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

Santiago Olivella† Tyler Pratt‡ Kosuke Imai§

July 12, 2018

Abstract

Many social scientists theorize how various factors influence the dynamic process of network evolution. These theories explain the ways in which nodal and dyadic characteristics play a role in the formation and evolution of relational ties over time. We develop a dynamic model of social networks by combining a Hidden Markov model with a mixed-membership stochastic blockmodel that identifies latent groups underlying the network structure. Unlike existing models, we incorporate covariates that predict both the dynamic changes in the node membership of latent groups and the direct formation of edges between dyads. Our motivating application is the dynamic modeling of international conflicts. While most existing work 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 coalitions. 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, is implemented through an open-source software package.

*The methods described in this paper can be implemented via the open-source statistical software, NetMix, available at `https://github.com/solivella/NetMix`.

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

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

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

1 Introduction

Social scientists often posit theories about the dynamic effects of latent groups of actors on relational outcomes of interest over time. In studying interstate conflict, for example, international relations scholars have examined the so-called "democratic peace" hypothesis, which states that blocs of actors, defined by their underlying disposition toward democratic values, rarely engage in wars amongst themselves (e.g., Maoz and Russett, 1993; Oneal and Russett, 1999). These social science theories often define groups of similar actors that underlie the structures of social networks, and stipulate how the formation and changes of these groups give rise to various actions and behaviors (Lorrain and H. C. White, 1971).

To aid the empirical testing of these theories, we develop a dynamic model of social networks that generalizes the mixed-membership stochastic blockmodel (MMSBM) originally proposed by Airoldi et al. (2008). The MMSBM is a popular generalization of the stochastic blockmodel (SBM; Wang and Wong, 1987; Snijders and Nowicki, 1997), which is a factor analytic model for network data characterized by latent groups of nodes. Unlike the SBM, the MMSBM allows nodes to instantiate different group memberships in their interactions with other nodes. Our proposed dynamic mixed-membership stochastic blockmodel, which we call dynMMSBM, enables the memberships of latent groups to evolve over time via a hidden Markov process while simultaneously incorporating both dyadic and nodal attributes that affect the dynamic formation of groups and ties.

Thus, our approach frees applied researchers from the need to resort to inefficient, two-step procedures to evaluate theories whereby memberships are first estimated, and then regressed on covariates of interest (Wasserman and Faust, 1994). Furthermore, the proposed model allows for prediction of group membership and future network ties of previously unobserved nodes. To facilitate the application of our proposed model, we develop a fast Bayesian inference algorithm by relying on a variational approximation to the collapsed posterior (Teh, Newman, and Welling, 2007). We offer an open-source software package, NetMix (available at `https://github.com/solivella/NetMix`), so that applied researchers can easily implement the proposed model in their own empirical analysis.

Our work builds upon the growing literature on dynamic modeling of social networks that exhibit some degree of stochastic equivalence. In addition to the SBM, a variety of models are generally available to accommodate such networks. For instance, the latent position cluster model (Handcock, Raftery, and Tantrum, 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, Birmelé, Ambroise, et al., 2011; Kim and Leskovec, 2013) or the mixed-membership SBM (MMSBM; Airoldi et al., 2008) fully address this issue by allowing nodes to belong to multiple equivalence classes. Typically, however, these models were limited by the fact that they imposed independence of group memberships over time and across nodes, as well as independence of dyads conditional on the equivalence structure. This made it difficult to accommodate networks that displayed both stochastic equivalence and some degree of heterogeneity across nodes (e.g. networks that had very skewed degree distributions). Subsequent work therefore focused on relaxing some of these independence assumptions. For instance, (e.g. Sweet, Thomas, and Junker, 2014) incorporated dyadic covariates into the MMSBM, thus allowing for connectivity patterns that where not exclusively the result of the equivalence structure. And even more recently, A. White and Murphy, 2016 incorporated node-specific attributes as predictors of the mixed-membership vectors, thus eliminating the assumption that all nodes in an equivalence class where exchageable. We incorporate both of these ideas in our own model, and allow 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, Fu, and Song, 2010; Ho and Xing, 2015; Fan, L. Cao, and Da Xu, 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 applied researchers typically resort to periodizing history into distinct "epochs" in order 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, our model relies on an M-state 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 issues of identification recently raised by Matias and Miele, 2017 that affect some of the earlier dynamic MMSBM specifications.

To the best of our knowledge, then, our model is the first to tackle both the need to incorporate dyadic and nodal attributes and the need to account for temporal dynamics simultaneously, in an effort to yield a tool of maximal use to applied researchers.

The development of our model is motivated by 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 political systems depress the rate of conflict among states. While many scholars provide evidence that democratic dyads are more peaceful than other pairs of states (Maoz and Russett, 1993; Oneal and Russett, 1999; Ray, 1998; Dafoe, 2011), others argue that the democratic peace was limited to a specific era when alliance patterns were highly correlated with democracy (Farber and Gowa, 1997; Gowa, 2011). The attribution of peaceful relations to democracy therefore depends on whether democratic political systems encourage states to enter the same geopolitical coalition — a question our model is designed to address.

When analyzing conflict data, the most common methodological approach is to assume the conditional independence of state dyad-year observations given some covariates within the generalized linear model framework (e.g., Maoz and Russett, 1993; Gleditsch and Hegre, 1997; Farber and Gowa, 1997; Mansfield and Snyder, 2002; Gartzke, 2007; Goldsmith, 2007; Mousseau, 2009; Gowa, 2011; Dafoe, Oneal, and Russett, 2013). Recent analyses, however, have turned to network models to relax this conditional independence assumption. Maoz, Kuperman, et al. (2006), for instance, use a measure of structural equivalence among dyads as a covariate of the standard logistic regression. Hoff and Ward (2004) employ random effects designed to explicitly model network dependence in dyadic data. Ward, Siverson, and X. Cao (2007) apply the latent space model developed by Hoff, Raftery, and Handcock (2002) to international conflict. Cranmer and Desmarais (2011) propose and apply a longitudinal extension of the exponential random graph model (ERGM) to conflict data. We build on this emerging body of scholarship that seeks to model complex dependencies in the conflict network.

Our proposed model offers important advantages over previous analytic strategies for studying the dynamic evolution of international conflicts. First, the proposed dynMMSBM is a natural approximation to the dynamic process of sorting states into geopolitical coalitions that shapes conflict behavior and characterizes distinctive periods in the history of international relations. These evolv-

ing coalitions often interact in unexpected ways that researchers find difficult to specify a priori. For example, although interstate conflicts during the Cold War are characterized by the overarching contest between the Eastern and Western blocs, the leading states in these coalitions rarely engaged directly in conflict. Instead, they clashed in a series of proxy wars among other states. Our model accommodates group-based relationships like these via heterogeneous edge formation among latent groups. We also embed nodal covariates directly into the sorting process, allowing monadic variables like regime type and military power to predict how states shift between coalitions over time. Such an analysis can provide theoretical insights that are difficult to obtain if one separately estimates network dependencies and covariate effects. Finally, the dynamic implementation allows the model to adjust to possibly discontinuous changes in the international system over time.

2 The Interstate Conflict Network

The study of interstate conflict is of great practical and theoretical interest to international relations scholars and policy makers. Practically, the ability to predict violent political clashes has attracted a large and growing literature on conflict forecasting (e.g., Schrodt, 1991; Beck, King, and Zeng, 2000; Ward, Metternich, et al., 2013; Chadefaux, 2014; Hegre et al., 2017). Theoretically, 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 Russett, 1999; Oneal and Tir, 2006; Hegre, 2008; Maoz, 2009).

Empirical studies of interstate conflict are most commonly conducted at the level of the state dyad-year. In these analyses, dyad-year observations are typically assumed to be independent conditional on the covariates included in the model (e.g., Maoz and Russett, 1993; Farber and Gowa, 1997; Mansfield and Snyder, 2002; Goldsmith, 2007; Mousseau, 2009; Gowa, 2011; Dafoe, Oneal, and Russett, 2013). However, there are strong reasons to believe conflict patterns violate this assumption. For centuries, states have actively managed conflict by constructing formal and informal coalitions. Alliances, for example, affect the probability of conflict both among allied states and between allies and non-allies. Many militarized conflicts (most notably, the World Wars) are *multilateral* in nature: states do not decide to engage in conflict as a series of disconnected dyads, but are drawn into war or maintain peace as a result of their membership in preexisting groups. To address this cross-sectional and temporal dependency, we propose a network model of interstate conflict that acknowledges the tendency of states to sort into geopolitical coalitions. In our model, which we call dynMMSBM and describe in detail in Section 3, nodal attributes influence the

formation of unobserved groups, and the effect of these attributes can vary over time, depending on which "epoch" of history the interaction takes place in.

We use the dynMMSBM to examine the onset of militarized disputes among 164 states in the years 1816–2010. To measure conflict, we draw on version 4.1 of the Militarized Interstate Dispute (MID) dataset (Ghosn and Bennett, 2003). A MID occurs when one state enages in an government-sanctioned "threat, display or use of military force" against "the government, official representatives, official forces, property, or territory of another state" (Jones, S. A. Bremer, and Singer, 1996, p. 168). The onset of a MID is a relatively rare event, occuring in approximately 0.5% of the 625,134 state dyad-year observations in our sample.

In defining the structural components of the dynMMSBM, we begin with the standard specification used in the literature on the democratic peace. This research agenda is among the most prominent theoretical debates in the study of interstate conflict, and it explores whether democracies engage in conflict at lower rates than other regime types. While recent network applications have re-examined the democratic peace debate (e.g., Hoff and Ward, 2004; Ward, Siverson, and X. Cao, 2007; Cranmer and Desmarais, 2011), the dynMMSBM offers several distinct advantages. By allowing democracy to shape the formation of latent groups, the proposed model closely mirrors the theoretical mechanism often cited by democratic peace theorists. Specifically, according to the theory, democracies represent a distinct community of states that have achieved a "separate peace" among themselves. The dynMMSBM could reveal such a community by identifying a latent group that exhibits low rates of intra-group conflict and that democratic states are more likely to join. The model also allows for the possibility of a similar "dictatorial peace" among autocratic states, as argued by Peceny, Beer, and Sanchez-Terry (2002). The direct inclusion of nodal variables also obviates the need to restructure monadic covariates to fit a dyadic dataset, which has exacerbated a debate in the democratic peace literature regarding the appropriate dyadic specification of democracy (see Dafoe, Oneal, and Russett, 2013). Finally, the dynamic implementation provides flexibility for the effect of democracy to vary over time, as hypothesized by Farber and Gowa (1997).

3 The Proposed Methodology

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

3.1 The Dynamic Mixed-Membership Stochastic Blockmodel

Suppose that we observe a social network as graph $G_t(V_t, E_t)$ for each time period $t \in \{1, \ldots, T\}$ where V_t and E_t represent a set of nodes and that of directed edges, respectively. We consider the case of undirected networks later in this section. We allow each node set V_t to possibly vary over time since nodes may enter and exit the network at different points in time. For example, in our application, some countries are born into or disappear from the international system during the study period. We use N_t to denote the number of nodes in V_t, i.e. $N_t = |V_t|$.

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

Like the SBM, the MMSBM makes the relational outcomes Y_{pqt} a function of K latent groups to which nodes belong. The key distinctive feature of the MMBSM, however, is that while a node can belong to a latent group when interacting with the other node of a given dyad, different groups may be instantiated by the same node in other relationships and depending on whether it is playing a role of sender or receiver. With the addition of a time dimension, this mixed membership framework can be taken one step further by allowing a node to belong to different latent groups across time periods even when it is interacting with the same node in the same role (either as a sender or a receiver). Formally, we define a K-dimensional indicator vector $\mathbf{z}_{p \to q,t}$ ($\mathbf{w}_{q \leftarrow p,t}$) whose kth element $z_{p \to q,t,k}$ ($w_{q \leftarrow p,t,k}$) is equal to one if node p (q) instantiates group k when interacting with node q (p) as a sender (receiver).

As in the standard MMSBM, the group-by-group propensities of edge formation can be collected in a $K \times K$ matrix $\mathbf{B}$ — the so called *blockmodel*. For the purpose of identification, and unlike other approaches to incorporating a time dimension in the MMSBM (e.g. Xing, Fu, and Song, 2010), we do not allow the blockmodel to vary over time (for a discussion of the problem, see Matias and Miele, 2017). That is, only node memberships into groups are allowed to differ across time periods.

We now turn to the description of the proposed model, dynMMSBM. We begin by modeling the edge indicator Y_{pqt} with a generalized linear model whose linear predictors consist of dyadic time-varying covariates $\mathbf{d}_{pqt}$ as well as a fixed effect specific to the interaction between two groups, to

which node p and q belong. This group interaction fixed effect is represented by the corresponding element of the blockmodel $\mathbf{B}$. Thus, this part of the model is given by,

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

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

In the dynMMSBM, each node has a time-specific probability of instantiating a group in any given interaction. We model these *mixed-membership* probability vectors, denoted $\boldsymbol{\pi}_{pt}$, as a time-specific mixture of M separate Dirichlet distributions with common concentration parameter ξ. We also let the mean of each Dirichlet distribution depend on the set of time-varying nodal covariates $\mathbf{x}_{pt}$ through the Softmax function. This enables researchers to predict group memberships for different nodes. Finally, we model the dynamic dimension of the social network by defining a first-order hidden Markov model with M hidden states for the mixed-membership vectors. Thus, the coefficients of the dyadic covariates in the mean of Dirichlet distribution are allowed to be different, depending on which hidden state the corresponding time period is in.

Formally, we have,

$$\begin{aligned} \mathbf{z}_{p\to q,t} &\sim \text{Multinom}(1,\boldsymbol{\pi}_{pt}) \\ \mathbf{w}_{q\leftarrow p,t} &\sim \text{Multinom}(1,\boldsymbol{\pi}_{qt}) \\ \boldsymbol{\pi}_{pt} &\sim \prod_{m=1}^{M}\left[P(s_{tm})\text{Dirichlet}\left(\text{SoftMax}(\boldsymbol{\beta}_m\mathbf{x}_{pt}),\xi\right)\right]^{s_{tm}} \\ \mathbf{s}_t &\sim \text{Multinom}(\mathbf{A}^{\top}\mathbf{s}_{t-1}) \\ \mathbf{s}_1 &\sim \text{Multinom}(\boldsymbol{\lambda},1) \end{aligned}$$

where $s_{tm}=1$ when time period t is in hidden state m (and zero otherwise), $\boldsymbol{\beta}_m$ is a $K\times J_x$ matrix of state-specific coefficients (with $\boldsymbol{\beta}_1=0$ for identification purposes), $\mathbf{A}$ is an $M\times M$, row-normalized matrix of state-transition probabilities, and $\boldsymbol{\lambda}$ is the M-dimensional vector of prior probabilities over initial states.

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

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

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

Thus, according to this model, the full joint distribution of data $\mathbf{Y} = \{\mathbf{Y}_t\}_{t=1}^T$, latent variables $\mathbf{Z} = \{\mathbf{z}_{p\to q,1}, \ldots, \mathbf{z}_{p\to q,T}\}_{p,q\in\mathbf{V}}$, $\mathbf{W} = \{\mathbf{w}_{q\leftarrow p,1}, \ldots, \mathbf{w}_{q\leftarrow p,T}\}_{p,q\in\mathbf{V}}$, $\mathbf{\Pi} = \{\boldsymbol{\pi}_{p1}, \ldots, \boldsymbol{\pi}_{pT}\}_{p\in\mathbf{V}}$, $\mathbf{S} = \{\mathbf{s}_t\}_{t=1}^T$, and parameters $\{\mathbf{B}, \mathbf{A}, \xi, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ is given by,

$$\begin{aligned}
&P(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{\Pi}, \mathbf{S}, \mathbf{B}, \mathbf{A}, \boldsymbol{\beta} \mid \mathbf{X}, \mathbf{D}) \\
&= P(s_1) \prod_{t=2}^{T} P(\mathbf{s}_t \mid \mathbf{s}_{t-1}, \mathbf{A}) \prod_{m=1}^{M} P(\mathbf{A}_m) \prod_{g=1}^{K} \prod_{h=1}^{K} P(B_{g,h}) \prod_{t=1}^{T} P(\mathbf{Y}_t, \mathbf{Z}_t, \mathbf{W}_t, \mathbf{\Pi}_t \mid \mathbf{s}_t, \mathbf{B}, \boldsymbol{\beta}_m, \boldsymbol{\gamma}, \mathbf{X}_t, \mathbf{D}_t) \\
&\quad \times P(\boldsymbol{\gamma}) P(\xi) \prod_{m=1}^{M} \prod_{k=1}^{K} P(\boldsymbol{\beta}_{km})
\end{aligned}$$

where

$$\begin{aligned}
&P(\mathbf{Y}_t, \mathbf{Z}_t, \mathbf{W}_t, \mathbf{\Pi}_t \mid \mathbf{S}_t, \mathbf{B}, \boldsymbol{\beta}_m, \boldsymbol{\gamma}, \mathbf{X}_t, \mathbf{D}_t) \\
&= \prod_{p,q\in V_t} P(Y_{pqt} \mid \mathbf{z}_{p\to q,t}, \mathbf{w}_{q\leftarrow p,t}, \mathbf{B}, \boldsymbol{\gamma}, \mathbf{D}_t) P(\mathbf{z}_{p\to q,t} \mid \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{q\leftarrow p,t} \mid \boldsymbol{\pi}_{qt}) \prod_{p\in V_t} \prod_{m=1}^{M} P(\boldsymbol{\pi}_{p,t} \mid \mathbf{s}_t, \boldsymbol{\beta}_m, \mathbf{x}_{pt})^{s_{tm}}
\end{aligned}$$

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

3.2 Marginalization

As we discuss in the next section, we define a factorized approximation to the posterior distribution of our model's parameters in order to drastically reduce the computation time needed to

learn about them. With the model as it currently stands, the approximating distribution would factorize over all parameters. In the true posterior, however, latent variables $\mathbf{z}_{p\to q,t}$ ($\mathbf{w}_{q\leftarrow p,t}$) and the mixed-membership parameters $\boldsymbol{\pi}_p$ ($\boldsymbol{\pi}_q$) are usually strongly dependent (Teh, Newman, and Welling, 2007). Similarly, the Markov state indicators $\mathbf{s}_t$ and parameters in the transition kernel $\mathbf{A}$ are normally strongly correlated in the true posterior. Therefore, and to improve the accuracy of the approximation, we first marginalize the latent mixed-membership vectors and the Markov transition probabilities, thus dealing with them exactly.

To do the marginalization, we first focus on the portion of the joint density that involves $\boldsymbol{\Pi}$. Define,

$$\alpha_{ptmk} \;=\; \frac{\xi \exp(\boldsymbol{\beta}_{km}\mathbf{x}_{pt})}{\sum\limits_{k'=1}^{K} \exp(\boldsymbol{\beta}_{k'm}\mathbf{x}_{pt})} \;=\; \xi\mu_{k'm} \tag{1}$$

as the kth element of a K-dimensional vector that serves as the parameter of the Dirichlet distribution from which mixed memberships are drawn. The term μ_{ptkm} is therefore the expected probability that node p instantiates group k at time t under state m, and is such that $\sum_{k=1}^{K}\mu_{ptkm}=1$. In turn, parameter $\xi>0$ controls the concentration of the Dirichlet distribution. Then, we can marginalize $\boldsymbol{\Pi}$ as follows,

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

where $\Gamma(\cdot)$ is the Gamma function, and $C_{ptk}=\sum_{q\in V_t}(z_{p\to q,t,k}+w_{p\leftarrow q,t,k})$ represents the number of times node p instantiates group k across its interactions with all other nodes q present at time t, whether as a sender or as a receiver. Note that we replace $\sum_{k=1}^{K}C_{ptk}$ with $2N_t$ because all nodes must instantiate exactly one group when interacting with other nodes at any given t—once as a sender and once more as a receiver. In the undirected case, this term reduces to N_t (see Appendix A.1).

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

$$\int\cdots\int\prod_{t=2}^{T}P(\mathbf{s}_t\mid\mathbf{s}_{t-1},\mathbf{A})\prod_{m=1}^{M}P(\mathbf{A}_m)\,\mathrm{d}\mathbf{A}_1\cdots\mathrm{d}\mathbf{A}_M \;=\; \prod_{m=1}^{M}\frac{\Gamma(M\gamma)}{\Gamma(M\gamma+U_m)}\prod_{n=1}^{M}\frac{\Gamma(\gamma+U_{mn})}{\Gamma(\gamma)}$$

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

Putting all together, the marginalized posterior distribution is proportional to the following joint density,

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

3.3 Estimation via Variational Approximation

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

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

where $\boldsymbol{\kappa}_t$, $\boldsymbol{\phi}_{p \to q,t}$, and $\boldsymbol{\psi}_{q \leftarrow p,t}$ are variational parameters.

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

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

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

where $C'_{ptk} = C_{ptk} - z_{p \to q,t,k}$ and the expectation is taken over the variational distribution of $\mathbf{Z}$. This corresponds to the (unnormalized) probability vector in a multinomial distribution. By symmetry, the update for $\psi_{q \leftarrow p,t,k}$ is similarly defined, and the two updates associated with a dyad can be computed in parallel to speed up computation. Also in the interest of speed, and in order to avoid costly computation of the Poisson-Bernoulli probability mass function, we approximate the expections in these updates by using a zeroth-order Taylor series expansion, so that $\mathbb{E}_{\widetilde{Q}_2}[\log(\alpha_{ptmk} + C'_{ptk})] \approx \log\left(\alpha_{ptmk} + \mathbb{E}_{\widetilde{Q}_2}\left[C'_{ptk} \right] \right)$ (Asuncion et al., 2009).

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

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

where $U'_m = U_m - s_{t,m}$ and $U'_{mn} = U_{mn} - s_{tm}s_{t+1,n}$.[1] Once again, this corresponds to the (unnormalized) probability vector in a multinomial distribution. Finally, note that there are special cases for updates at the first and last time periods, i.e., $t = 1$ and $t = T$. Their details are available in Appendix A.2.

Finally, to obtain the estimates of the regression parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and the blockmodel $\mathbf{B}$, as well as estimates of the concentration parameter ξ, we find optimal values with respect to the approximate lower bound, defined as the log expectation of equation (3) over the variational distribution. The resulting product-of-multinomials form of $\widetilde{Q}$ (which relied only on a factorizing assumption), allows us to compute the necessary expectations. To find optimal values, we use an iterative quasi-Newton algorithm, and provide the gradients required for this step in Appendix A.2.

4 Empirical Analysis

We apply the proposed dynMMSBM to the interstate conflict network data described in Section 2. We use the open-source statistical software NetMix to fit the model. As shown below, the dynMMSBM recovers the essential geopolitical coalitions that drive conflict patterns and generates novel insights into the heterogeneous effect of key covariates, like democracy. It also outperforms the standard logistic regression model in forecasting future conflicts.

4.1 Specification

We model conflict as an undirected network in which ties arise from states' evolving membership in four latent groups. While the substantive results presented below are not conditional on the number of latent groups, we found that four provided sufficient flexibility to reflect different types of coalitions that can be qualitatively interpreted while avoiding near-empty groups. We include in $\mathbf{x}_t$ two node-level covariates, the degree of democracy in a state's domestic government and the

[1] This definition of the term U'_{mn} is valid whenever $m \neq n$ and $t \neq T$. For other cases, see Appendix A.2.

state's military capability, that are hypothesized to influence membership in these groups (Maoz and Russett, 1993; Hegre, 2008). The degree to which these covariates are associated with latent group memberships depends on two hidden Markov states. In addition, we include four dyadic predictors — geographic distance, the presence of a shared border, alliance ties, and the existence of a defense pact — that are expected to directly affect the probability of conflict among state dyads (Gleditsch, 1995; Huth, 2009; Leeds, 2003). These variables form $\mathbf{d}_t$ in the model.

We measure levels of democracy using the variable `POLITY`, from the Polity IV dataset (Marshall, Gurr, and Jaggers, 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 -0.05. Five percent of state years are assigned the minimum score of -10, and 16.6% receive the maximum of 10. We also include a measure of states' national military capability (`MILITARY CAPABILITY`) as a monadic covariate. We use version 5.0 of the composite index (CINC scores) originally developed by Singer, S. Bremer, and Stuckey (1972), and take the log to mitigate the skewed distribution. These variables enter the model as monadic predictors that influence the process of group formation.

Our analysis further incorporates several dyadic variables that are expected to be predictive of conflict between states beyond the effects of the equivalence classes induced by the blockmodel. These include dichotomous indicators for the existence of a formal alliance (`ALLIANCE`) and defense pact (`DEFENSE PACT`) between states in a given year; data comes from version 4.1 of the COW Formal Alliances dataset (Gibler, 2009). In addition, we control for geographic distance (`DISTANCE`) and the presence of a contiguous border (`BORDER`) between states (Stinnett et al., 2002). Following the convention in the literature, a count of years since the last militarized dispute between each dyad and a cubic spline control for temporal dependencies (Beck, Katz, and Tucker, 1998).

4.2 Results

4.2.1 Conflict propensities of latent groups

Figure 1 graphically summarizes the resulting blockmodel where the width of the edge is proportional to the estimated probability of conflict between groups. If an edge originates from one node and ends up in the same node, it represents the estimated probability of conflict with other countries in the same group. The exact estimates of the blockmodel, on which this figure is based, are available in Table 2 of Appendix A.4. In the figure, the node size is proportional to the estimated frequency with which states instantiate membership in each group, $\sum_{t=1}^{T} \sum_{p \in V_t} \boldsymbol{\pi}_{pt,g}$.

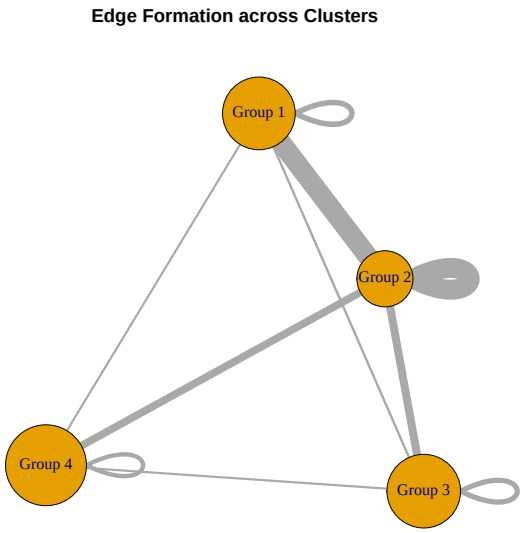

Figure 1: **Network Graph of Edge Formation Probabilities**. The nodes (circles) in the figure represent the four latent groups. Node size reflects the frequency with which states instantiate membership in each group. Weighted edges (lines) represent the probability of conflict between groups.

The results reveal that a single latent group, Group 2, is responsible for the predominant share of conflict among states. States in Group 2 engage each other in militarized disputes at a very high rate (probability 0.594 on average), and these states also frequently enter conflicts with other groups. Examples of node-years with the highest membership in Group 2 are the United States in 1938 and the Soviet Union in 1937 — two states poised to engage in a large number of MIDs at the onset of World War II. Group 2 also has the smallest estimated membership; nodes instantiate membership in this group 19.8% of the time. This is substantially lower than the corresponding rates of the other blocks: for Groups 1, 3, and 4, the rates are 25.7%, 26%, and 28.5%, respectively. Unlike Group 2, the other three groups are comparatively peaceful. States in these groups rarely fight among themselves (see diagonal estimates of Table 2) or with each other (off-diagonal). Group 4 has the lowest probability of intra-group conflict, at 0.02%. As we discuss below, states with democratic political systems are more likely to instantiate Group 4 than any other group.

Figure 2 shows the relationship between membership in the most conflictual group (Group 2) and overall conflict rates over time. The black line depicts the estimated average probability of membership in Group 2 for each year. There is significant variation over time. Group 2 membership spikes in the run up to WWI and WWII, corresponding to sharp increases in the rate of militarized dispute onset (red line). Membership in this group declines in the post-war period, which is

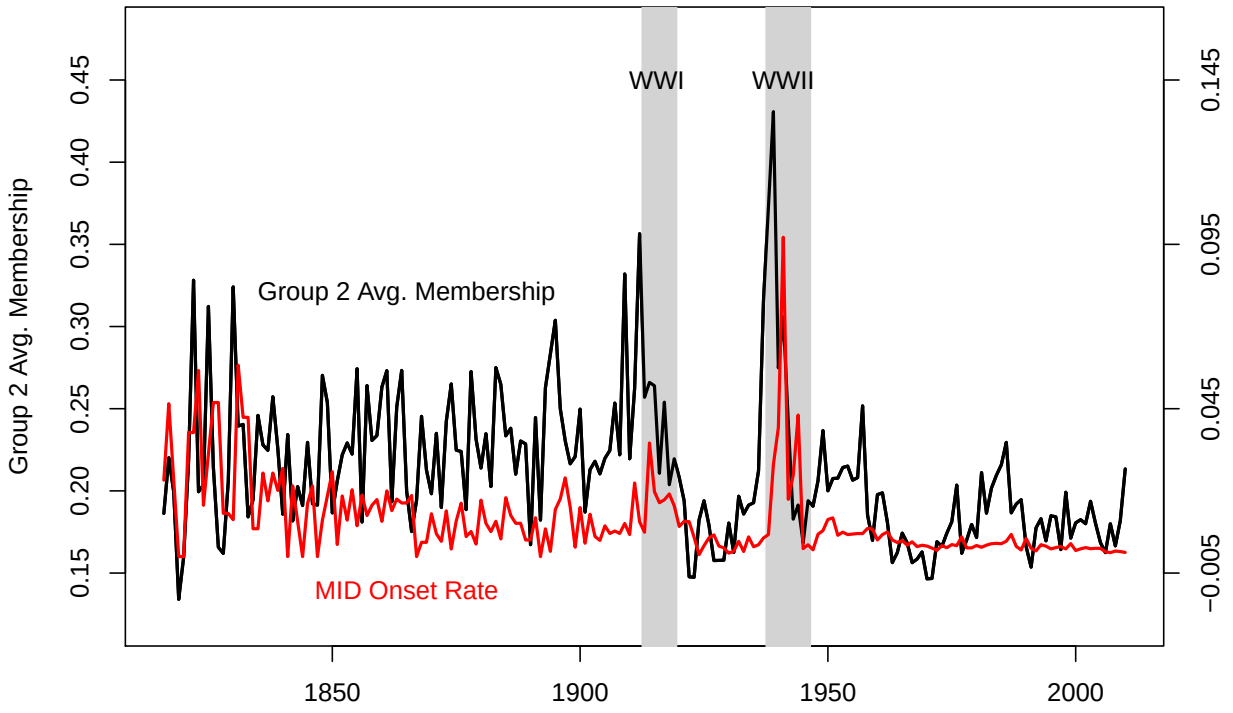

Figure 2: **Group 2 Membership vs. Conflict Rate**. The black line depicts the average probability of state membership in Group 2 over time. The red line shows the observed rate of conflict in each year. The World Wars are depicted via gray bars.

characterized by low rates of conflict.

4.2.2 Memberships of latent groups

The dynMMSBM allows us to examine latent group membership at the node, year, and node-year level. Figure 3 displays average group membership by year for a select group of states. As the figure demonstrates, group membership can shift markedly from one year to the next, particularly in response to instances of conflict. The Russian Federation, for example, shows notable increases in Group 2 membership during the outbreak of World War I and World War II, when the country experienced MIDs with a large number of adversaries.

Differences across states are also apparent in Figure 3. Costa Rica and Switzerland, states that have largely avoided conflict by adhering to a foreign policy of neutrality, are less prone to membership in Group 2 compared to great powers like the United States, Russia, and China. The liberal democracies pictured (the United States, Switzerland, and Costa Rica) tend to have higher membership in Group 4, while autocratic states like Russia and China tend to instantiate membership in Group 1.

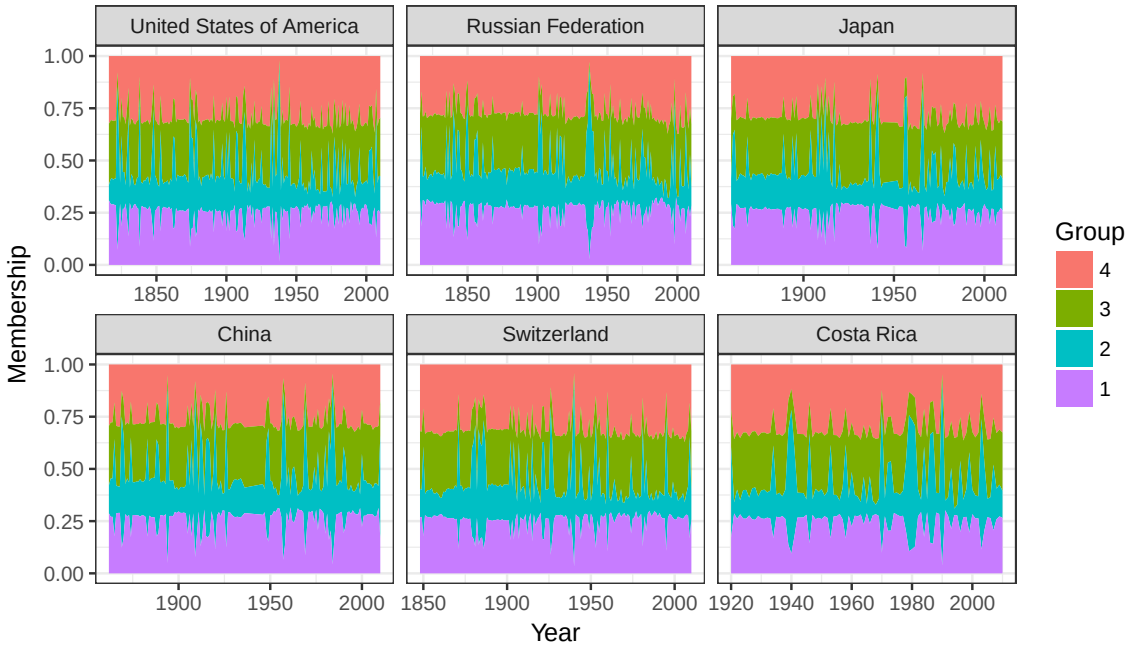

Figure 3: **Average Node Membership over Time, Select States**. The figure shows, for six states, the average rate of membership in four latent groups in each year the state is present in the network.

To probe the plausibility of the estimated group assignments, we examine latent group membership during the Cold War. 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. While these dueling coalitions did not frequently engage in direct conflict, they are believed to have shaped conflict patterns for much of the post-War era. To see if the dynMMSBM can recover the underlying geopolitical structure of the Cold War era, we identify the 25 states with the highest average membership probability in each latent group during the period of 1955–1990. For each latent group g, we calculate the following quantity for every state p: $\frac{1}{35}\sum_{t=1955}^{1990}\pi_{ptg}$. The states with the highest membership in each latent group are listed in Table 3 of Appendix A.4.

The distribution of states across the latent groups is consistent with presence of competing geopolitical coalitions during the Cold War. Group 1 features several members of the Warsaw Pact (Russia/Soviet Union, Hungary, and Romania), the defining alliance in the Eastern bloc. Soviet-leaning states such as North Korea, Syria, and Algeria also had higher membership in this group. Group 4, on the other hand, is populated largely by the United States, its NATO allies (Australia, Germany, Denmark, Belgium, Italy, New Zealand), and several of the neutral European states that leaned toward the Western bloc (Finland, Austria, Sweden).

Monadic Variable	Group 1	Group 2	Group 3	Group 4
INTERCEPT	0.000	-0.100	0.018	0.080
POLITY	0.000	0.002	0.004	0.010
MILITARY CAPABILITY	0.000	0.014	0.000	-0.005
Dyadic Variable	**Coefficient**			
DISTANCE	-0.169			
BORDERS	0.549			
ALLIANCE	0.833			
DEFENSE PACT	0.613			
PEACE YRS	-0.114			

Table 1: **Estimated Coefficients, Interstate Conflict Network**. The table shows the coefficients associated with the two monadic predictors for each of four latent groups, as well as the coefficients for the dyadic predictors. Cubic splines not shown.

The composition of these two groups, in addition to the blockmodel shown in Figure 1, provides further evidence that the model is recovering the essential conflict patterns of the Cold War era. States in Groups 1 (Eastern Bloc) and 4 (Western Bloc) have a low probability of conflict with each other. Both, however, tend to engage in conflict with Group 2 at high rates. Glancing at the composition of Group 2 (the second column of Table 3), we see many of the contested areas that experienced "proxy wars" between the two major geopolitical coalitions. Afghanistan, Cambodia, Chile, Yugoslavia, Bangladesh, Chad, Equatorial Guinea were violent flashpoints in the Cold War as the Eastern and Western blocs competed for influence. The elevated rate of between Group 2 states and those in Groups 1 and 4 reflects intervention by the Eastern and Western bloc, respectively, in these states.

4.2.3 Other characterization of latent groups

Examining the covariate relations can also helps characterize the nature of each latent group. 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 relative to Group 1. We display the coefficients only for Markov state 1, since almost the entire time period, i.e., 97.2%, is estimated to derive from this state. Compared to the baseline (Group 1), more democratic states (i.e., those with high `POLITY` scores) are most likely to instantiate membership in Group 4. This is consistent with the interpretation of Group 4 as the Western alliance of liberal democracies during the Cold War. The positive coefficient on `POLITY` also offers some evidence in support of the democratic peace theory: democracies are more likely to join Group 4, which has the lowest rate of intra-group conflict of any of the latent groups. This pattern

is consistent with a separate peace among democracies. Autocratic states, on the other hand, tend to sort into Group 1. Greater military capability is positively associated with membership in Group 2 and negatively associated with membership in Group 4.

In addition to the coefficients, we can also predict how the probability of edge formation changes as a node's monadic covariates shift. In the generative process of the model, group memberships are instantiated for each dyad in each time period. As a result, states in the conflict network are assigned a latent group each time they interact with another state in each 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 for each dyad-year.

For example, consider the prediction of conflict propensity when each node's `Polity` score is greater by one standard deviation (7.14) than what is actually observed.[2] The overall average effect of this change on the probability of edge formation, calculated as

$$\frac{1}{T}\sum_{t=1}^{T}\frac{1}{|V_t \times V_t|}\sum_{p,q\in V_t}[\,\mathbb{E}[Y_{pqt}|\texttt{POLITY}+7.14]-\mathbb{E}[Y_{pqt}]]$$

is positive but negligible in size (0.0009). However, there is significant heterogeneity in the effect across states and over time. Figure 4 shows, for each state, the average difference in the probability of interstate conflict due to an increase in `POLITY` score. Some states such as Poland, Ethiopia, and Cuba are predicted to be more peaceful, on average, if they are more democratic. Others, however, actually are estimated to be more conflict-prone in this scenario. These include some of the most powerful states in the system (e.g., Germany, Russia, and the United Kingdom).

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 other nodes in the network. Figure 5 looks within states to gauge the effect of the shift in `POLITY` over time, revealing additional heterogeneity. The figure also reveals a clear trend over time that shows attenuation of the effect of democracy on the average probability of a tie.

4.2.4 Dyadic covariates

Dyadic predictors operate outside the latent group membership structure, directly influencing the probability of conflict among states. The dyadic coefficient estimates are shown in Table 1. Consistent with existing work, greater geographic distance between states tends to depress conflict, as

[2] During this exercise, we allow POLITY scores to increase up to the maximum value (10).

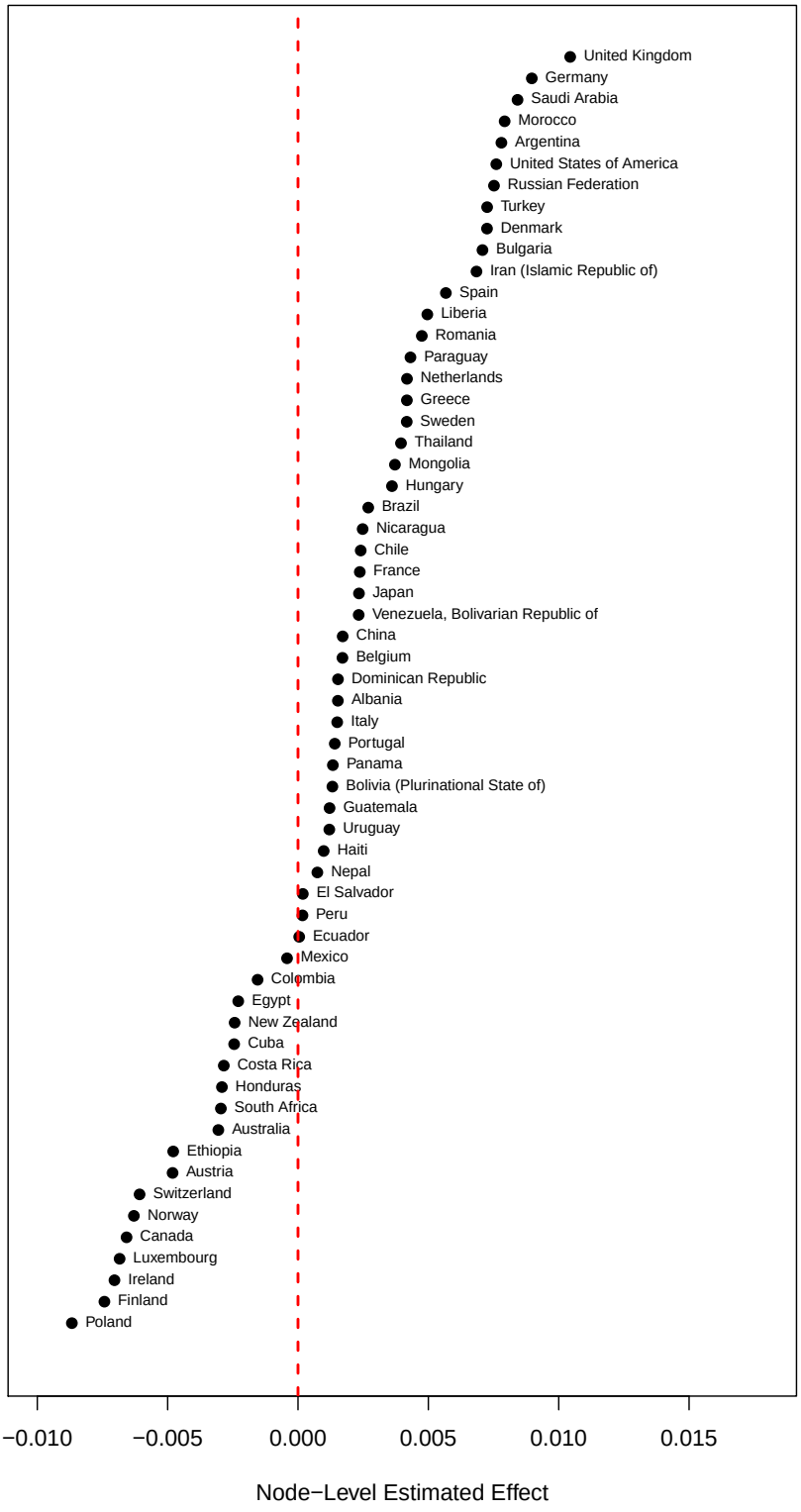

Figure 4: **Effect of Shift in Polity by State.** The figure shows the estimated change in the probability of interstate conflict for each state when its POLITY score is increased by one standard deviation (7.14) from its observed value.

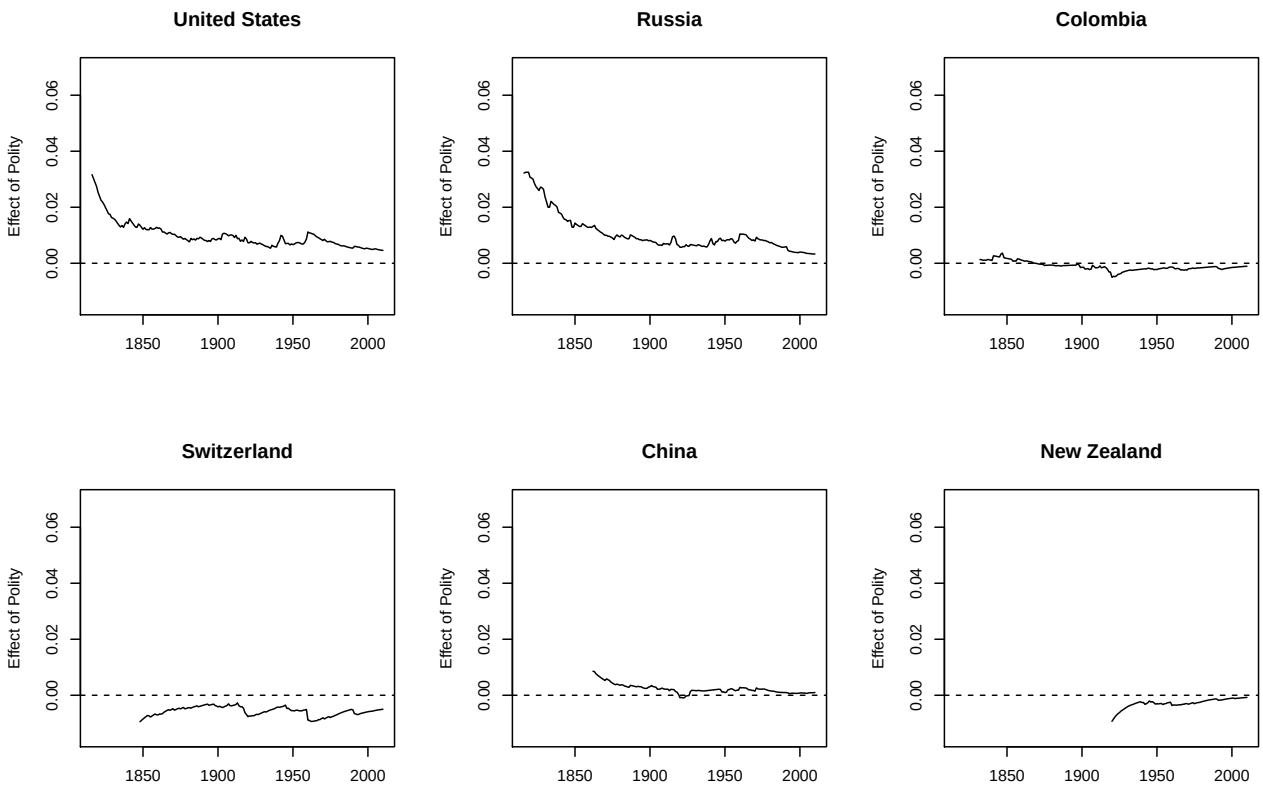

Figure 5: **Effect of Shift in Polity over Time, Select States**. The figure shows the estimated change in the probability of interstate conflict over time if a state's POLITY score is increased by one standard deviation (7.14) from its observed value.

does the length of time since two states have engaged in a militarized dispute. Sharing a border increase the likelihood of conflict. Somewhat surprisingly, militarized disputes are more likely among states that share an alliance or defense pact.

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 "epochs" or periods in time when a network exhibits distinctive patterns of interactions among actors.

Using a network defined by 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 typical expectations — with liberal democracies democracies aligned in one bloc, and more authoritarian regimes aligned in another. The dynMMSBM also reveals more nuanced structures, like blocs of states where proxy conflicts were fought during the Cold War period.

The main goal of this paper is to provide applied researchers with a model that can accommodate a variety of theorized relationships for dynamic network outcomes that display some form of stochastic equivalence. To this end, we make available an open-source software package that implements the dynMMSBM. In future, we plan to further extend the dynMMSBM'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.

A Appendix

A.1 Marginalizing the membership vectors and the transition probabilities

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

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

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,

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

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 nomalizing constant,

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

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 yields equation (2) in Section 3.2.

A.2 Details of the Variational EM Algorithm

A.2.1 E-step

E-step 1: Z and W

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

$$
\log P(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{S}, \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{X}, \mathbf{D})
$$

$$= z_{p\to q,t,k} \sum_{g=1}^{K} w_{q\leftarrow p,t,g} \left\{ Y_{pqt} \log(\theta_{pqtkh}) + (1 - Y_{pqt}) \log(1 - \theta_{pqtkh}) \right\}$$
$$+ \sum_{m=1}^{M} s_{tm} \log \Gamma(\alpha_{ptmk} + C_{ptg}) + \text{const.}$$

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 $\log \Gamma(\alpha_{ptmk} + C_{ptk}) = z_{p\to q,t,k} \log(\alpha_{ptmk} + C'_{ptk}) + \log \Gamma(\alpha_{ptmk})$ and thus simplify the expression to,

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

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

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

The exponential of this expression corresponds to the 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 $\mathbf{w}_{q\leftarrow q,t}$ is similarly derived.

E-step 2: S

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

$$P(\mathbf{Y}, \mathbf{Z}, \mathbf{W}, \mathbf{s}, \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{D}, \mathbf{X}) = \Gamma(M\eta + U_m)^{-1} \prod_{n=1}^{M} \Gamma(\eta + U_{mn}) \prod_{m' \neq m}^{M} \Gamma(\eta + U_{m'm})$$
$$\times \prod_{p \in V_t} \left[\frac{\Gamma(\xi)}{\Gamma(\xi + 2N_t)} \prod_{k=1}^{K} \frac{\Gamma(\alpha_{ptmk} + C_{ptk})}{\Gamma(\alpha_{ptmk})} \right]^{s_{tm}} + \text{const.}$$

To isolate terms that depend on s_{tm} for specific t and m, first define $U'_m = U_m - s_{tm}$ and

$$U'_{mn} = \begin{cases} U_{mn} - s_{tm} s_{t+1,n} & \text{if } m \neq n \text{ and } t \neq T, \text{ or } m = n \text{ and } t = 1 \\ U_{mn} & \text{if } m \neq n \text{ and } t = T \\ U_{mn} - s_{tm} s_{t+1,n} - s_{t-1,m} s_{tn} & \text{if } m = n \text{ and } 1 < t < T \\ U_{mn} - s_{t-1,m} s_{tn} & \text{if } m = n \text{ and } t = T \end{cases}$$

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

$$
\begin{aligned}
&P(\mathbf{Y}, \mathbf{Z}, \mathbf{S}, \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{D}, \mathbf{X}) \\
&= (M\eta + U'_m)^{-s_{tm}}\left\{(\eta + U'_{mm} + 1)^{s_{t+1,m}s_{t-1,m}}(\eta + U'_{mm})^{s_{t-1,m} - s_{t-1,m}s_{t+1,m} + s_{t+1,m}}\right\}^{s_{tm}} \\
&\quad\times \prod_{n\neq m}^{M}(\eta + U'_{mn})^{s_{t+1,n}s_{tm}}(\eta + U'_{nm})^{s_{tm}s_{t-1,n}} \prod_{p\in V_t}\left[\frac{\Gamma(\xi)}{\Gamma(\xi + 2N_t)}\prod_{k=1}^{K}\frac{\Gamma(\alpha_{ptmk} + C_{ptk})}{\Gamma(\alpha_{ptmk})}\right]^{s_{tm}} + \text{const.}
\end{aligned}
$$

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

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

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

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

which has to be normalized so that it sums to one. When $t = T$, the term simplifies to,

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

Finally, when $t = 1$, the term is given by,

$$\hat{\kappa}_{1m} \ \propto\ \exp\left[-\mathbb{E}_{\tilde{Q}_1}[\log(M\eta + U'_m)]\right] \prod_{n=1}^{M} \exp\left[\kappa_{2n}\ \mathbb{E}_{\tilde{Q}_1}[\log(\eta + U'_{mn})]\right]$$
$$\times \prod_{p \in V_2} \left[\frac{\Gamma(\xi)}{\Gamma(\xi + 2N_1)} \prod_{k=1}^{K} \frac{\mathbb{E}_{\widetilde{Q}_1}[\Gamma(\alpha_{p1mk} + C_{p1k})]}{\Gamma(\alpha_{p1mk})}\right]$$

A.2.2 M-step

Lower Bound

We first provide the expression for the lower bound,

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

M-step 1: update for B

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

$$\mathcal{L}(\widetilde{Q}) \ = \ \sum_{t=1}^{T}\sum_{p \in V_t}\sum_{q \in V_t}\sum_{g,h=1}^{K} \phi_{p\to q,t,g}\psi_{q\leftarrow p,t,h}\{Y_{pqt}\log\theta_{pqtgh} + (1 - Y_{pqt})\log(1 - \theta_{pqtgh})\}$$

$$-\sum_{g=1}^{K}\sum_{h=1}^{K}\frac{(B_{gh}-\mu_{gh})^2}{2\sigma_{gh}^2}+\text{const.}$$

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

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

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

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

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

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

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

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

Recall that $\alpha_{ptmk}=\xi\mu_{ptkm}$, where the mean $\mu_{ptkm}=\frac{\exp(\mathbf{x}_{pt}\boldsymbol{\beta}_{km})}{\sum_{k'}\exp(\mathbf{x}_{pt}\boldsymbol{\beta}_{k',m})}$ and $\xi>0$ is a concentration parameter. To find the optimal value of $\boldsymbol{\beta}_{mk}$, we first impose $\boldsymbol{\beta}_{1m}\equiv\mathbf{0}\;\forall m$ for identification purposes and restrict the lower bound to terms involving the remaining $\boldsymbol{\beta}_{mk}$:

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

which we can optimize with respect to β_{mkj} — the jth element of the $\boldsymbol{\beta}_{mk}$ coefficient vector. No closed form solution exists for the maximum of $\mathcal{L}(\tilde{Q})_{\boldsymbol{\beta}_m}$, but a gradient-based algorithm can be implemented to maximize it. The corresponding gradient with respect to each element of $\boldsymbol{\beta}_{mk}$ is given by,

$$\frac{\partial\mathcal{L}(\tilde{Q})}{\partial\beta_{mkj}} \;=\; \sum_{t=1}^{T}\kappa_{tm}\sum_{p\in V_t}\frac{\alpha_{ptmk}x_{ptr_1}}{\sum_{k'=1}^{K}\exp(\mathbf{x}_{pt}\boldsymbol{\beta}_{k'm})}$$

$$\times \sum_{g \neq k} \exp(\mathbf{x}_{pt}\boldsymbol{\beta}_{mg}) \Big(\breve{\psi}(\alpha_{ptmg}) - \mathbb{E}_{\widetilde{Q}_2}[\breve{\psi}(\alpha_{ptmg} + C_{ptg})] + \mathbb{E}_{\widetilde{Q}_2}[\breve{\psi}(\alpha_{ptmk} + C_{ptk})] - \breve{\psi}(\alpha_{ptmk}) \Big) - \frac{\beta_{mkj} - \mu_\beta}{\sigma_\beta^2}$$

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.

M-step 4: update for $\boldsymbol{\xi}$

Finally, the update for concentration parameter ξ is given by an optimization of the lower bound,

$$\begin{aligned}\mathcal{L}(\widetilde{Q}) \ = \ & \sum_{t=1}^{T}\sum_{m=1}^{M} \kappa_{t,m} \sum_{p \in V_t} \log\Gamma(\xi) - \sum_{t=1}^{T}\sum_{m=1}^{M} \kappa_{tm} \sum_{p \in V_t} \log\Gamma(\xi + 2N_t) \\ & + \sum_{t=1}^{T}\sum_{m=1}^{M} \kappa_{tm} \sum_{p \in V_t} \sum_{k=1}^{K} \mathbb{E}_{\tilde{Q}_2}[\log\Gamma(\alpha_{ptmk} + C_{ptk})] - \log\Gamma(\alpha_{ptmk}) - \frac{(\xi - \mu_\xi)^2}{2\sigma_\xi^2} + \text{const.}\end{aligned}$$

using the corresponding gradient, defined by

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \xi} = & \sum_{t=1}^{T}\sum_{m=1}^{M} \kappa_{tm} \Big\{ \breve{\psi}(\xi) - \breve{\psi}(\xi + 2N_t) \Big\} \\ & + \sum_{t=1}^{T}\sum_{m=1}^{M} \kappa_{tm} \sum_{p \in V_t} \sum_{k=1}^{K} \frac{\exp(\mathbf{x}_{pt}\boldsymbol{\beta}_{k,m})}{\sum_{k'=1}^{K} \exp(\mathbf{x}_{pt}\boldsymbol{\beta}_{k',m})} \Big\{ \mathbb{E}_{\tilde{Q}_2}[\breve{\psi}(\alpha_{ptmk} - C_{ptk})] - \breve{\psi}(\alpha_{ptmk}) \Big\}\end{aligned}$$

A.3 Initial values for ϕ and ψ

Implementation of the dynamic model requires defining good starting values for the mixed-membership vectors. While k-means offers good starting values for ϕ and ψ in the non-dynamic setting, applying it to sociomatrices at different time points introduces a type of identification problem commonly known as *label switching*: $P(Y_{pqt} \mid z_{p \to q,t,g}, w_{q \leftarrow p,t,h}) = P(Y_{pqt} \mid z_{p \to q,t,\tau_g}, w_{q \leftarrow p,t,\tau_h})$, where τ is a permutation of $1, \ldots, K$. When applying k-means to each time period, label switching forces us to ask: is a node assigned to a different group at two different times because its membership changes, or because the algorithm settled on an alternative permutation of the group labels? Answering this question requires resolving the label ambiguity problem.

To do so, define a simple approximation to the blockmodel $\mathbf{B}$ at a given time period t,

$$\tilde{B}_{gh}^{\tau_t} \ = \ \frac{\sum_{p \in V_t} \sum_{q \in V_t} Y_{pqt} \phi_{p \to q,t,g} \psi_{q \leftarrow p,t,h}}{\sum_{p \in V_t} \sum_{q \in V_t} \phi_{p \to q,t,g} \psi_{q \leftarrow p,t,h}}$$

which depends on group assignments, and is therefore permutation-specific. Since we have assumed that the blockmodel $\mathbf{B}$ is time-invariant, these approximations should coincide, in expectation, across time periods — provided their labels are not scrambled. Accordingly, we can unscramble permuted approximations to the blockmodel by finding, for each time period, the permutation that minimizes the difference between each element of $\tilde{\mathbf{B}}^{\tau_t}$ and the corresponding element of $\mathbf{B}$.

For an initial (potentially scrambled) set of node allocations to groups in each time period t, Stephens (2000) proposes the following EM-style iterative algorithm to solve the above optimization problem:

1. Find the B'_{gh} that minimize $\sum_{t=1}^{T}\sum_{g=1}^{K}\sum_{h=1}^{K}\tilde{B}_{gh}^{\tau_t}\log\left(\frac{\tilde{B}_{gh}^{\tau_t}}{B'_{gh}}\right)$
2. For $t = 1, \ldots, T$, find τ_t that minimizes $\sum_{g=1}^{K}\sum_{h=1}^{K}\tilde{B}_{gh}^{\tau_t}\log\left(\frac{\tilde{B}_{gh}^{\tau_t}}{B'_{gh}}\right)$

repeating steps 1 and 2 until convergence.

The argmin of step 1 is given by $B'_{gh} = \frac{1}{T}B_{gh}^{\tau_t}$ — the average blockmodel across time periods. While the solution for step 2 can be found by examining all $K!$ full permutations, a more efficient approach takes advantage of the fact that step 2 is equivalent to the assignment problem in combinatorial optimization, and uses integer programming techniques (viz. the so-called Hungarian algorithm) to find a solution without having to test all permutations (see Stephens, 2000, for details).

A.4 Additional Empirical Results

	Group 1	Group 2	Group 3	Group 4
Group 1	0.0003	0.0010	0.0001	0.0001
Group 2	0.0010	0.5940	0.0004	0.0004
Group 3	0.0001	0.0004	0.0003	0.0001
Group 4	0.0001	0.0004	0.0001	0.0002

Table 2: **Group-Level Edge Formation Probabilities**. The table displays the probability of interstate conflict between nodes that instantiate membership in each of four latent groups. The diagonal shows rates of intra-group conflict and off-diagonal shows rates of conflict between groups.

Group 1	**Group 2**	**Group 3**	**Group 4**
0.296 Gabon	0.318 Qatar	0.302 Germany	0.326 Finland
0.293 Syria	0.301 Comoros	0.299 Namibia	0.323 Jamaica
0.288 UAE	0.298 Chile	0.296 Yemen	0.322 Mauritius
0.288 Togo	0.287 Papua New Guinea	0.292 Gabon	0.320 Switzerland
0.287 Ghana	0.263 China	0.290 Syria	0.319 Australia
0.287 Ethiopia	0.259 Fiji	0.286 Ghana	0.318 Germany
0.287 Guinea Bissau	0.255 Mexico	0.284 Togo	0.316 Trinidad-Tobago
0.284 Romania	0.252 Chad	0.284 Ethiopia	0.315 Namibia
0.283 Malawi	0.249 Bangladesh	0.283 Laos	0.315 Gambia
0.282 Guinea	0.248 Norway	0.283 Pakistan	0.314 Denmark
0.282 Bhutan	0.247 Costa Rica	0.283 Myanmar	0.311 Austria
0.282 Angola	0.244 Iran	0.281 UAE	0.310 Botswana
0.282 Myanmar	0.242 Lesotho	0.280 Zimbabwe	0.310 Sweden
0.282 Djibouti	0.236 Cambodia	0.280 Dominican Repub	0.309 Sri Lanka
0.280 Taiwan	0.235 Yugoslavia	0.280 Finland	0.309 Zimbabwe
0.290 Pakistan	0.234 Peru	0.280 Romania	0.307 United States
0.280 North Korea	0.231 Saudi Arabia	0.279 Angola	0.307 Belgium
0.280 Laos	0.230 Madagascar	0.279 Taiwan	0.307 Gabon
0.278 Algeria	0.227 Bahrain	0.279 Guinea Bissau	0.306 Dominican Repub
0.278 Haiti	0.226 Equatorial Guinea	0.278 Guinea	0.306 Pakistan
0.278 Hungary	0.222 Afghanistan	0.278 Australia	0.306 Laos
0.277 Russia	0.222 Guyana	0.277 North Korea	0.305 Uruguay
0.277 Kuwait	0.221 Solomon Islands	0.277 Jamaica	0.305 Ghana
0.277 Oman	0.220 Nepal	0.276 Mauritius	0.305 Italy
0.277 Burundi	0.219 Zambia	0.276 Sierra Leone	0.305 New Zealand

Table 3: **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 1955-1990. Average group membership is reported beside the state name for the top 25 states in each latent group. The group assignments are consistent with known geopolitical coalitions in the Cold War, with Eastern bloc countries in Group 1, Western allies clustered in Group 4, and states that experienced proxy wars in Group 2.

References

Airoldi, Edoardo M et al. (2008). "Mixed membership stochastic blockmodels". In: *Journal of Machine Learning Research* 9, pp. 1981–2014.

Asuncion, Arthur et al. (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, Katherine (1996). "Economic interdependence: A Path to Peace or a Source of Interstate Conflict?" In: *Journal of Peace Research* 33.1, pp. 29–49.

Beck, Nathaniel, Jonathan N Katz, and Richard Tucker (1998). "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable". In: *American Journal of Political Science* 42.4, pp. 1260–1288.

Beck, Nathaniel, Gary King, and Langche Zeng (2000). "Improving Quantitative Studies of International Conflict: A Conjecture". In: *American Political Science Review* 94.1, pp. 21–35.

Chadefaux, Thomas (2014). "Early Warning Signals for War in the News". In: *Journal of Peace Research* 51.1, pp. 5–18.

Cranmer, Skyler J and Bruce A Desmarais (2011). "Inferential Network Analysis with Exponential Random Graph Models". In: *Political analysis* 19.1, pp. 66–86.

Dafoe, Allan (2011). "Statistical critiques of the democratic peace: Caveat emptor". In: *American Journal of Political Science* 55.2, pp. 247–262.

Dafoe, Allan, John R Oneal, and Bruce Russett (2013). "The Democratic Peace: Weighing the Evidence and Cautious Inference". In: *International Studies Quarterly* 57.1, pp. 201–214.

Fan, Xuhui, Longbing Cao, and Richard Yi Da Xu (2015). "Dynamic infinite mixed-membership stochastic blockmodel". In: *IEEE transactions on neural networks and learning systems* 26.9, pp. 2072–2085.

Farber, Henry S and Joanne Gowa (1997). "Common interests or common polities? Reinterpreting the democratic peace". In: *The Journal of Politics* 59.2, pp. 393–417.

Gartzke, Erik (2007). "The capitalist peace". In: *American journal of political science* 51.1, pp. 166–191.

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

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

Gleditsch, Nils Petter (1995). "Geography, Democracy, and Peace". In: *International Interactions* 20.4, pp. 297–323.

Gleditsch, Nils Petter and Haavard Hegre (1997). "Peace and democracy: Three levels of analysis". In: *Journal of Conflict Resolution* 41.2, pp. 283–310.

Goldsmith, Benjamin E (2007). "A liberal peace in Asia?" In: *Journal of Peace Research* 44.1, pp. 5–27.

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

Handcock, Mark S, Adrian E Raftery, and Jeremy M Tantrum (2007). "Model-based clustering for social networks". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170.2, pp. 301–354.

Hegre, Haavard (2008). "Gravitating Toward War: Preponderance May Pacify, but Power Kills". In: *Journal of Conflict Resolution* 52.4, pp. 566–589.

Hegre, Haavard et al. (2017). "Introduction: Forecasting in Peace Research". In: *Journal of Peace Research* 54.2, pp. 5–18.

Ho, Qirong and Eric P Xing (2015). "Handbook of Mixed Membership Models and Their Applications". In: ed. by Edoardo Airoldi et al. CRC Press. Chap. Analyzing time-evolving networks using an evolving cluster mixed membership blockmodel, pp. 489–525.

Hoff, Peter D, Adrian E Raftery, and Mark S Handcock (2002). "Latent space approaches to social network analysis". In: *Journal of the american Statistical association* 97.460, pp. 1090–1098.

Hoff, Peter D and Michael D Ward (2004). "Modeling Dependencies in International Relations Networks". In: *Political Analysis* 12.2, pp. 160–175.

Huth, Paul K (2009). *Standing Your Ground: Territorial Disputes and International Conflict*. University of Michigan Press.

Jones, Daniel M, Stuart A Bremer, and J David Singer (1996). "Militarized interstate disputes, 1816–1992: Rationale, coding rules, and empirical patterns". In: *Conflict Management and Peace Science* 15.2, pp. 163–213.

Jordan, Michael I et al. (1999). "An introduction to variational methods for graphical models". In: *Machine learning* 37.2, pp. 183–233.

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

Latouche, Pierre, Etienne Birmelé, Christophe Ambroise, et al. (2011). "Overlapping stochastic block models with application to the french political blogosphere". In: *The Annals of Applied Statistics* 5.1, pp. 309–336.

Leeds, Brett Ashley (2003). "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes". In: *American Journal of Political Science* 47.3, pp. 427–439.

Lorrain, Francois and Harrison C White (1971). "Structural equivalence of individuals in social networks". In: *The Journal of mathematical sociology* 1.1, pp. 49–80.

Mansfield, Edward D and Jack Snyder (2002). "Incomplete democratization and the outbreak of military disputes". In: *International Studies Quarterly* 46.4, pp. 529–549.

Maoz, Zeev (2009). "The Effects of Strategic and Economic Interdependence on International Conflict Across Levels of Analysis". In: *American Journal of Political Science* 53.1, pp. 223–240.

Maoz, Zeev, Ranan D Kuperman, et al. (2006). "Structural equivalence and international conflict: A social networks analysis". In: *Journal of Conflict Resolution* 50.5, pp. 664–689.

Maoz, Zeev and Bruce Russett (1993). "Normative and structural causes of democratic peace, 1946–1986". In: *American Political Science Review* 87.3, pp. 624–638.

Marshall, Monty, Ted Robert Gurr, and Keith Jaggers (2017). "Polity IV Project, Political Regime Characteristics and Transitions, 1800-2016." In: *Polity IV Project-Dataset Users' Manual*.

Matias, Catherine and Vincent Miele (2017). "Statistical clustering of temporal networks through a dynamic stochastic block model". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.4, pp. 1119–1141.

Mousseau, Michael (2009). "The social market roots of democratic peace". In: *International Security* 33.4, pp. 52–86.

Oneal, John R and Bruce Russett (1999). "The Kantian peace: The pacific benefits of democracy, interdependence, and international organizations, 1885–1992". In: *World politics* 52.1, pp. 1–37.

Oneal, John R and Jaroslav Tir (2006). "Does the Diversionary Use of Force Threaten the Democratic Peace? Assessing the Effect of Economic Growth on Interstate Conflict, 1921–2001". In: *International Studies Quarterly* 50.4, pp. 755–779.

Peceny, Mark, Caroline C Beer, and Shannon Sanchez-Terry (2002). "Dictatorial Peace?" In: *American Political Science Review* 96.1, pp. 15–26.

Ray, James Lee (1998). "Does democracy cause peace?" In: *Annual Review of Political Science* 1.1, pp. 27–46.

Salter-Townshend, Michael and Thomas Brendan Murphy (2015). "Role analysis in networks using mixtures of exponential random graph models". In: *Journal of Computational and Graphical Statistics* 24.2, pp. 520–538.

Schrodt, Philip A (1991). "Prediction of Interstate Conflict Outcomes Using a Neural Network". In: *Social Science Computer Review* 9.3, pp. 359–380.

Singer, J David, Stuart Bremer, and John Stuckey (1972). "Capability Distribution, Uncertainty, and Major Power War, 1820-1965". In: *Peace, war, and numbers* 19, p. 48.

Snijders, Tom AB and Krzysztof Nowicki (1997). "Estimation and prediction for stochastic blockmodels for graphs with latent block structure". In: *Journal of classification* 14.1, pp. 75–100.

Stephens, Matthew (2000). "Dealing with label switching in mixture models". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62.4, pp. 795–809.

Stinnett, Douglas M et al. (2002). "The Correlates of War (COW) Project Direct Contiguity Data, version 3.0". In: *Conflict Management and Peace Science* 19.2, pp. 59–67.

Sweet, Tracy, Andrew Thomas, and Brian Junker (2014). "Handbook of mixed membership models and their applications". In: ed. by Edoardo Airoldi et al. CRC Press. Chap. Hierarchical mixed membership stochastic blockmodels fir nultiple networks and experimental interventions.

Teh, Yee W., David Newman, and Max Welling (2007). "A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation". English. In: *Advances in Neural Information Processing Systems*, pp. 1353–1360.

Wang, Yuchung J and George Y Wong (1987). "Stochastic blockmodels for directed graphs". In: *Journal of the American Statistical Association* 82.397, pp. 8–19.

Ward, Michael D, Nils W Metternich, et al. (2013). "Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction". In: *International Studies Review* 15.4, pp. 473–490.

Ward, Michael D, Randolph M Siverson, and Xun Cao (2007). "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace". In: *American Journal of Political Science* 51.3, pp. 583–601.

Wasserman, Stanley and Katherine Faust (1994). *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press.

White, Arthur and Thomas Brendan Murphy (2016). "Mixed-membership of experts stochastic blockmodel". In: *Network Science* 4.1, pp. 48–80.

Xing, Eric P, Wenjie Fu, and Le Song (2010). "A state-space mixed membership blockmodel for dynamic network tomography". In: *The Annals of Applied Statistics* 4.2, pp. 535–566.