	**Group 2**	**Group 3**
0.999 Germany	0.999 China	0.999 Trinidad-Tobago
0.998 Canada	0.999 Indonesia	0.999 New Zealand
0.998 Australia	0.999 Russia	0.999 Ireland
0.998 Netherlands	0.999 Poland	0.999 Norway
0.998 Italy	0.999 South Africa	0.999 Finland
0.998 West Germany	0.996 Mexico	0.999 Denmark
0.998 Belgium	0.974 Czechoslovakia	0.999 Switzerland
0.998 UK	0.968 Romania	0.999 Luxembourg
0.998 Japan	0.952 Korea South	0.999 Austria
0.997 India	0.945 East Germany	0.998 Jamaica
0.994 USA	0.854 Brazil	0.994 Costa Rica
0.731 France	0.725 Egypt	0.816 Sweden
0.706 Pakistan	0.682 Spain	0.775 Mauritius
0.537 Turkey	0.642 Nigeria	0.755 Israel
0.403 South Vietnam	0.541 Taiwan	0.663 Sri Lanka
Group 4	**Group 5**	**Group 6**
0.999 Brunei	0.999 Singapore	1 Liechtenstein
0.999 Bahamas	0.999 Yemen	1 Swaziland
0.769 Malta	0.999 Morocco	1 Equatorial Guinea
0.295 Iceland	0.997 Congo DR/Zaire	1 Bhutan
0.244 Afghanistan	0.997 Algeria	1 Djibouti
0.237 Cambodia	0.996 Ethiopia	1 Cape Verde
0.192 Zimbabwe	0.975 Bulgaria	1 Comoros
0.166 Barbados	0.958 Cuba	1 St Kitts-Nevis
0.088 Ghana	0.95 Hungary	1 Gabon
0.086 Tunisia	0.949 Chile	1 Solomon Is
0.067 Kuwait	0.815 Syria	1 Guinea-Bissau
0.035 Uganda	0.803 Iraq	1 Dominica
0.026 Laos	0.803 Zimbabwe	1 Antigua-Barbuda
0.026 Syria	0.731 Peru	1 Mauritania
0.024 Lebanon	0.707 Myanmar	1 Central African Rep

Table 4: **States with Highest Membership in Latent Groups, Cold War period**. To identify the states with highest membership in each latent group, we average over each states' latent membership probabilities in the years 1950-1990. Average group membership is reported beside the state name for the top 15 states in each latent group. The group assignments are consistent with known geopolitical coalitions in the Cold War, with Western allies in Group 1, Eastern bloc countries clustered in Group 2, Western-leaning neutral states in Group 3, and states engulfed in proxy conflicts in Group 4.