Dynamic Stochastic Blockmodel Regression for Network Data: Application to International Militarized Conflicts*

Santiago Olivella† Tyler Pratt‡ Kosuke Imai§

First Draft: July 12, 2018
This Draft: February 24, 2021

Abstract

A primary goal of social science research is to understand how latent group memberships predict the dynamic process of network evolution. In the modeling of international conflicts, for example, 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 network data 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. The proposed methodology, which relies on a variational approximation to a collapsed posterior distribution as well as stochastic optimization for scalability, is implemented through an open-source software package.

Keywords: hidden Markov model, mixed-membership stochastic blockmodel, social networks, stochastic optimization, variational approximation

*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 and political networks, and stipulate how the formation and evolution of these groups give rise to various 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. We extend the classical MMSBM in three ways. First, we allow memberships in latent groups to evolve over time according to a hidden Markov process. Second, we define a regression model for both latent memberships and observed ties, incorporating both dyadic and nodal attributes to explain the formation of groups and to relax the strict assumption of stochastic equivalence for members of the same groups. Finally, we define the model in a collapsed parameter space, using sufficient statistics to summarize large networks and improve computational scalability.

Our model, which we call dynMMSBM, therefore 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 ap-

plication 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), using stochastic gradient descent to accommodate very large-scale networks while retaining both theoretical properties of the approximation and practical run times (Hoffman et al., 2013; Gopalan and Blei, 2013).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 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. In turn, Hoff and Ward (2004) employ random effects to explicitly model network dependence in dyadic data, and Ward et al. (2007) apply the latent space model developed by Hoff et al. (2002) to international conflict. Similarly, 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 network data 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 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 assume independence of group memberships over time and across nodes, as well as 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 stochastic 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 the need to incorporate dyadic and nodal attributes as well as 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 Challenges In The Study of 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).

While empirical studies of interstate conflict are commonly conducted assuming conditional independence of dyad-year observations (e.g., Maoz and Russett, 1993; Farber and Gowa, 1997; Goldsmith, 2007; Gowa, 2011; Dafoe et al., 2013), 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, often unobserved groups.

Statistical models of network formation are a natural way to account for these cross-sectional and temporal dependencies. While recent applications use network models to re-examine the democratic peace debate (e.g., Hoff and Ward, 2004; Ward et al., 2007; Cranmer and Desmarais, 2011), existing methods are hindered by several challenges. First, they do not directly model the evolving geopolitical coalitions that shape patterns of conflict. Such a model would more closely reflect the theoretical mechanisms explaining 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).

A second limitation is the need to restructure monadic covariates like democracy to fit a dyadic level of analysis. This problem has exacerbated a debate in the democratic peace literature regarding the appropriate dyadic specification of democracy (see Dafoe et al., 2013). An ideal model would directly incorporate nodal variables at the country level by embedding them within the generative process of group formation. Finally, most existing methods do not provide flexibility for the effect of democracy to vary over time, despite theoretical claims that it should do so (Farber and Gowa, 1997; Cederman, 2001).

In the following section, we propose a model that overcomes these shortcomings. The dynMMSBM could uncover a democratic peace by identifying a latent group that exhibits low rates of intra-group conflict and that democratic states are more likely to join. Other hypotheses in this literature — for example, the possibility of a similar "dictatorial peace" among autocratic states (Peceny et al., 2002), interactions between democracy and power asymmetries (Bueno de Mesquita et al., 2004), and variation in the strength of the democratic peace over time (Gleditsch and Hegre, 1997; Cederman, 2001) — are also easily accommodated by the model structure. Each latent group is associated with its own set of nodal covariates, obviating the need to restructure monadic variables, and the dynamic implementation provides flexibility for covariate effects to vary over time.

3 The Proposed Methodology

Using the history of interstate conflict networks to study the democratic peace theory requires a model that not only defines the probability of conflict as a function of membership into latent groups of countries, but that also enables the exploration of how these memberships evolve over time and how they are informed by country-level characteristics — particularly regime type. Furthermore, for practical use, the model must deal with the

computational complexity involved in estimating a dynamic network model with a large number of nodes.

In this section, we describe a modeling approach that addresses these needs. We first define a general regression model for networked data, and then derive a fast estimation algorithm based on a stochastic variational approximation to the collapsed posterior distribution over the model's parameters. While we focus our exposition on directed networks, our model applies to undirected networks with minimal modifications, as we illustrate in our application.

3.1 The Dynamic Mixed-Membership Stochastic Blockmodel

Let $G_t = (V_t, E_t)$ be a directed network observed at time t, with node-set V_t and edge-set E_t. For a pair of nodes $p, q \in V_t$, let $Y_{pqt} = 1$ if there exists a directed edge from node p to q, and $Y_{pqt} = 0$ otherwise. Each node $i \in V_t$ is assumed to be associated with a K-dimensional mixed-membership vector π_{it}, encoding the extent to which i belongs to each of K latent groups at time t.

To study how these mixed-memberships vary as a function of node-level predictors, and to allow such memberships to evolve over time, we further assume that the network at time t is in one of M latent states, and that a Markov process governs transitions from one state to the next. We then model each mixed-membership vector as a draw from the following Markov-dependent mixture,

$$\boldsymbol{\pi}_{it} \sim \sum_{m=1}^{M} \Pr(S_t = m \mid S_{t-1}) \times \text{Dirichlet}\left(\{\exp(\mathbf{x}_{it}^{\top}\boldsymbol{\beta}_{km})\}_{k=1}^{K}\right) \tag{1}$$

where the vector of predictors $\mathbf{x}_{it}$ is allowed to vary over time and the vector of coefficients $\boldsymbol{\beta}_{km}$ for group k is indexed by state m in the Markov process. Our model thus extends the traditional MMSBM by allowing the mixed membership vectors to not only be a function of node-level predictors, but also by letting these vectors to change over time as the Markov states evolve. Specifically, these random states are generated according to $S_t \mid S_{t-1} = n \sim$

Categorical($\mathbf{A}_n$), which is governed by a transition matrix $\mathbf{A}$ and the state at the previous time period, S_{t-1}. We define a uniform prior over the initial state S_1 and independent symmetric Dirichlet prior distributions for the rows of $\mathbf{A}$.

The model is completed by defining a $K \times K$ blockmodel matrix $\mathbf{B}$, with its $B_{gh} \in \mathbb{R}$ element giving the propensity of a member of group g to form a tie to a member of group h (for undirected network data, $\mathbf{B}$ is a symmetric matrix). Thus, we have,

$$Y_{pqt} \sim \text{Bernoulli}\left(g^{-1}\left(\mathbf{z}_{p\to q,t}^{\top}\mathbf{B}\mathbf{w}_{q\leftarrow p,t} + \mathbf{d}_{pqt}^{\top}\boldsymbol{\gamma}\right)\right) \tag{2}$$

where g^{-1} is the logistic function, and $\mathbf{z}_{p\to q,t} \sim \text{Multinomial}(\boldsymbol{\pi}_{pt})$ is an indicator vector for the group that node p chooses when interacting with node q at time t (and similarly for $\mathbf{w}_{q\leftarrow p,t}$). To relax the assumption of strict stochastic equivalence commonly used in other variants of the stochastic blockmodel, we also incorporate dyadic predictors $\mathbf{d}_{pqt}$ into the regression equation for the probability of a tie, with regression coefficients $\boldsymbol{\gamma}$.

Put together, the data generating process can be summarized as follows:

1. For each time period $t > 1$, draw a historical state $S_t \mid S_{t-1} = n \sim \text{Categorical}(\mathbf{A}_n)$.

2. For each node i at time t, draw state-dependent mixed-membership vector $\boldsymbol{\pi}_{it} \mid S_t = m \sim \text{Dirichlet}\left(\{\exp(\mathbf{x}_{it}^{\top}\boldsymbol{\beta}_{k,m}\}_{k=1}^{K})\right)$.

3. For each pair of nodes p and q at time t,

    - Sample a group indicator $\mathbf{z}_{p\to q,t} \sim \text{Multinomial}(\boldsymbol{\pi}_{pt})$.

    - Sample a group indicator $\mathbf{w}_{q\leftarrow p,t} \sim \text{Multinomial}(\boldsymbol{\pi}_{qt})$.

    - Sample a link between them $Y_{pqt} \sim \text{Bernoulli}\left(g^{-1}\left(\mathbf{z}_{p\to q,t}^{\top}\mathbf{B}\mathbf{w}_{q\leftarrow p,t} + \mathbf{d}_{pqt}^{\top}\boldsymbol{\gamma}\right)\right)$.

This data generating process results in the following joint distribution of observed network data and latent variables given a set of global hyper-parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B})$ and observed

covariates $(\mathbf{D}, \mathbf{X})$:

$$
\begin{aligned}
&P(\mathbf{Y}, \mathbf{L}, \mathbf{\Pi}, \mathbf{A} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B}, \mathbf{D}, \mathbf{X}) \\
&= P(S_1) \left[\prod_{t=2}^{T} P(S_t \mid S_{t-1}, \mathbf{A}) \right] \left[\prod_{t=1}^{T} \prod_{it \in V_t} P(\boldsymbol{\pi}_{it} \mid \mathbf{X}, \boldsymbol{\beta}, S_t) \right] \prod_{m=1}^{M} P(\mathbf{A}_m) \\
&\quad \times \left[\prod_{t=1}^{T} \prod_{p,q \in V_t} \Big[P(Y_{pqt} \mid \mathbf{z}_{p \to q,t}, \mathbf{w}_{q \leftarrow p,t}, \mathbf{B}, \boldsymbol{\gamma}, \mathbf{D}) P(\mathbf{z}_{p \to q,t} \mid \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{q \leftarrow p,t} \mid \boldsymbol{\pi}_{qt}) \right] \right]
\end{aligned}
\tag{3}
$$

where $\mathbf{L} := \{\mathbf{Z}, \mathbf{W}, \mathbf{S}\}$ collects all latent group memberships and hidden Markov states, $\mathbf{\Pi} := \{\boldsymbol{\pi}_{it}\}_{it \in V_t} \forall t$ collects all mixed-membership vectors, and transition matrix $\mathbf{A}$ is defined as before.

3.2 Marginalization

As we discuss in more detail in Section 3.3, we derive a factorized approximation to the posterior distribution proportional to Equation (3) in order to drastically reduce the computation time required for inference. A typical approximating distribution would factorize over all latent variables. In the true posterior, however, latent group indicators $\mathbf{z}_{p \to q,t}$ ($\mathbf{w}_{q \leftarrow p,t}$) and the mixed-membership parameters $\boldsymbol{\pi}_{pt}$ ($\boldsymbol{\pi}_{qt}$) are usually strongly correlated (Teh et al., 2007). Similarly, the Markov states S_t and parameters in the transition kernel $\mathbf{A}$ are typically highly correlated in the true posterior.

Therefore, and to avoid the strong assumption of independence induced by the standard factorized approximating distribution, we marginalize out the latent mixed-membership vectors and the Markov transition probabilities and then approximate the marginalized posterior. The details of the marginalization can be found in Appendix A.1. Letting $\alpha_{itkm} = \exp(\mathbf{x}_{it}^{\top} \boldsymbol{\beta}_{km})$, $\alpha_{it \cdot m} = \sum_{k=1}^{K} \alpha_{itkm}$, and $\theta_{pqtgh} = g^{-1}(B_{gh} + \mathbf{d}_{pqt}^{\top} \boldsymbol{\gamma})$, the result-

ing collapsed posterior is proportional to:

$$
\begin{aligned}
&P(\mathbf{Y}, \mathbf{L} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B}, \mathbf{X}) \\
&\quad \propto \prod_{m=1}^{M} \left[\frac{\Gamma(M\eta)}{\Gamma(M\eta + U_{m\cdot})} \prod_{n=1}^{M} \frac{\Gamma(\eta + U_{mn})}{\Gamma(\eta)} \right] \\
&\quad \times P(\mathbf{s}_1) \prod_{t=2}^{T} \prod_{m=1}^{M} \prod_{it \in V_t} \left[\frac{\Gamma(\alpha_{it\cdot m})}{\Gamma(\alpha_{it\cdot m} + 2N_t)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{itmk} + C_{itk})}{\Gamma(\alpha_{itmk})} \right]^{I(S_t = m)} \\
&\quad \times \prod_{t=1}^{T} \prod_{p,q \in V_t} \prod_{g,h=1}^{K} \left(\theta_{pqtgh}^{y_{pqt}} (1 - \theta_{pqtgh})^{1 - y_{pqt}} \right)^{z_{p \to q,t,g} \times w_{q \leftarrow p,t,h}}
\end{aligned}
\tag{4}
$$

where $I(\cdot)$ is the binary indicator function, and $\Gamma(\cdot)$ is the Gamma function. The marginalized joint distribution makes explicit use of a number of sufficient statistics: $C_{itk} = \sum_{q \in V_t} (z_{i \to q,t,k} + w_{i \leftarrow q,t,k})$, which represent the number of times node i instantiates group k across its interactions with all other nodes q present at time t (whether as a sender or as a receiver); $U_{mn} = \sum_{t=2}^{T} I(S_t = n) I(S_{t-1} = m)$, which counts the number of times the hidden Markov process transitions from state m to state n; and $U_{m\cdot} = \sum_{t=2}^{T} \sum_{n} I(S_t = n) I(S_{t-1} = m)$, which tracks the total number of times the Markov process transitions from m (potentially to stay at m).

3.3 Estimation via Variational Expectation-Maximization

For posterior inference, we rely on a mean-field variational approximation to the collapsed posterior distribution (Jordan et al., 1999; Teh et al., 2007). To do so, we define a factorized distribution over the latent variables $\mathbf{L}$ as

$$
\widetilde{Q}(\mathbf{L} \mid \mathbf{K}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) = \prod_{t=1}^{T} Q_1(\mathbf{s}_t \mid \boldsymbol{\kappa}_t) \prod_{p,q \in V_t} Q_2(\mathbf{z}_{p \to q,t} \mid \boldsymbol{\phi}_{p \to q,t}) Q_2(\mathbf{w}_{q \leftarrow p,t} \mid \boldsymbol{\psi}_{q \leftarrow p,t}), \tag{5}
$$

where $\boldsymbol{\kappa}_t$, $\boldsymbol{\phi}_{p \to q,t}$, and $\boldsymbol{\psi}_{q \leftarrow p,t}$ are variational parameters. Our factorized approximation assumes the latent state variables are independent in the collapsed space. This is a strong assumption, but one that has been found to strike a good balance between accuracy and scalability (see **?**).

We then apply Jensen's inequality to derive a lower bound for the log marginal proba-

bility of our network data $\mathbf{Y}$

$$P(\mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B}, \mathbf{X}) \geq \mathcal{L} \triangleq \mathbb{E}_{\widetilde{Q}}[\log P(\mathbf{Y}, \mathbf{L} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B}, \mathbf{X})] - \mathbb{E}_{\widetilde{Q}}[\log \widetilde{Q}(\mathbf{L} \mid \mathbf{K}, \boldsymbol{\Phi}, \boldsymbol{\Psi})] \quad (6)$$

and optimize this lower bound with respect to the variational parameters to approximate the true posterior over our latent variables (Jordan et al., 1999). To do so, we iterate between finding an optimal $\widetilde{Q}$ (the E-step) and optimizing the corresponding lower bound with respect to the hyper-parameters $\mathbf{B}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ (the M-step).

After initializing all sufficient statistics and variational parameters, our E-step begins by updating the $\boldsymbol{\phi}$ parameters for all (pt, qt) dyads in our data as follows:

$$\hat{\phi}^{(s)}_{p \to q,t,k} \propto \prod_{m=1}^{M} \left[\exp\left[\mathbb{E}_{f,\widetilde{Q}_2}[\log(\alpha_{ptmk} + C'_{ptk})]\right]\right]^{\kappa_{tm}} \prod_{g=1}^{K} \left(\theta_{pqtkg}^{y_{pqt}}(1 - \theta_{pqtkg})^{1-y_{pqt}}\right)^{\psi_{q \leftarrow p,t,g}} \quad (7)$$

where $C'_{ptk} = C_{ptk} - z_{p \to q,t,k}$ and the expectation is taken over the variational distribution of $\mathbf{Z}$. By symmetry, the update for $\psi_{q \leftarrow p,t,k}$ is similarly defined.

In turn, and for $t = 2, \ldots, T-1$, we update all hidden Markov state variational parameters according to

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

where $U'_{m\cdot} = U_{m\cdot} - s_{t,m}$ 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).

In order to avoid a costly computation of the Poisson-Binomial probability mass function (which is required when computing expected values that involve sufficient statistics), we approximate the expectations in these updates by using a zeroth-order Taylor series expansion, so that $\mathbb{E}_{f,\widetilde{Q}_2}[\log(\alpha_{ptkm} + C'_{ptk})] \approx \log\left(\alpha_{ptkm} + \mathbb{E}_{f,\widetilde{Q}_2}\left[C'_{ptk}\right]\right)$ and similarly for terms involving all $U'_{\cdot}$ counts (Asuncion et al., 2009).

Finally, during the M-step, we find locally optimal values of $\mathbf{B}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ with respect to the lower bound, given by

$$\begin{aligned}\mathcal{L}_{\phi,\kappa}(\mathbf{B},\boldsymbol{\beta},\boldsymbol{\gamma}) \triangleq & \sum_{t=1}^{T}\sum_{m=1}^{M}\kappa_{tm}\sum_{p\in V_t}\log\Gamma\left(\xi_{ptm}\right)-\log\Gamma\left(\xi_{ptm}+2N_t\right) \\ & +\sum_{t=1}^{T}\sum_{m=1}^{M}\kappa_{tm}\sum_{p\in V_t}\sum_{k=1}^{K}\mathbb{E}_{\tilde{Q}}[\log\Gamma(\alpha_{ptmk}+C_{ptk})]-\log\Gamma(\alpha_{ptmk}) \\ & +\sum_{t=1}^{T}\sum_{(p,q)\in E_t}\sum_{g,h=1}^{K}\phi_{p\to q,t,g}\psi_{q\leftarrow p,t,h}\left\{y_{pqt}\log(\theta_{pqtgh})+(1-y_{pqt})\log(1-\theta_{pqtgh})\right\} \\ & -\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{(p,q)\in E_t}\sum_{k=1}^{K}\{\phi_{p\to q,t,k}\log(\phi_{p\to q,t,k})-\psi_{q\leftarrow p,t,h}\log(\psi_{q\leftarrow p,t,k})\}\end{aligned} \tag{8}$$

using a quasi-Newton method.[1] To regularize the fit, we define independent standard Normal priors for all parameters. When required, standard errors for these quantities are obtained 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 $\mathbf{B}$.

3.3.1 Stochastic VI Algorithm

For problems involving large networks, the above variational approximation can be very computationally intensive (even after parallelization — see section 3.4). To enable fast inference on a set of networks with a very large number of nodes over multiple time periods, we define an alternative optimization strategy which relies on the stochastic gradient ascent approach proposed by Hoffman et al. (2013), as applied to our collapsed variational target (Foulds et al., 2013; Dulac et al., 2020).

Like other stochastic VI (SVI) algorithms, ours follows a random gradient with expected value equal to the true gradient of the lower bound in Equation 8. To form this unbiased gradient, and at each step of the algorithm, we sample a mini-batch of nodes within each time period t uniformly at random, and form subgraphs $\mathbf{Y}_t^{(s)}$ among all dyads

[1] We provide the needed gradients in Appendix A.2.

containing the sampled nodes. The algorithm proceeds by optimizing the local variational parameters (i.e. $\mathbf{\Phi}$ and $\mathbf{K}$) for all dyads (p, q) in each $\mathbf{Y}_t^{(s)}$ using the updates given in the previous section, holding global counts constant at their most current values. We then condition on these locally updated variational parameters and obtain an intermediate value of all global counts (i.e. $\mathbf{C}$ and $\mathbf{U}$) by computing their expected value under the mini-batch sampling distribution. We finalize each step by updating these global counts using a weighted average, as follows:

$$\mathbf{C}^{(s)} = (1-\rho_s)\mathbf{C}_t^{(s-1)} + \rho_s\, \mathbb{E}_f\left[\mathbf{C}_t\right]; \quad \mathbf{U}_t^{(s)} = (1-\rho_s)\mathbf{U}^{(s-1)} + \rho_s\, \mathbb{E}_f\left[\mathbf{U}\right] \tag{9}$$

where we set the step-size $\rho_s = (\tau + s)^{-p}$, and $p \in (0.5, 1.0]$ and $\tau \geq 0$ are researcher-set arguments controlling the extent to which previous iterations affect current values of the sufficient statistics (Hoffman et al., 2013; Cappé and Moulines, 2009). To set the values of our hyperparameters we once again follow an empirical Bayes approach, updating the hyper-parameters along with the global sufficient statistics by taking a step in the direction of the gradient of the stochastic lower bound. For instance, and for the $\boldsymbol{\gamma}$ coefficients, we have:

$$\boldsymbol{\gamma}^{(s)} = \boldsymbol{\gamma}^{(s-1)} + \rho_s \nabla_\gamma \mathcal{L}^{(s)}_{\hat{\phi},\hat{\kappa}}(\boldsymbol{\gamma}) \tag{10}$$

where

$$\mathcal{L}^{(s)}_{\phi,\kappa}(\boldsymbol{\gamma}) = \sum_{t=1}^{T} \frac{|E_t|}{|E_t^{(s)}|} \sum_{(p,q)\in E_t^{(s)}} \sum_{g,h=1}^{K} \phi_{p\to q,t,g}\psi_{q\leftarrow p,t,h}\big\{ y_{pqt}\log(\theta_{pqtgh}) + (1-y_{pqt})\log(1-\theta_{pqtgh})\big\}$$

is a random function that is equal to the third line in Equation 8 in expectation. The updates for all other hyper-parameters are similarly defined (Hoffman et al., 2013), and provide the required gradients in Appendix A.2.

When using the correct schedule for the step-sizes ρ_s, this procedure is guaranteed to find a local optimum of the lower bound without the need to perform a costly update over

the parameters associated with alls dyads at every iteration (Gopalan and Blei, 2013), and we illustrate gains in scalability of our SVI approach in Appendix A.3.4.

3.4 Implementation Details

Like other mixed-membership models, there are important practical considerations when fitting the dynMMSBM. First, finding good starting values is essential. In particular, the quality of starting values for the sufficient statistics in the $\mathbf{C}$ global terms 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).

Second, and to establish convergence of our collapsed variational algorithm, we evaluate absolute change in the estimated hyper-parameters, and stop iterating when all changes fall below a user-defined tolerance level ($1.0e^{-4}$ in our application). In the case of the SVI algorithm, we retain a small sample of dyads (viz. 1% of all pairs in our application) before initialization and evaluate its log-likelihood after each iteration, stopping when change falls below a tolerance of $a0^{-3}$ or when no improvement has been observed in the past 10 iterations. The stopping rule based on a held-out sample helps us avoid overfitting, and reduces the amount of "jitter" induced by the stochastic gradient descent. Finally, and to maximize computational efficiency, we exploit the assumption of conditional independence across edges and optimize local parameters $\mathbf{\Phi}$ in parallel across (subsampled) dyads.

In Appendix A.3, we conduct a series of validation simulations, in which we evaluate the estimation accuracy using a set of simulated dynamic networks, and compare the results of fitting a fully specified dynMMSBM and fitting a separate MMSBM (without covariates)

to each time period. We show the substantial gains in error reduction resulting from the use of our proposed model. Next, and having defined a suitable model for our data, we turn to the study of the issue of the democratic peace.

4 Empirical Analysis

We now apply the dynMMSBM to study the onset of militarized disputes among 216 states in the years 1816–2010, based on the Militarized Interstate Dispute (MID) dataset version 4.1 (Ghosn and Bennett, 2003).[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 well in out-of-sample prediction tests (see Table A2 in Appendix A.4 for prediction tests and Figures A7 and A8 for a visualization of blockmodel estimates for specifications with five and seven groups).

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

[2]Our primary results reflect a batch analysis of the data, taking all years into consideration. In Appendix A.5, we use out-of-sample prediction to evaluate forecasting performance. In Appendix A.4, we replicate our analysis via online updating, where we iteratively expand the time window to update estimates in real time (see Table A7 and Figures A9 and A10).

formed when a new dispute 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 842,685 state dyad-year observations in our sample.

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 between states in a given year (`ALLIANCE`); 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).[3] 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, we control for further temporal trends using a count of years since the last militarized dispute between each dyad and a cubic spline (Beck et al., 1998). Fi-

[3] As an alternative way to address geographic effects, we estimate a specification that includes a set of regional indicator variables (see Table A5 and Figures A5 and A6 in the appendix).

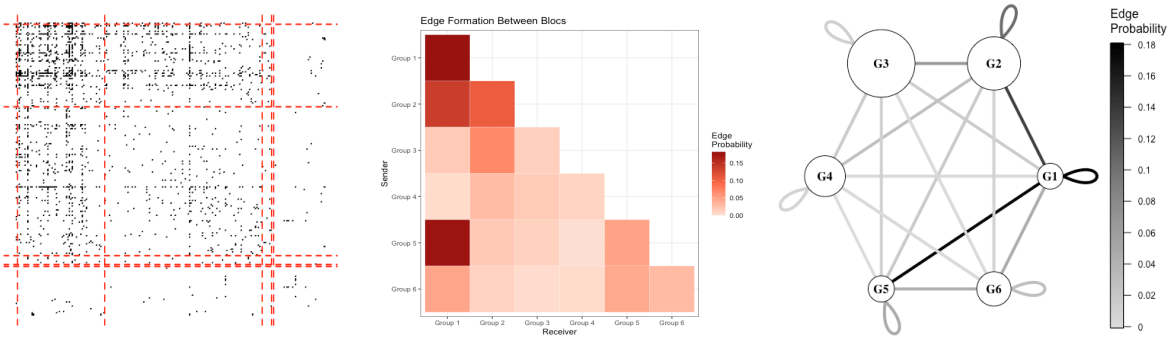

Figure 1: **Estimated blockmodel in the conflict network**. The left panel displays the adjacency matrix of militarized disputes between 216 states. Black squares indicate the existence of at least one MID between the states in row x and column y; dotted lines separate states by estimated group membership. The middle panel displays the estimated probability of conflict between groups as a heat map. The right panel is a network graph summarizing the estimated blockmodel, where size of the nodes (circles) reflects aggregate membership in each group and weighted edges (lines) reflect the probability of conflict.

nally, 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.

The model is fitted using our open-source software package NetMix. Estimation took one hour and eighteen minutes on a computer with a 3.6Ghz CPU, converging after 709 EM iterations. Note that the estimation time drops to approximately 55 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 illustrates how patterns of interstate conflict inform the estimation of group memberships. The left panel shows the 216 $\times$ 216 adjacency matrix of militarized disputes between countries, aggregated over the entire time period. Black squares indicate the existence of at least one MID between the country represented by row x and the country in column y. The dynMMSBM assigns each country to a mixture of the six latent groups, each of which initiates disputes at unique rates. In

the matrix, we sort countries by estimated group membership – demarcated in the figure by dotted lines – to demonstrate the varying rates of conflict within and between groups.

The middle panel of Figure 1 shows the estimated rates of conflict between groups. For example, group 1 has elevated rates of intra-group conflict as well as frequent conflict with groups 2 and 5, as evidenced by the darker shade of these cells in the figure. Groups 4 and 5 have the most peaceful relations, initiating disputes with each other only 0.14% of the time. Table A3 of Appendix A.4 presents the exact blockmodel estimates used to create the figure.

The right panel combines information on group membership and dispute rates, depicting each latent group as a node on a graph. The size of the nodes (circles) reflects the estimated membership size of the group. Group 3 is the most populous, representing 39.9% of country-year observations in the sample. Group 2 is the second largest (27.2%), followed by Groups 4 (16.2%), 6 (10.4%), 5 (3.3%), and 1 (2.9%). The edges (lines) depict the estimated rates of conflict between groups, with darker-shaded edges indicating a higher propensity of conflict onset.

To gauge the validity of these estimates, we examine whether the group assignments and dispute probabilities correspond to known historical conflict patterns. Our model estimates that when a country from Group 1 interacts with a country from Group 2, there is an unusually high probability (13.7%) 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, Canada, United Kingdom, and their Western European allies often instantiate Group 1, while China, Russia, and other Eastern bloc countries tend to instantiate membership in Group 2.

Other groups also reveal important structure in the international system. Group 3 includes many countries that maintained a foreign policy of neutrality throughout much of the 19th and 20th centuries (e.g., Norway, Finland, Ireland, 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 Group 3 countries have a low rate of conflict with Group 1 (1.7%) and are less bellicose overall. Group 4 includes many countries that were caught in the crossfire of the intense geopolitical conflict between the Western and Eastern coalitions represented by Groups 1 and 2. Afghanistan, Angola, and Cambodia are among the countries with high membership in Group 4 that were sites of proxy conflicts during the Cold War period. Group 5 is composed of many autocratic countries in the Middle East and Africa, while Group 6 features small or geographically remote countries.

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 the underlying geopolitical structure of the Cold War era, we identify the 15 countries with the highest average membership probability in each latent group during the period of 1950–1990. We do this by computing $\frac{1}{T}\sum_{t=1950}^{1990}\pi_{ptg}$ for every country in a given latent group g. The countries with the highest membership in each group are listed in Table A4 of Appendix A.4.

The distribution of countries across the groups is consistent with presence of competing geopolitical coalitions during the Cold War. Group 1 contains the major NATO allies, including the United States, United Kingdom, West 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, Poland, Czechoslovakia, and Romania). The estimated blockmodel indicates the competing coalitions experience abnormally high rates of conflict.

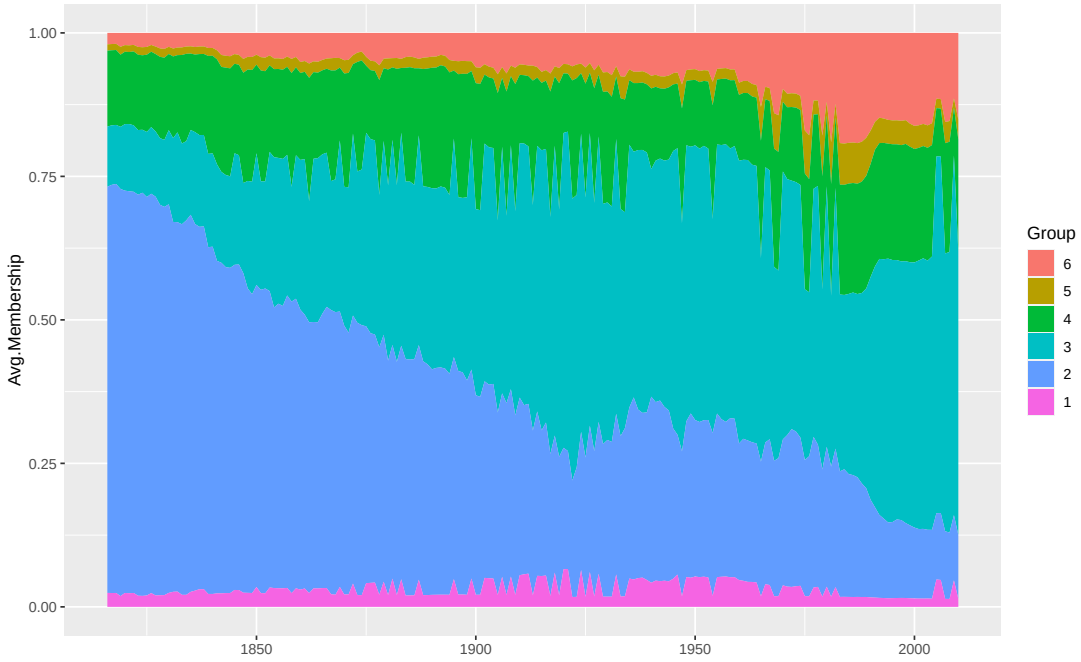

Figure 2: **Membership in Latent Groups over Time**. The figure shows the average proportion of membership in six latent groups for each year from 1816–2010.

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 from 1816-2010. Latent groups expand and contract as countries move in and out of geopolitical coalitions. Group 2 — populated by autocratic countries with high military capacity — noticeably declines in membership throughout the period. This reflects a general trend toward democratization among industrialized countries, as well as geopolitical transitions of the Soviet client states after the Cold War concluded. The most peaceful clusters, Group 3 and 4, increase in membership over the period, which may be attributable to the consolidation of norms against military aggression. In the post-World War II era, decolonization and independence movements led to a substantial increase in the number of independent countries. This likely accounts for the late growth of Group 6 — a cluster representing small countries with limited military capability.

The evolution of groups shown in Figure 2 are consistent with international relations

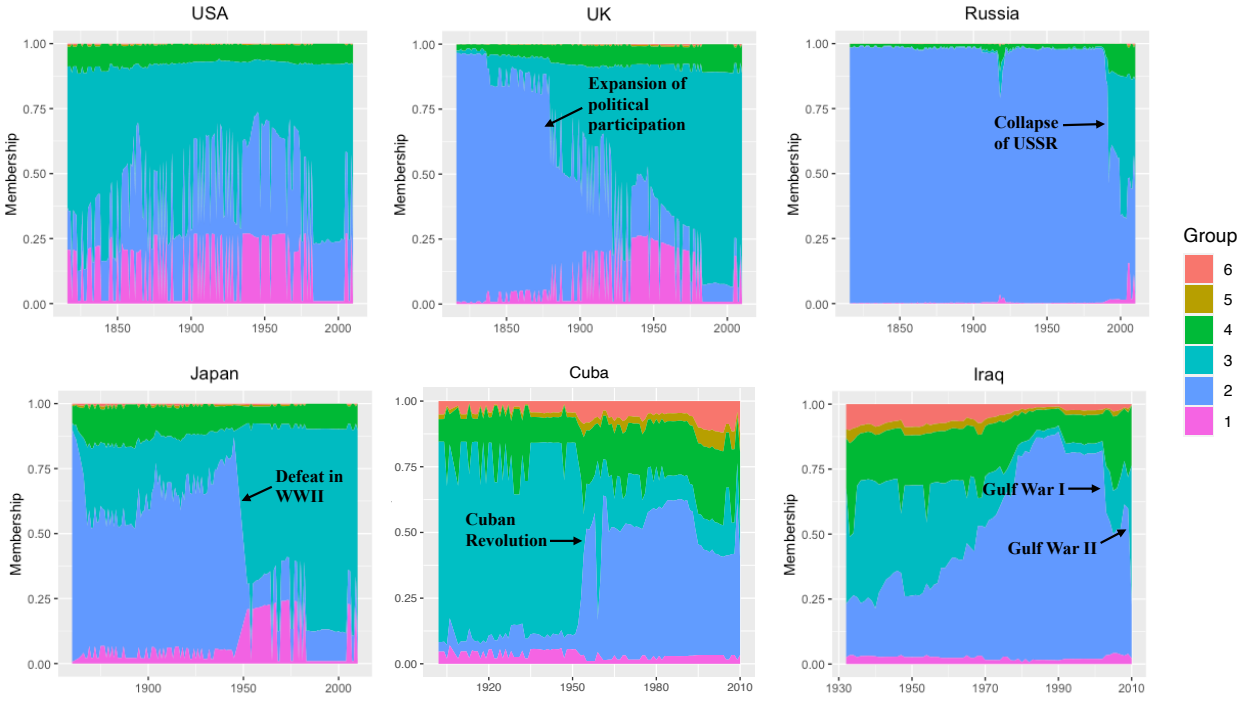

Figure 3: **Average Node Membership over Time, Select Countries**. The figure shows, for six countries, the average rate of membership in four latent groups in each year the country is present in the network.

scholarship emphasizing dynamic change in conflict patterns. Cederman (2001), for example, proposes a dynamic learning process in which democratic countries consolidate peaceful relations over time. The observed growth of Group 3 — a cluster populated by democracies with very low rates of conflict — supports this hypothesis.

Figure 3 displays the evolution of group membership for a select group of countries. There is significant variation across countries and within some countries over time. The United States and United Kingdom feature relatively high membership in Group 1 compared to other countries, as discussed above. They also exhibit significant membership in Group 3, the other Western-leaning and democratic cluster. US membership is comparably stable over the period of the study, while the UK consolidates its membership in these groups after transitioning to a democratic political system. For example, we observe a sharp increase in the UK's membership in Group 3 following the 1867 Reform Act, which newly enfranchised parts of the urban working class. Russia's membership is overwhelmingly

dominated by Group 2. At the end of the Cold War, the implosion of the Soviet system shifts Russian membership toward Group 3 with a slight reversion in the last few years.

Japan, Cuba, and Iraq further demonstrate how political shocks like revolution and foreign intervention affect conflict patterns in ways that are reflected in latent membership. Japan experiences a sudden shift from Group 2 to Groups 1 and 3 upon its defeat in World War II and subsequent occupation by American forces. The shift in membership corresponds with a clear change in the country's conflict patterns. Japan's overall rate of conflict declined from 2.7% prior to 1945 to 0.7% thereafter. More than 60% of Japan's disputes in the post-1945 period were with Group 2 members Russia, China, and North Korea.

Cuba's membership in Group 2 increases sharply following the onset of the Cuban Revolution and the ascension of the Castro regime. The country experiences consistently high Group 2 membership since the 1950s, with a slight attenuation in the last few decades. In turn, Iraq features two breaks in latent membership that correspond to conflicts with the United States. Following the first Gulf War in 1990-1991, we observe reduced membership in Group 2 and increases in Groups 3 and 4. A similar shift in 2003 reflects the invasion by the US and allied countries and the installation of a new 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. In the interest of space, and since the majority of the time period under study (viz. 51.3%) is estimated to derive from this state, we display the coefficients only for Markov state 1. See Table A6 in Appendix A.4 for Markov state 2 coefficients.

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

Predictor	Dyadic	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
INTERCEPT		12.016 (1.069)	16.539 (1.069)	11.383 (1.069)	12.376 (1.069)	8.836 (1.074)	7.389 (1.066)
POLITY		0.083 (1.084)	-0.251 (1.083)	0.076 (1.084)	-0.115 (1.083)	-0.091 (1.096)	-0.091 (1.079)
MILITARY CAPABILITY		0.638 (1.029)	1.192 (1.029)	0.130 (1.025)	0.513 (1.029)	0.235 (1.048)	-0.134 (1.059)
BORDERS	2.123 (0.001)						
DISTANCE	-0.0001 (0.002)						
ALLIANCE	0.087 (0.001)						
IO CO-MEMBERS	0.009 (0.002)						
PEACE YRS	-0.021 (0.002)						
N nodes: 216; N dyad-years: $842,685$; N time periods: 195 Lower bound at convergence: $-527,587.7$							

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 the majority of the time period. The estimated coefficients for cubic splines and indicators for variable missingness are not shown.

the Western alliance of liberal democracies during the Cold War, and Group 3 as Western-leaning neutral states. Notably, these two democratic clusters exhibit significantly different patterns of conflict. Group 1 countries have a high rate of military disputes, both with other Group 1 members (18.2%) and with other groups (7.7%). Group 3 countries are significantly more consistent with the democratic peace hypothesis. Predicted conflict between members of this group are rare (0.14%), and they also have a lower dispute rate with other latent groups (2.3%).

Other monadic coefficients are largely consistent with the descriptive patterns discussed above. Autocratic regimes sort into Group 2 at the highest rate. Greater military capability is negatively associated with membership in Group 6 and positively 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 conflict 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, countries in the conflict network are assigned to a latent group each time they interact with another country 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 country's POLITY score is increased by one standard deviation (6.78), making sure scores increase only up to the maximum value (10). The overall average effect of this change on the probability of edge formation, averaging all dyadic interactions and time periods,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{|V_t \times V_t|}\sum_{p,q\in V_t}[\,\mathbb{E}(y_{pqt} \mid \texttt{POLITY} + 6.78) - \mathbb{E}(y_{pqt})]$$

is negative but negligible in size: -0.001. Thus, increasing the degree of democracy in a country results in a minor decrease in overall conflict, given the underlying geopolitical coalitions throughout the time period.

There is, however, a significant amount of heterogeneity in this effect across countries and over time. Figure 4 shows, for a large set of countries, the difference in expected probability of interstate conflict due to an increase of one standard deviation in POLITY score. Many countries (such as Germany, Russia, and Iraq) are predicted to be substantially more peaceful, on average, if they were more democratic during the period of the study. Others, however, experience very little change in conflict behavior (e.g., Australia and Nicaragua). A handful of countries are estimated to become *more* conflict prone (e.g., Kosovo, Montenegro, and Brunei). An increase in polity shifts these countries 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

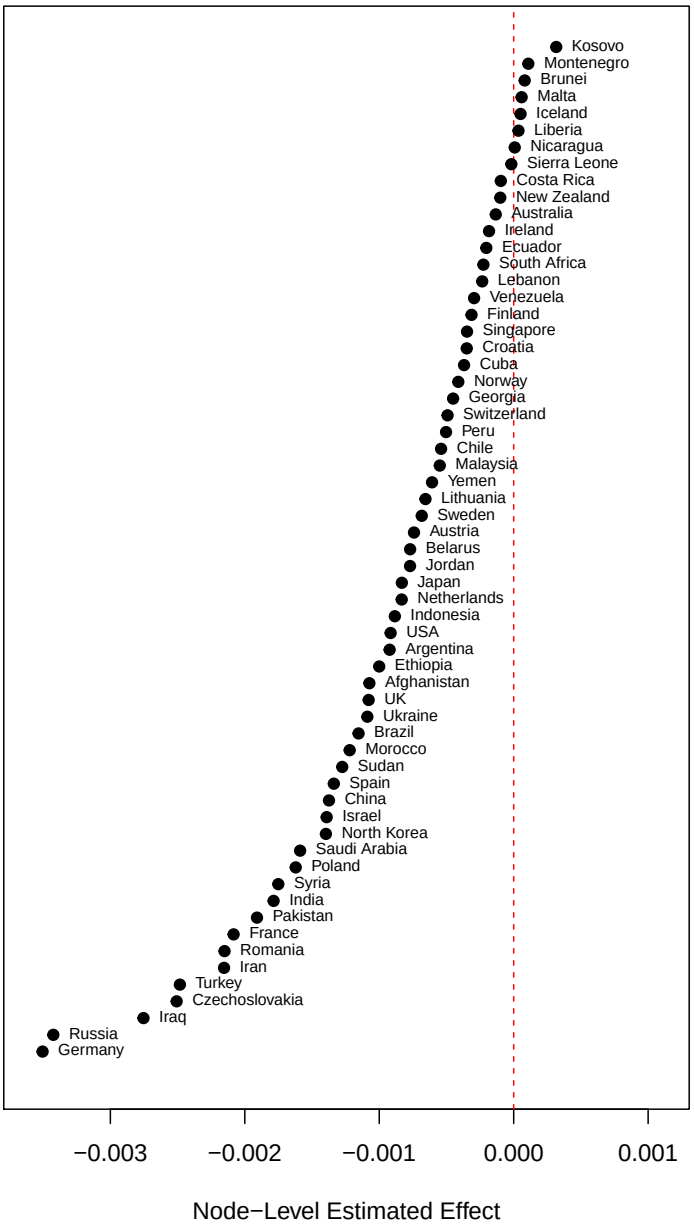

Figure 4: **Estimated Effects of Covariate Shift in Polity over Time, Select States**. The figure shows the estimated change in the probability of interstate conflict if a state's POLITY score is increased by one standard deviation (6.78) from its observed value.

network. Figure 5 looks within countries to gauge the effect of the shift in POLITY over time, revealing additional heterogeneity. To illustrate how monadic effects can vary within countries, consider the sharp drop in the estimated effect of POLITY for Russia from 1918-1921. This period is preceded by the ascendance of the Bolshevik government, which took power in November 1917. Over the next few years, the government engaged in a series

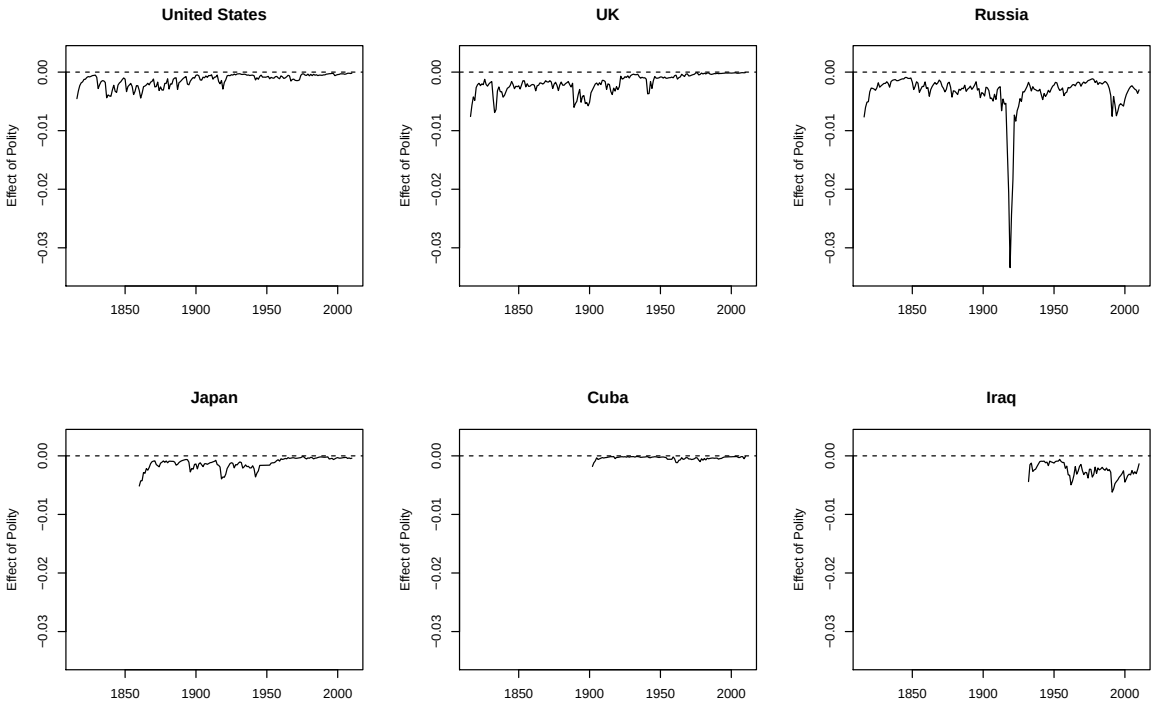

Figure 5: **Effect of Shift in Polity over Time, Select States**. The figure shows the estimated change in the probability of interstate conflict over time if a country's POLITY score is increased by one standard deviation (6.78) from its observed value (up to a maximum of 10).

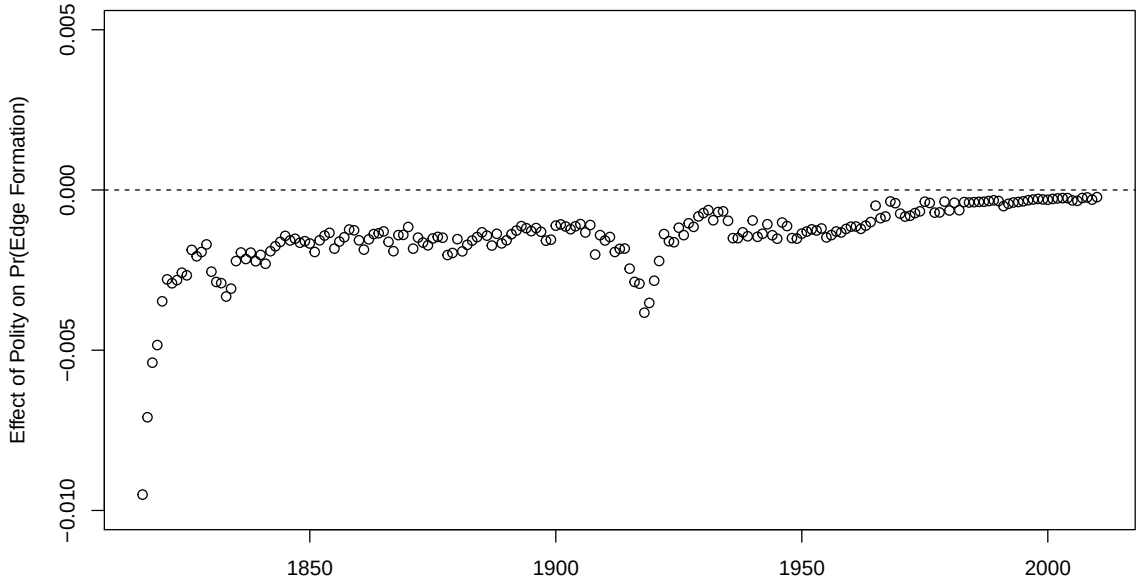

Figure 6: **Estimated Aggregate Effect of Shift in Polity over Time**. The figure shows the estimated average change in the probability of interstate conflict when countries' POLITY scores are increased by one standard deviation (6.78) up to the maximum POLITY score.

of militarized disputes with the Allied Powers of WWI, who supported anti-communist forces during the Russian Civil War. This pattern of disputes is consistent with the estimated blockmodel, which predicts an elevated rate of conflict between Group 1 (US, UK, France, Japan) and Group 2 (Russia). The estimates in Figure 5 compare these patterns of conflict to a counterfactual world in which Russia had a more democratic political system. Increasing Russia's POLITY score from its observed value in 1918 (-1) to a higher value (6) shifts the expected group membership for Russia away from Group 2 (from 75.4% to 32.4%) and toward Group 3 (from 10.3% to 40.7%). This reduces the likelihood of disputes, since Group 3 has significantly lower rates of inter- and intra-group conflict. By 1922 the Bolshevik regime consolidated power and the country's POLITY score drops to -7, after which an equivalent increase in POLITY has a smaller effect.

Figure 6 displays the average effect of POLITY for each year in the time period. An increase in democracy induces less conflict, on average, throughout most of the sample. The effect is noticeably lower during the pre-WWII period, hitting a local minimum in 1918 (-0.004). The impact of polity has 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 has no statistically discernible effect on conflict propensity. Somewhat surprisingly, the presence of a formal alliance and joint membership in international organizations increase the likelihood of conflict, though these effects are substantively small.

In Appendix A.5, we compare the results of our empirical analysis with those of the standard logistic regression model, which assumes all dyad-years are conditionally inde-

pendent and forces all node-level predictors to be transformed into dyadic form. We furthermore emulate the process of analyzing data in real-time by estimating both models using data from 1816-2008 and then evaluating model performance on what the forecasting predictions would have been during the two following years, 2009 and 2010. We find that the dynMMSBM significantly outperforms the conventional approach in the Diebold-Mariano test for forecasting comparison (Diebold and Mariano, 1995). It also marginally improves on the logistic model in area under the ROC curve, though the difference is not statistically significant.

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.

Cappé, O. and Moulines, E. (2009), "On-line expectation–maximization algorithm for latent data models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 593–613.

Cederman, L.-E. (2001), "Back to Kant: Reinterpreting the democratic peace as a macro-historical learning process," *American Political Science Review*, 15–31.

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.

Diebold, F. and Mariano, R. (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Dulac, A., Gaussier, E., and Largeron, C. (2020), "Mixed-Membership Stochastic Block Models for Weighted Networks," in *Conference on Uncertainty in Artificial Intelligence*, PMLR, pp. 679–688.

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.

Foulds, J., Boyles, L., DuBois, C., Smyth, P., and Welling, M. (2013), "Stochastic collapsed variational Bayesian inference for latent Dirichlet allocation," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 446–454.

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.

Gopalan, P. K. and Blei, D. M. (2013), "Efficient discovery of overlapping communities in massive networks," *Proceedings of the National Academy of Sciences*, 110, 14534–14539.

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.

Harvey, D., Leybourne, S., and Newbold, P. (1997), "Testing the equality of prediction mean squared errors. International Journal of forecasting," *International Journal of forecasting*, 13, 253–263.

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.

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013), "Stochastic variational inference," *The Journal of Machine Learning Research*, 14, 1303–1347.

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 (**??**) in Section 3.2.

A.2 Details of the Collapsed Variational Algorithm

A.2.1 Expectation Steps

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 (**??**) 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 **??** 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 Maximization steps

Lower Bound

M-step 1: update for B

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

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

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

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

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:

$$\mathcal{L}(\widetilde{Q}) = \sum_{t=1}^{T} \sum_{m=1}^{M} \kappa_{tm} \sum_{p \in V_t} \left[\log \Gamma(\xi_{ptm}) - \log \Gamma(\xi_{ptm} + 2N_t) \right]$$

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

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

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.

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 A1.

Generating a single, 9-period network under each of these scenarios results in the mixed memberships depicted in Figure A1, 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

	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 A1: **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.

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.

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 A2 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 A2 reveals. For each cell of the blockmodels, the true probability of an

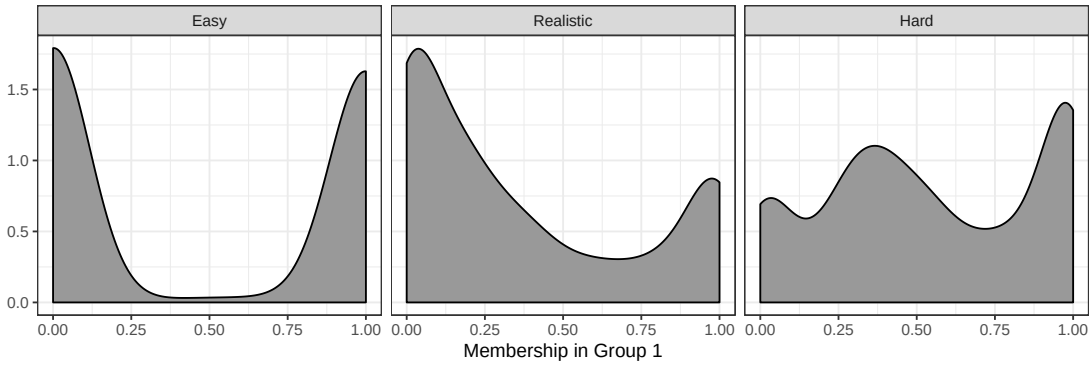

Figure A1: **Simulated mixed-memberships in synthetic networks**. The plots depict the mixed-membership vectors of nodes in three simulated networks, each with 100 nodes observed over 9 time periods. It shows the memberships of nodes in networks generated under an 'easy' DGP (i.e. one where memberships 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.

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.

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 A3 shows, for each time period, the distribution of estimated effect sizes of the 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

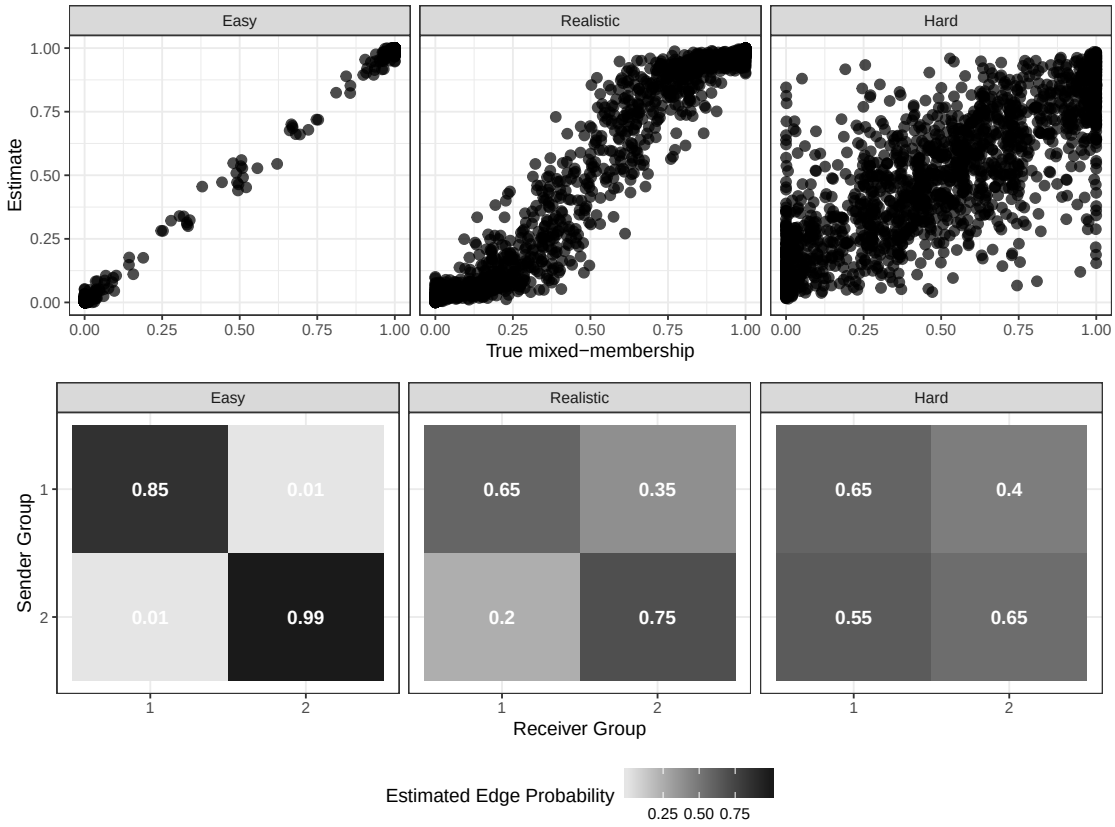

Figure A2: **Estimation accuracy**. For each DGP scenario, the figure shows the estimated mixed-membership vectors (top row) and the estimated blockmodels (bottom row) against their known values (indicated by the white numbers in each cell of the blockmodel for the bottom row). Overall, accuracy of retrieval both sets of parameters depends on the complexity of the learning problem, although recovery is generally very good, even under 'hard' inferential conditions.

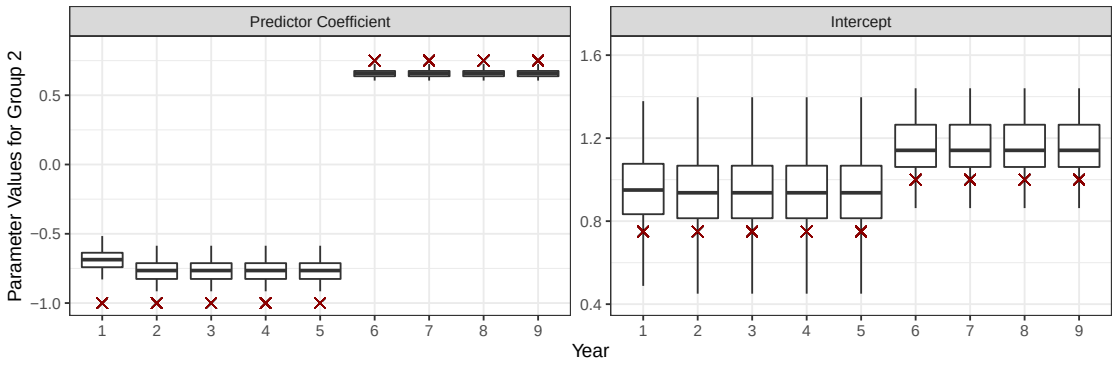

Figure A3: **Estimated parameters of block membership regression**. The figure shows, for each time period, the HMM-weighted effect of a continuous predictor on the probability of instantiating latent group 2 (left panel), and the HMM-weighted intercept of the corresponding regression line (right panel), 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.

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 A2).

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 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 A4.

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 realign-

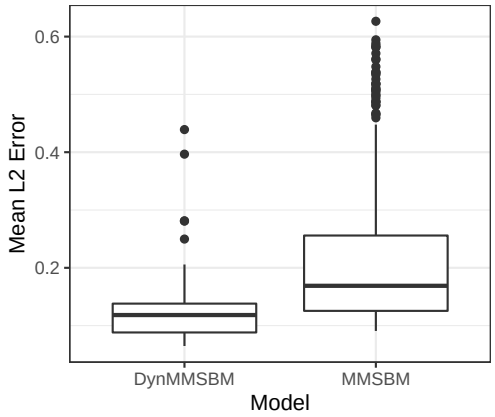

Figure A4: **Error for estimated mixed membership vectors**. The figure shows average L_2 distances between estimated and true mixed-membership vectors for all nodes in each of 50 replicated dynamic networks. On the left, estimates are generated using dynMMSBM. On the right, estimates are generated using the canonical MMSBM, fit separately to the nine time periods in each simulated network.

ment 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.3.4 Results with Stochastic Variational Inference

A.4 Additional Empirical Results

# Groups	AUROC
2	0.966
	(0.020)
3	0.989
	(0.012)
4	0.986
	(0.013)
5	0.984
	(0.014)
6	0.986
	(0.013)
7	0.975
	(0.017)

Table A2: **Out of Sample Prediction, Different Latent Groups**. The table displays the area under the ROC curve (AUROC) and associated standard error for specifications with 2-7 Latent Groups. Each model is fit on data from 1816-2008 and used to forecast conflict in the period 2009-2010.

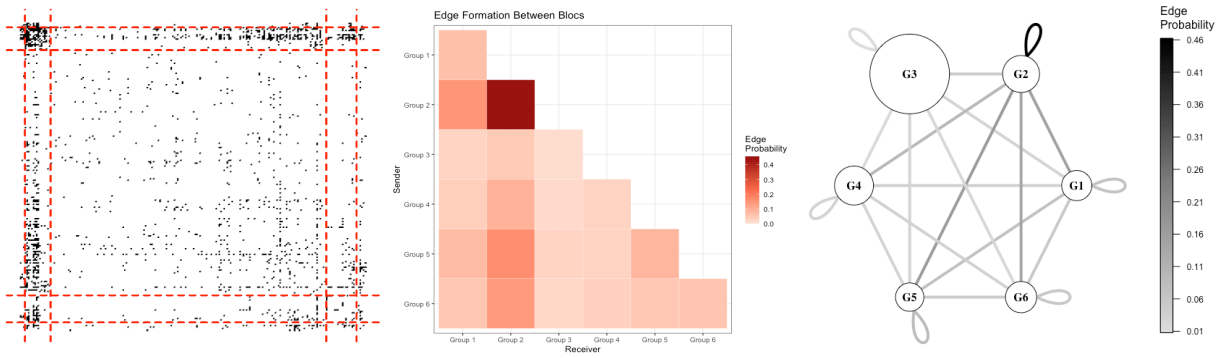

Figure A5: **Estimated blockmodel in the conflict network (regional model)**. Blockmodel visualizations for a specification including an indicator for state region. The left panel displays the adjacency matrix of militarized disputes. The middle panel displays the estimated probability of conflict between groups as a heat map. The right panel is a network graph summarizing the estimated blockmodel. The blockmodel in this specification is moderately correlated (0.30) with the primary model.

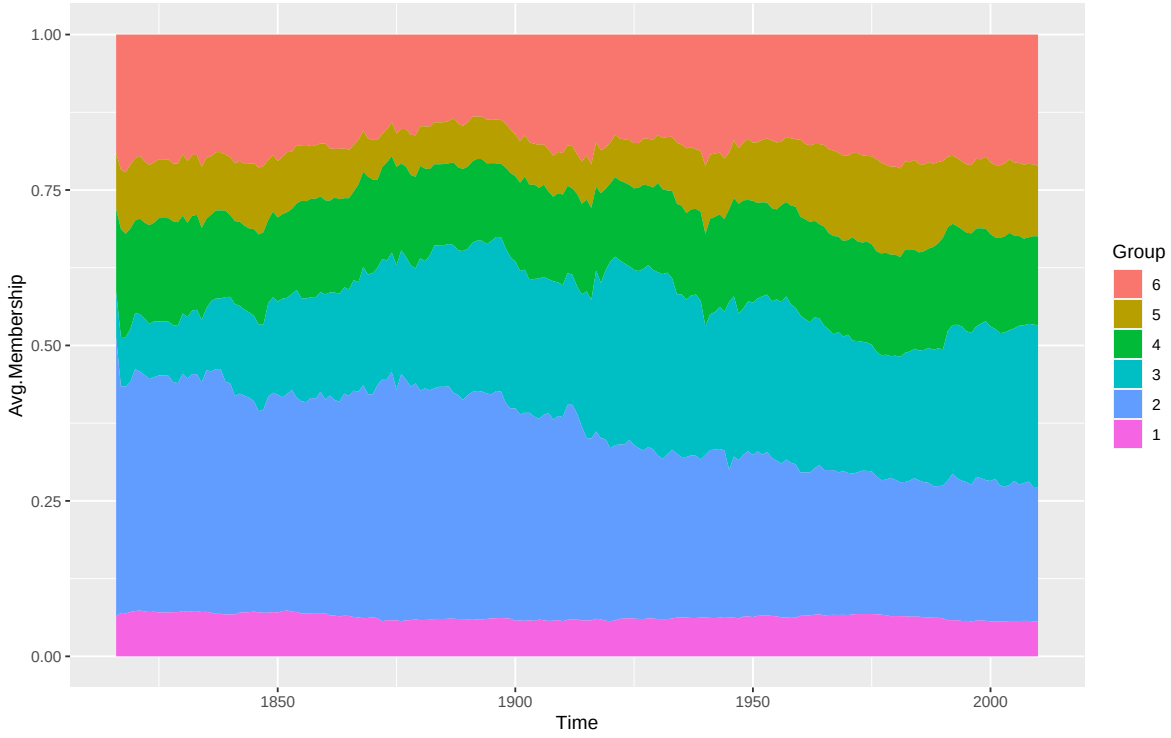

Figure A6: **Membership in Latent Groups over Time (regional model)**. The figure shows the average proportion of membership in six latent groups for each year from 1816–2010. The estimated evolution of membership in this specification is similar in some respects to the primary specification (e.g., the steady decline of Group 2 membership over the time period), but differs in others (e.g., the larger size of Group 6 in this model). Notably, this specification does not experience transitions in the hidden Markov state. This may be attributable to the addition of indicators for state region, which are static over time.

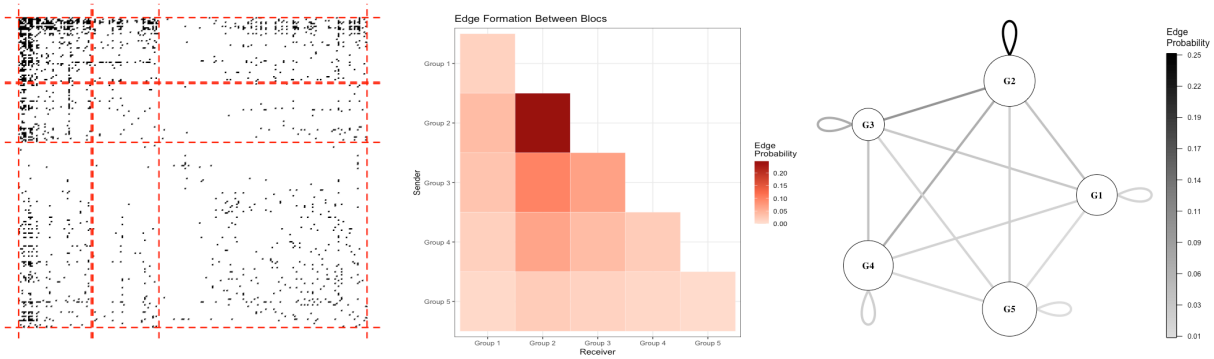

Figure A7: **Estimated blockmodel in the conflict network (5-group specification)**. The left panel displays the adjacency matrix of militarized disputes between 216 states. The middle panel displays the estimated probability of conflict between groups as a heat map. The right panel is a network graph summarizing the estimated blockmodel.

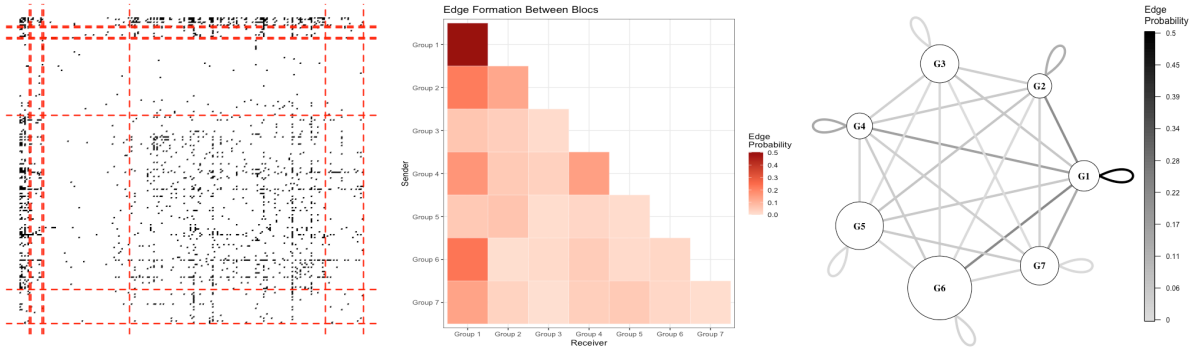

Figure A8: **Estimated blockmodel in the conflict network (7-group specification)**. The left panel displays the adjacency matrix of militarized disputes between 216 states. Dotted lines separate states by estimated group membership; some groups are not visible in the adacency matrix since they have very low membership. The middle panel displays the estimated probability of conflict between groups as a heat map. The right panel is a network graph summarizing the estimated blockmodel.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Group 1	0.182	0.137	0.017	0.004	0.180	0.047
Group 2	0.137	0.105	0.068	0.028	0.018	0.011
Group 3	0.017	0.068	0.014	0.017	0.011	0.003
Group 4	0.004	0.028	0.017	0.009	0.001	0.004
Group 5	0.180	0.018	0.011	0.001	0.049	0.044
Group 6	0.047	0.011	0.003	0.004	0.044	0.030

Table A3: **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.

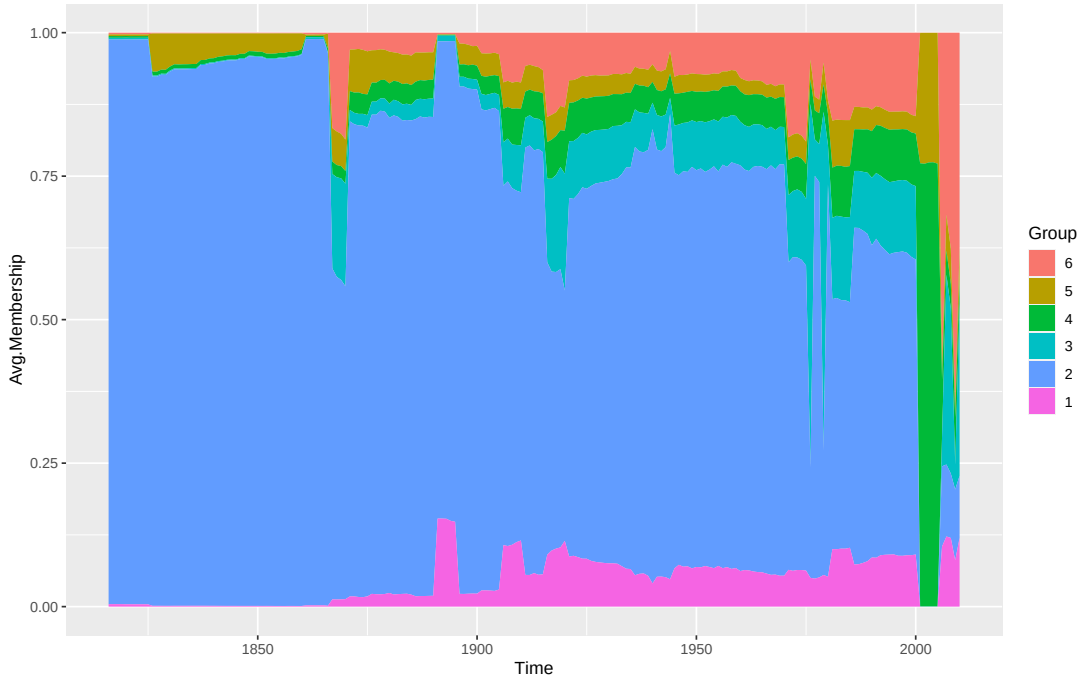

Figure A9: **Membership in Latent Groups over Time (online update model)**. The figure shows the average proportion of membership in six latent groups for each year from 1816–2010. Estimates are derived from specifications using expanding five-year windows. We first fit a model for the years 1816–1820, then use the resulting mixed membership estimates as starting values for the next window (1821-1825). We repeat until all years are included. Membership patterns are positively correlated with the primary model (0.37), but the evolution of membership differs in several ways. Group 2 is significantly larger throughout the period, and the late increases in Group 4 (beginning in 2000) and Group 6 (beginning in 2005) are more pronounced.

Group 1	**Group 2**	**Group 3**
0.158 USA	0.972 Russia	0.888 Costa Rica
0.138 UK	0.969 China	0.88 New Zealand
0.138 Japan	0.827 Germany East	0.878 Ireland
0.137 India	0.81 Poland	0.874 Jamaica
0.132 West Germany	0.766 Czechoslovakia	0.868 Norway
0.115 Italy	0.763 Korea North	0.866 Finland
0.109 France	0.747 Romania	0.863 Denmark
0.100 Canada	0.744 Iran	0.862 Switzerland
0.082 Belgium	0.743 Indonesia	0.859 Luxembourg
0.081 Australia	0.728 Taiwan	0.847 Mauritius
0.08 Netherlands	0.703 Egypt	0.843 Austria
0.069 Turkey	0.69 Saudi Arabia	0.838 Trinidad-Tobago
0.065 Sweden	0.685 Mexico	0.832 Sweden
0.062 South Africa	0.682 Yugoslavia	0.829 Cyprus
0.056 Austria	0.68 Vietnam North	0.827 Israel
Group 4	**Group 5**	**Group 6**
0.319 Yemen	0.188 Djibouti	0.849 Liechtenstein
0.296 Brunei	0.187 Bhutan	0.825 St Kitts-Nevis
0.276 Bahamas	0.173 Guinea-Bissau	0.775 Antigua-Barbuda
0.274 Singapore	0.173 Swaziland	0.747 Vanuatu
0.27 Cambodia	0.155 Comoros	0.737 Dominica
0.269 Angola	0.151 Equatorial Guinea	0.737 St Vincent-Grenadines
0.263 Senegal	0.145 Qatar	0.726 St Lucia
0.261 Mozambique	0.145 Bahrain	0.678 Western Samoa
0.257 Tanzania	0.142 Gabon	0.674 Grenada
0.253 Tunisia	0.136 Cape Verde	0.671 Seychelles
0.251 Namibia	0.134 Malawi	0.65 Belize
0.249 Afghanistan	0.13 Oman	0.644 Sao Tome Principe
0.248 Nepal	0.125 Lesotho	0.618 Maldives
0.244 Ghana	0.122 St Kitts-Nevis	0.516 Barbados
0.243 Kenya	0.117 Mauritania	0.515 Comoros

Table A4: **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.

Group 1	**Group 2**	**Group 3**
0.159 Zanzibar	0.659 Russia	0.771 New Zealand
0.099 Djibouti	0.636 Egypt/UAR	0.715 Solomon Is
0.094 Bhutan	0.55 China	0.71 Ireland
0.092 Cape Verde	0.532 USA	0.704 Papua New Guinea
0.09 Mauritania	0.531 Syria	0.689 Norway
0.089 Central African Rep	0.523 Israel	0.686 Finland
0.089 Niger	0.52 Iran	0.685 Luxembourg
0.089 Burundi	0.514 Iraq	0.684 Denmark
0.089 Mali	0.497 Algeria	0.682 Switzerland
0.089 Nepal	0.483 Namibia	0.646 Malaysia
0.089 Botswana	0.48 Morocco	0.643 Cyprus
0.088 Libyan	0.464 Libyan	0.64 Austria
0.087 Equatorial Guinea	0.46 Brazil	0.626 Australia
0.087 Sri Lanka/Ceylon	0.425 Poland	0.617 Sweden
0.087 Burkina Faso/UV	0.416 North Yemen	0.575 Belgium
Group 4	**Group 5**	**Group 6**
0.525 Yemen	0.47 Vanuatu	0.561 St Lucia
0.47 Bhutan	0.466 Western Samoa	0.561 St Vincent-Grenadines
0.44 Oman	0.381 Brunei Darussalam	0.561 Grenada
0.438 Qatar	0.362 Mongolia	0.56 Dominica
0.437 Pakistan	0.344 Gabon	0.56 Antigua-Barbuda
0.436 Saudi Arabia	0.334 Maldives	0.558 St Kitts-Nevis
0.42 Afghanistan	0.328 Malawi	0.557 Belize
0.416 Bahrain	0.317 Liechtenstein	0.555 Barbados
0.391 Bangladesh	0.315 Cambodia/Kampuchea	0.517 Bahamas
0.335 Iran	0.309 Korea North	0.469 Zanzibar
0.334 Tanzania/Tanganyka	0.309 Cape Verde	0.443 Mauritius
0.326 Trinidad-Tobago	0.301 Cote d'Ivoire	0.431 Panama
0.326 Guatemala	0.301 Guinea	0.406 Malta
0.314 India	0.292 Equatorial Guinea	0.404 Trinidad-Tobago
0.3 Lebanon	0.287 Seychelles	0.397 Iceland

Table A5: **States with Highest Membership in Latent Groups, Cold War period (regional model)**. Average group membership in the years 1950-1990 is reported beside the state name for the top 15 states in each latent group. The group assignments are largely similar to the main specification. Nodes with high membership in Groups 3-6, for example, reflect many of the same countries as in A4. The most notable difference in this specification is the grouping of Western and Eastern bloc countries into a single, highly belligerent latent group (Group 2): Russia, China, and the United States are prominent members of this group, as are France and the United Kingdom (not shown).

Predictor	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
INTERCEPT	10.420	16.239	12.357	11.622	4.457	4.549
	(1.021)	(1.021)	(1.021)	(1.021)	(1.060)	(1.064)
POLITY	−0.005	−0.137	0.209	0.052	−0.201	−0.157
	(0.914)	(0.913)	(0.913)	(0.913)	(1.047)	(1.062)
MILITARY	0.363	1.017	0.237	0.163	−0.443	−0.556
CAPABILITY	(1.063)	(1.062)	(1.062)	(1.061)	(1.062)	(1.064)

N nodes: 216; N dyad-years: 842, 685; N time periods: 195
Lower bound at convergence: −527, 587.7

Table A6: **Estimated Coefficients and their Standard Errors, Markov State 2**. The table shows the estimated coefficients (and standard errors) of the two monadic predictors for each of six latent groups in the second Markov state. The estimated coefficients for cubic splines and indicators for variable missingness are not shown.

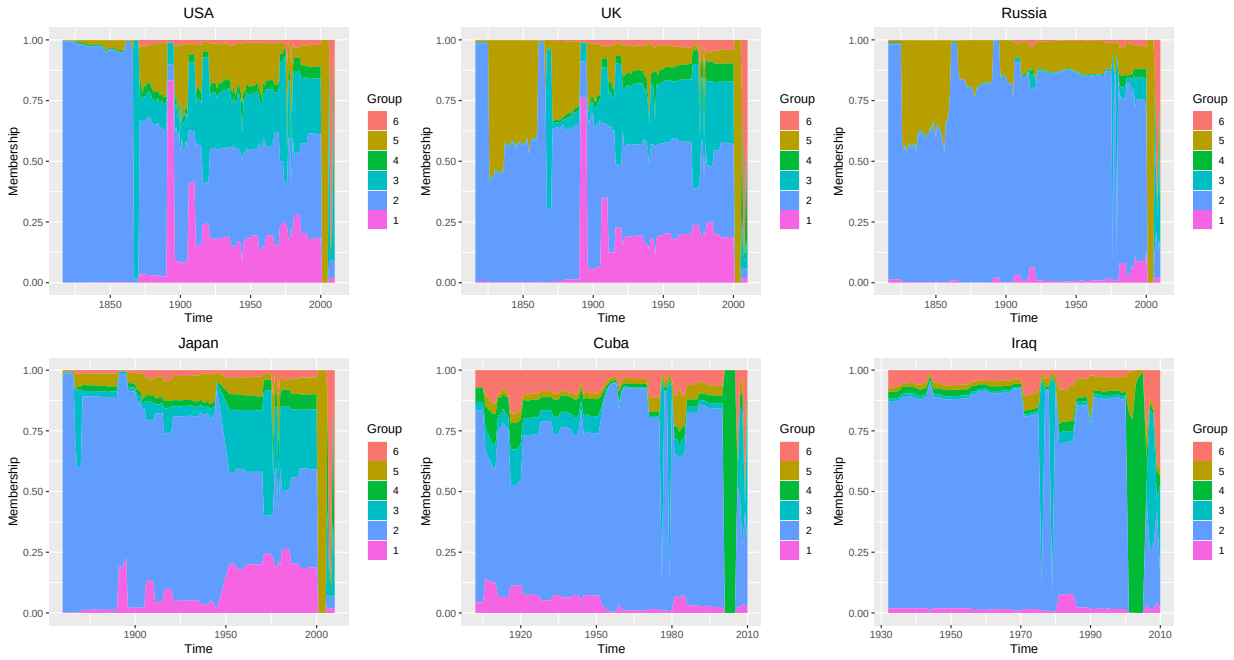

Figure A10: **Average Node Membership over Time, Select States (online update model)**. The figure shows, for six states, the average rate of membership in six latent groups in each year the state is present in the network. Estimates are derived from specifications using expanding five-year windows. We first fit a model for the years 1816–1820, then use the resulting mixed membership estimates as starting values for the next window (1821-1825). We repeat until all years are included. The estimated evolution of state membership is broadly consistent with the primary specification used in Figure 3, with some previously observed structural breaks more pronounced (e.g., Russia at the end of the Cold War, Iraq in 1991) and others attenuated (Japan in 1945) or absent (Cuba in the 1950s).

Predictor	Dyadic	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
INTERCEPT		3.802 (1.066)	6.544 (1.069)	3.531 (1.069)	2.223 (1.069)	5.223 (1.074)	1.804 (1.069)
POLITY		−0.009 (1.079)	−0.104 (1.083)	0.002 (1.083)	0.013 (1.084)	0.0164 (1.096)	0.011 (1.084)
MILITARY CAPABILITY		0.005 (1.059)	0.309 (1.028)	0.021 (1.029)	−0.112 (1.029)	0.175 (1.048)	−0.005 (1.025)
BORDERS	2.453 (0.002)						
DISTANCE	-0.0001 (0.004)						
ALLIANCE	0.084 (0.003)						
IO CO-MEMBERS	0.001 (0.004)						
PEACE YRS	-0.019 (0.004)						
N nodes: 216; N dyad-years: $842,685$; N time periods: 195 Lower bound at convergence: $-155,058.1$							

Table A7: **Estimated Coefficients and their Standard Errors (online update model)**. 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. Estimates are derived from specifications using expanding five-year windows. We first fit a model for the years 1816–1820, then use the resulting mixed membership estimates as starting values for the next window (1821-1825). We repeat until all years are included. We report coefficient estimates for the final model (2006-2010.

A.5 Comparison with Logistic Regression

In this appendix, 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 forecasting 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. First, we conduct a Diebold-Mariano test of comparative forecasting accuracy (Diebold and Mariano, 1995; Harvey et al., 1997). The test, which compares mean-squared forecasting error of the two methods, indicates that the dynMMSBM significantly outperforms the logit model in dispute prediction (DM statistic $= -2.12$, $p = 0.034$).

Second, we compare the receiver operating characteristic curves (ROCs) for each model.

We display the ROC curves in Figure A11 and show the area under the ROC curves in A8. By this criterion, the dynMMSBM continues to outperform the logit model but only marginally. The dynMMSBM has a larger area under the ROC curve, though the difference is not statistically significant.

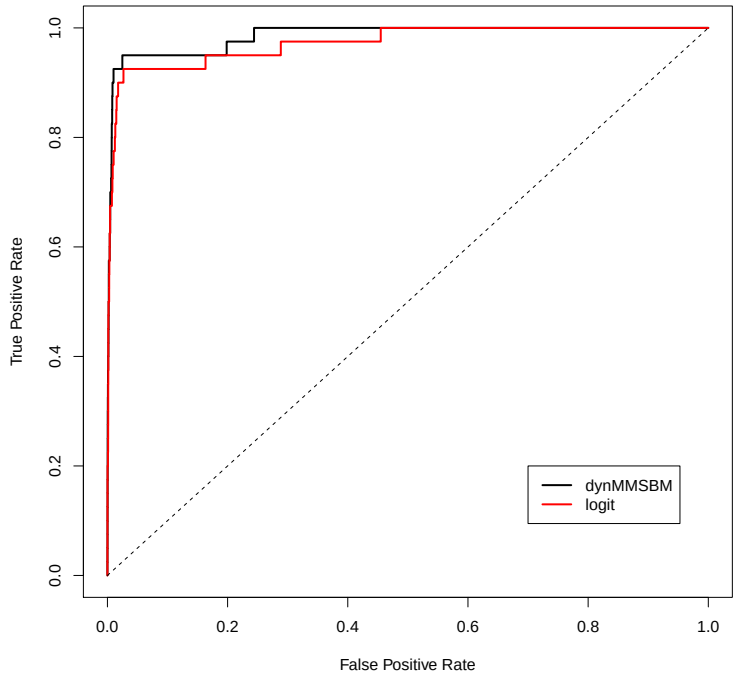

Figure A11: **ROC Curve: Logit, Dynamic Mixed-membership SBM Models**. To perform the forecast, we exclude the final two years (2009-2010) from the dataset and estimate each model on the preceding years (1816-2008). Then we predict the missing years based solely on the covariate data.

Model	AUROC
dynMMSBM	0.986
	(0.013)
Logit	0.973
	(0.018)

Table A8: **Out of Sample Prediction, dynMMSBM vs. Logit**. The table displays the area under the ROC curve (AUROC) and associated standard error for the two models. Each model is fit on data from 1816-2008 and used to forecast conflict in the period 2009-2010.