



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

Santiago Olivella^a, Tyler Pratt^b, and Kosuke Imai^c 

^aUNC-Chapel Hill, Chapel Hill, NC; ^bYale University, New Haven, CT; ^cDepartment of Government and Department of Statistics, Institute for Quantitative Social Science, Harvard University, Cambridge, MA

ABSTRACT

The decision to engage in military conflict is shaped by many factors, including state- and dyad-level characteristics as well as the state's membership in geopolitical coalitions. Supporters of the democratic peace theory, for example, hypothesize that the community of democratic states is less likely to wage war with each other. Such theories explain the ways in which nodal and dyadic characteristics affect the evolution of conflict patterns over time via their effects on group memberships. To test 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 dynamic node memberships in latent groups as well as the direct formation of edges between dyads. While prior substantive research often assumes the decision to engage in international militarized conflict is independent across states and static over time, we demonstrate that conflict is driven by states' evolving membership in geopolitical blocs. Our analysis of militarized disputes from 1816 to 2010 identifies two distinct blocs of democratic states, only one of which exhibits unusually low rates of conflict. Changes in monadic covariates like democracy shift states between coalitions, making some states more pacific but others more belligerent. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received December 2019
Accepted December 2021

KEYWORDS

Democratic peace; Hidden Markov model; International conflict; Mixed-membership stochastic blockmodel; Social networks



1. Introduction

Social scientists often posit theories about the effects of latent groups of actors on relational outcomes of interest over time. In the study of international conflict, scholars debate the so-called “democratic peace” hypothesis, which states that a specific bloc of actors—defined by their democratic institutions—rarely engage in wars amongst themselves (e.g., Oneal and Russett 1999). Others argue that militarized conflict is driven by state membership in geopolitical coalitions that evolve over time (Farber and Gowa 1997). These theories 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 test 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. This relaxes the strict assumption of stochastic equivalence for members of the same groups. Finally, we apply collapsed variational inference and improve computational scalability of the model. Our approach, 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 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), using stochastic gradient descent to accommodate 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 a dynamic analysis of international conflicts among states over the last two centuries.

CONTACT Kosuke Imai  imai@harvard.edu  Department of Government and Department of Statistics, Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138.

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>.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

© 2021 American Statistical Association

Political scientists have long sought to explain the causes of interstate conflict and predict its outbreak. In the study of the aforementioned democratic peace hypothesis, 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; Imai and Lo 2021). 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.

Our findings provide several new insights into the origins of conflict in the international system. First, our model identifies two distinct blocs of democracies that exhibit disparate rates of conflict. One group, composed of states with lower levels of military capacity, rarely engages in conflict with other democracies. The other group exhibits no such pacific tendency, regularly engaging in militarized disputes among themselves and with others. Second, we demonstrate that the effect of democracy on conflict varies both across states and over time. Changes in domestic political institutions shift states between latent groups, making some states more pacific (e.g., Germany) but others more belligerent (Kosovo). Over time, the evolution of the group structure has reduced the average effect of democratization on conflict.

1.1. Related Models

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, Raftery, and Tantrum 2007) and the recently developed ego-ERGM (Salter-Townshend and 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, relaxes some of these independence assumptions. For instance, Sweet, Thomas, and Junker (2014) incorporate 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) incorporate 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) shows 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, Fu, and Song 2010; Ho and Xing 2015; Fan, 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 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 simultaneously address the need to incorporate dyadic and nodal attributes as well as the need to account for temporal dynamics, in an effort to develop a model that can be readily employed in applied research.

2. Challenges of Modeling 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 literature on conflict forecasting (e.g., Schrodtt 1991; Beck, King, and Zeng 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).

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; Gartzke 2007; Dafoe, Oneal, and Russett 2013). However, there are reasons to believe conflict patterns violate this conditional independence 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 nonallies. 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.

Recent analyses 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, Siverson, and Cao (2007) apply the latent space model developed by Hoff, Raftery, and Handcock (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. While we build on this emerging body of scholarship that seeks to model complex dependencies in the conflict network, our approach addresses several challenges faced by these existing network modeling strategies.

First, and although existing approaches can capture higher order dependencies in conflict relations, 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 of network analyses of international conflict is the need to restructure monadic covariates like democracy to fit a dyadic analysis. This problem has exacerbated a debate in the democratic peace literature regarding the appropriate dyadic specification of democracy (see Dafoe, Oneal, and Russett 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, Beer, and Sanchez-Terry 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 accommodated by the model structure. Each latent group is directly associated with its own set of nodal covariates, and the dynamic implementation provides flexibility for covariate effects to vary over time.

3. The Proposed Model

Analyzing the interstate conflict network to study the democratic peace theory requires a model that defines the probability of conflict as a function of membership in latent groups of countries. In addition, the model must enable 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 should deal with the computational complexity involved in estimating a dynamic network model with a large number of nodes.

Below, 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. 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,

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

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 MMSBM by allowing the mixed membership vectors to not only be a function of node-level predictors, but also by letting these vectors change over time as the Markov states evolve. Specifically, these random states are generated according to $S_t | S_{t-1} = n \sim \text{Categorical}(\mathbf{A}_n)$, 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,

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

where g^{-1} is the logistic function, and $\mathbf{z}_{p \rightarrow q,t} \sim \text{Multinomial}(1, \boldsymbol{\pi}_{pt})$ 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 $S_t | S_{t-1} = n \sim \text{Categorical}(\mathbf{A}_n)$.
2. For each node i at time t , draw state-dependent mixed-membership vector $\boldsymbol{\pi}_{it} | S_t = m \sim \text{Dirichlet}(\{\exp(\mathbf{x}_{it}^\top \boldsymbol{\beta}_{k,m})\}_{k=1}^K)$.
3. For each pair of nodes p and q at time t ,
 - Sample a group indicator $\mathbf{z}_{p \rightarrow q,t} \sim \text{Multinomial}(1, \boldsymbol{\pi}_{pt})$.
 - Sample a group indicator $\mathbf{w}_{q \leftarrow p,t} \sim \text{Multinomial}(1, \boldsymbol{\pi}_{qt})$.
 - Sample a link between them
$$Y_{pqt} \sim \text{Bernoulli}\left(g^{-1}\left(\mathbf{z}_{p \rightarrow q,t}^\top \mathbf{B} \mathbf{w}_{q \leftarrow p,t} + \mathbf{d}_{pqt}^\top \boldsymbol{\gamma}\right)\right).$$

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

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

where $\mathbf{L} := \{\mathbf{Z}, \mathbf{W}, \mathbf{S}\}$ collects all latent group memberships and hidden Markov states, $\boldsymbol{\Pi} := \{\boldsymbol{\pi}_{it}\}_{i \in V_t} \forall t$ 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 $\mathbf{z}_{p \rightarrow q,t}$ ($\mathbf{w}_{q \leftarrow p,t}$) and the mixed-membership parameters $\boldsymbol{\pi}_{pt}$ ($\boldsymbol{\pi}_{qt}$) are usually strongly correlated (Teh, Newman, and Welling 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 distri-

bution, 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 Section A of the supplementary information. Letting $\alpha_{itkm} = \exp(\mathbf{x}_{it}^\top \boldsymbol{\beta}_{km})$, $\alpha_{it \cdot m} = \sum_{k=1}^K \alpha_{itkm}$, and $\theta_{pqth} = g^{-1}(\mathbf{B}_{gh} + \mathbf{d}_{pqt}^\top \boldsymbol{\gamma})$, the resulting collapsed posterior is proportional to:

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

where $I(\cdot)$ is the binary indicator function, and $\Gamma(\cdot)$ is the Gamma function.

The marginalized joint distribution explicitly use a number of sufficient statistics: $C_{itk} = \sum_{q \in V_t} (z_{i \rightarrow q,t,k} + w_{i \leftarrow q,t,k})$, 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 $U_m = \sum_{t=2}^T \sum_n I(S_t = n) I(S_{t-1} = m)$, 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, Newman, and Welling 2007). To do so, we define a factorized distribution over the latent variables \mathbf{L} as

$$\begin{aligned} \tilde{Q}(\mathbf{L} | \mathbf{K}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) &= \prod_{t=1}^T Q_1(\mathbf{s}_t | \boldsymbol{\kappa}_t) \\ &\times \prod_{p,q \in V_t} Q_2(\mathbf{z}_{p \rightarrow q,t} | \boldsymbol{\phi}_{p \rightarrow q,t}) Q_2(\mathbf{w}_{q \leftarrow p,t} | \boldsymbol{\psi}_{q \leftarrow p,t}), \end{aligned} \quad (5)$$

where $\boldsymbol{\kappa}_t$, $\boldsymbol{\phi}_{p \rightarrow q,t}$, 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 Wang and Blunsom 2013).

We then apply Jensen's inequality to derive a lower bound for the log marginal probability of our network data \mathbf{Y}

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

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 \tilde{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 ϕ parameters for all (pt, qt) dyads in our data as follows:

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

where $C'_{ptk} = C_{ptk} - z_{p \rightarrow q, t, k}$ 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 $t = 2, \dots, T - 1$, we update all hidden Markov state variational parameters according to

$$\begin{aligned} \hat{\kappa}_{tm}^{(s)} &\propto \exp \left[-\mathbb{E}_{\tilde{Q}_1} [\log(M\eta + U'_{m-})] \right] \\ &\times \exp \left[\kappa_{t+1, m} \kappa_{t-1, m} \mathbb{E}_{f, \tilde{Q}_1} [\log(\eta + U'_{mm} + 1)] \right] \\ &\times \exp \left[(\kappa_{t-1, m} - \kappa_{t-1, m} \kappa_{t+1, m} + \kappa_{t+1, m}) \mathbb{E}_{f, \tilde{Q}_1} [\log(\eta + U'_{mm})] \right] \\ &\times \prod_{n \neq m} \exp \left[\kappa_{t+1, n} \mathbb{E}_{f, \tilde{Q}_1} [\log(\eta + U'_{nn})] \right] \\ &\times \prod_{n \neq m} \exp \left[\kappa_{t-1, n} \mathbb{E}_{f, \tilde{Q}_1} [\log(\eta + U'_{nn})] \right] \\ &\times \prod_{pt \in V_t} \left[\frac{\Gamma(\alpha_{it-m})}{\Gamma(\alpha_{it-m} + 2N_t)} \prod_{k=1}^K \frac{\mathbb{E}_{\tilde{Q}_1} [\Gamma(\alpha_{ptmk} + C_{ptk})]}{\Gamma(\alpha_{ptmk})} \right], \end{aligned} \quad (8)$$

where $U'_{m-} = U_m - s_{t, m}$ and $U'_{mn} = U_{mn} - s_{tm} s_{t+1, n}$. This definition of the term U'_{mn} is valid whenever $m \neq n$ and $t \neq T$ (for other cases, see Section B of the supplementary material).

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 $\mathbb{E}_{f, \tilde{Q}_2} [\log(\alpha_{ptmk} + C'_{ptk})] \approx \log(\alpha_{ptmk} + \mathbb{E}_{f, \tilde{Q}_2} [C'_{ptk}])$ and similarly for terms involving all U' 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 following lower bound, using a quasi-Newton method (see Section B of the supplementary material for the expressions of the required gradients),

$$\begin{aligned} \mathcal{L}_{\phi, \kappa}(\mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma}) &\triangleq \sum_{t=1}^T \sum_{m=1}^M \kappa_{tm} \sum_{p \in V_t} \log \Gamma(\xi_{ptm}) \\ &- \log \Gamma(\xi_{ptm} + 2N_t) \\ &+ \sum_{t=1}^T \sum_{m=1}^M \kappa_{tm} \sum_{p \in V_t} \sum_{k=1}^K \mathbb{E}_{\tilde{Q}} [\log \Gamma(\alpha_{ptmk} + C_{ptk})] \\ &- \log \Gamma(\alpha_{ptmk}) + \end{aligned}$$

$$\begin{aligned} &+ \sum_{t=1}^T \sum_{(p, q) \in E_t} \sum_{g, h=1}^K \phi_{p \rightarrow q, t, g} \psi_{q \leftarrow p, t, h} \\ &\times \{ \gamma_{pqtkg} \log(\theta_{pqtkg}) + (1 - \gamma_{pqtkg}) \log(1 - \theta_{pqtkg}) \} \\ &- \sum_{t=1}^T \sum_{m=1}^M \sum_{(p, q) \in E_t} \sum_{k=1}^K \{ \phi_{p \rightarrow q, t, k} \log(\phi_{p \rightarrow q, t, k}) \\ &- \psi_{q \leftarrow p, t, h} \log(\psi_{q \leftarrow p, t, k}) \}. \end{aligned} \quad (9)$$

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} .

3.4. Stochastic VI Algorithm

For problems involving large networks, the above variational approximation can be computationally intensive even after parallelization (see Section 3.5). To enable fast inference on networks with a 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, Gaussier, and Langeron 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 6. 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 containing the sampled nodes. The algorithm proceeds by optimizing the local variational parameters (i.e., Φ 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:

$$\begin{aligned} \mathbf{C}_t^{(s)} &= (1 - \rho_s) \mathbf{C}_t^{(s-1)} + \rho_s \mathbb{E}_f [\mathbf{C}_t]; \\ \mathbf{U}^{(s)} &= (1 - \rho_s) \mathbf{U}^{(s-1)} + \rho_s \mathbb{E}_f [\mathbf{U}] \end{aligned} \quad (10)$$

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 (Cappé and Moulines 2009; Hoffman et al. 2013). 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. As an example, for $\boldsymbol{\gamma}$, we have:

$$\boldsymbol{\gamma}^{(s)} = \boldsymbol{\gamma}^{(s-1)} + \rho_s \nabla_{\boldsymbol{\gamma}} \mathcal{L}_{\phi, \kappa}^{(s)}(\boldsymbol{\gamma}) \quad (11)$$

where

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

is a random function that is equal to the third line in Equation 9 in expectation. The updates for all other hyper-parameters are similarly defined (Hoffman et al. 2013). Section B of the supplementary material provides the required gradients.

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 all dyads at every iteration (Gopalan and Blei 2013).

3.5. 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, Ke, and Luo 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 realigning the (assumed constant) blockmodels using a graph matching algorithm (Lyzinski, Fishkind, and Priebe 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 (10.0^{-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 average change falls below a tolerance of 10.0^{-3} or when no improvement has been observed in the past 20 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 Φ in parallel across (subsampling) dyads.

In Section C of the supplementary material, 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.

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 (Palmer et al. 2015).¹ 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 sensitive to 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 S2 in Section D of the supplementary material for prediction tests and Figures S7 and S8 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, Bremer, and Singer 1996, p. 168). Ties in the network are 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 rare event, only 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, 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 -0.43 . Roughly 6% 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, Bremer, and Stuckey 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

¹ The MID data are available at <https://correlatesofwar.org/data-sets/MIDs>

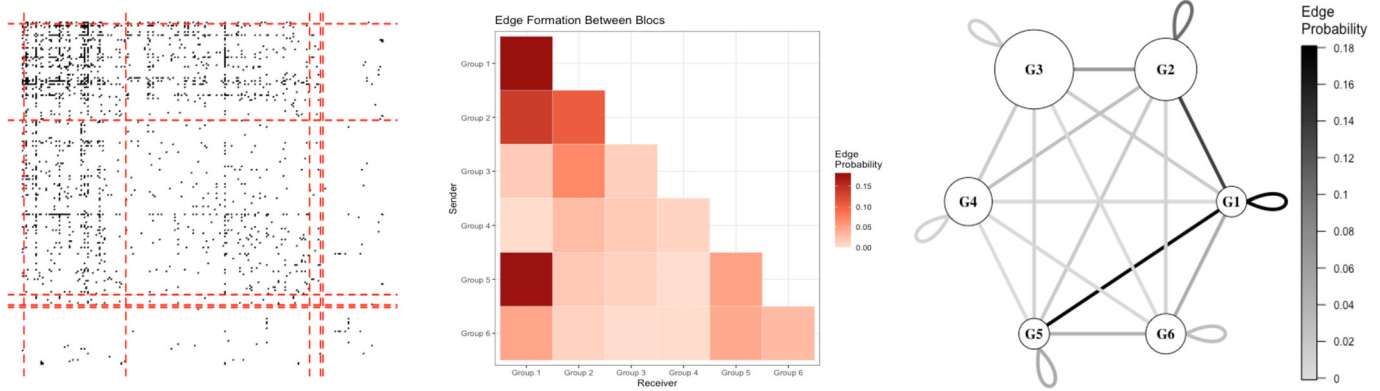


Figure 1. Estimated blockmodel in the conflict network. The left panel displays the adjacency matrix of militarized disputes between 216 states. Black squares indicate the existence of at least one MID between the states in row x and column y ; dotted lines separate states by estimated group membership. The middle panel displays the estimated probability of conflict between members of 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.

(Stinnett et al. 2002).² 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, Katz, and Tucker 1998). Finally, to account for the missing values of some predictors, we rely on a missing-indicator approach, adding dummy variables that indicate which observations have missing values in the corresponding variable, and replacing all missing values with zero.

The model is fitted using our open-source software package NetMix. Estimation took 1 hr and 18 min on a computer with a 3.6 Ghz CPU, converging after 709 EM iterations. Note that the estimation time drops to approximately 55 min 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 216×216 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 0.14% of the time. Table S4 of Section D in the supplementary material presents the estimated blockmodel 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 neutral stance, these states maintained close diplomatic relations with the Western allies that populate Group 1. According to the blockmodel, Group 3 has a low rate of conflict with Group 1 (1.7%) and is 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

²As an alternative way to address geographic effects, we estimate a specification that includes a set of regional indicator variables (see Table S3 and Figures S5 and S6 in the supplementary material).



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.

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 era lends further credence to the validity of the model. As noted earlier, this 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 recovers the underlying geopolitical structure of the Cold War, 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 S5 of Section D of the supplementary material.

The group memberships of countries are 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 (e.g., China, East Germany, Poland, Czechoslovakia, and Romania). The estimated blockmodel indicates the competing coalitions experience abnormally high rates of conflict.

4.3. The Dynamics of Membership

The dynMMSBM further allows us to examine how latent group membership changes over time. Figure 2 displays the evolution of group membership 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 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. U.S. membership is comparably stable over the period of the study, while the United Kingdom consolidates its membership in these groups after transitioning to a democratic political system. For example, we observe a sharp increase in the U.K.’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

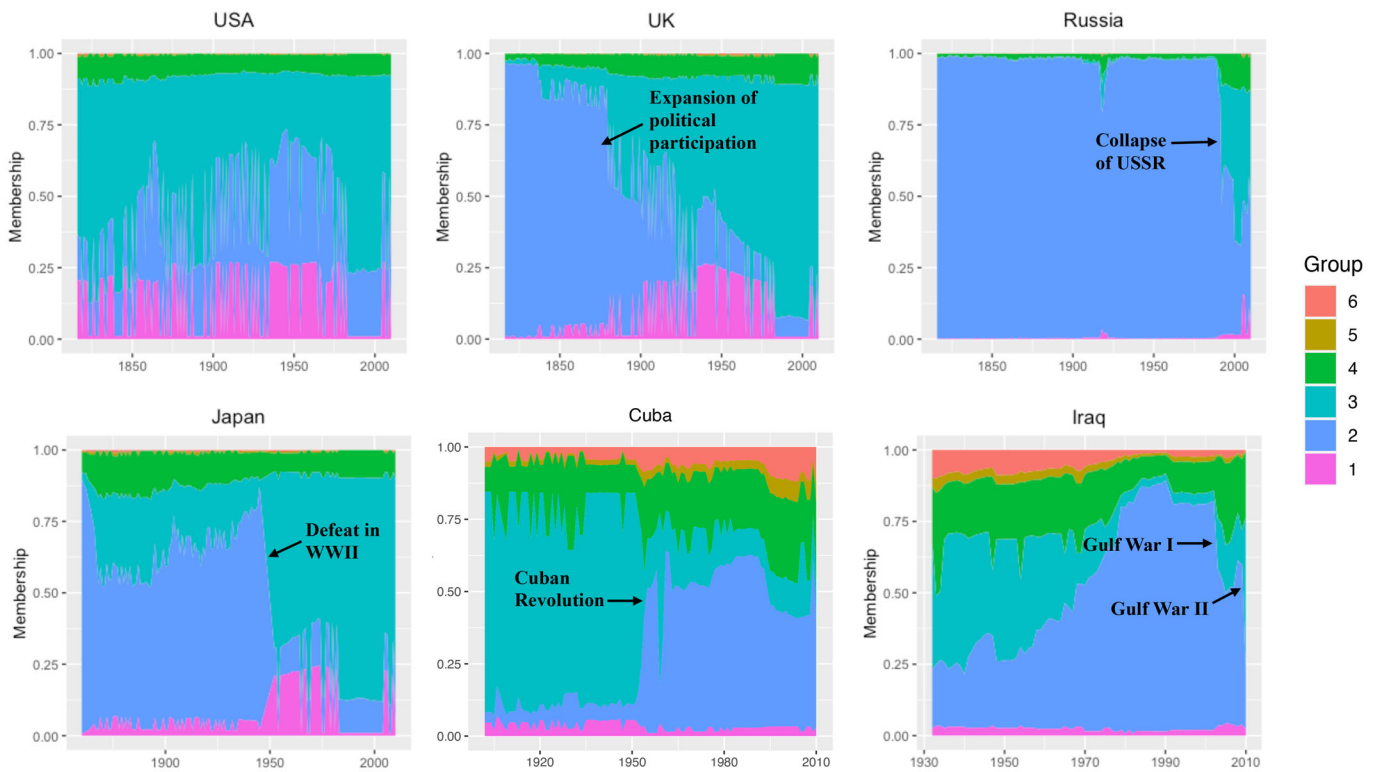


Figure 3. Average group membership over time, select countries. The figure shows, for six countries, the average rate of membership in four latent groups in each year the country is present in the network.

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 United States 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 S6 in Section D of the supplementary material 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 the Western alliance of liberal democracies during the Cold War, and Group 3 as Western-leaning neutral states. Notably, these two democratic clusters exhibit 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 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 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

Table 1. Estimated coefficients and their standard errors.

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)						

Note: 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.

N nodes: 216; N dyad-years: 842, 685; N time periods: 195

Lower bound at convergence: −527, 587.7

of this change on the probability of edge formation, averaging all dyadic interactions and time periods,

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

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 nonlinear and contingent upon the existing group membership of the node in question and the other nodes in the 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 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

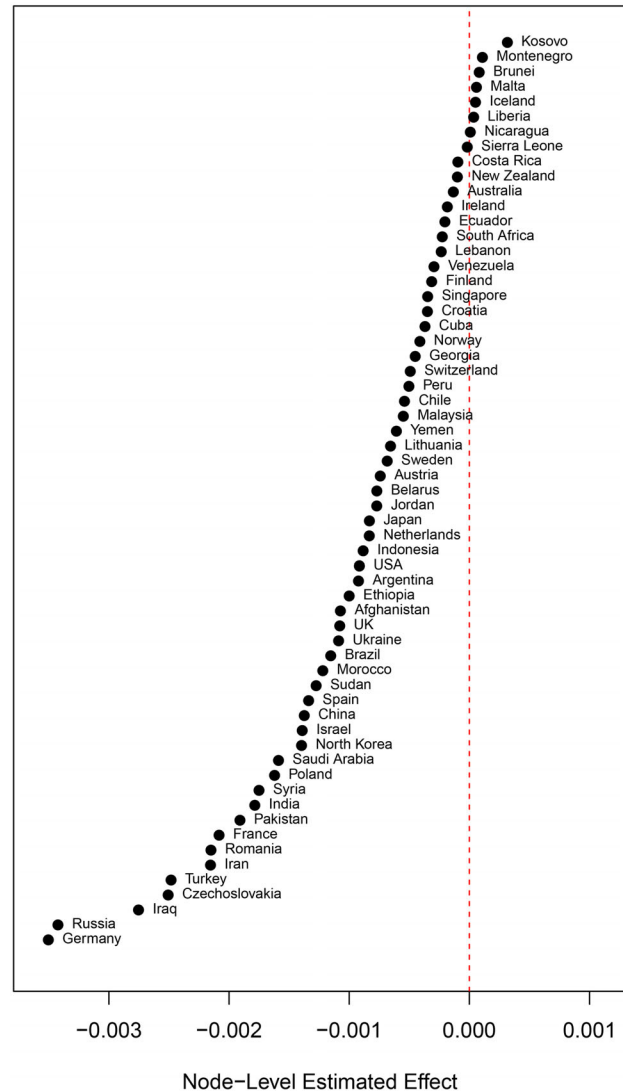


Figure 4. Estimated effects of covariate shift in polity over time, select states. The figure shows the estimated change in the probability of interstate conflict if a state's POLITY score is increased by one standard deviation (6.78) from its observed value.

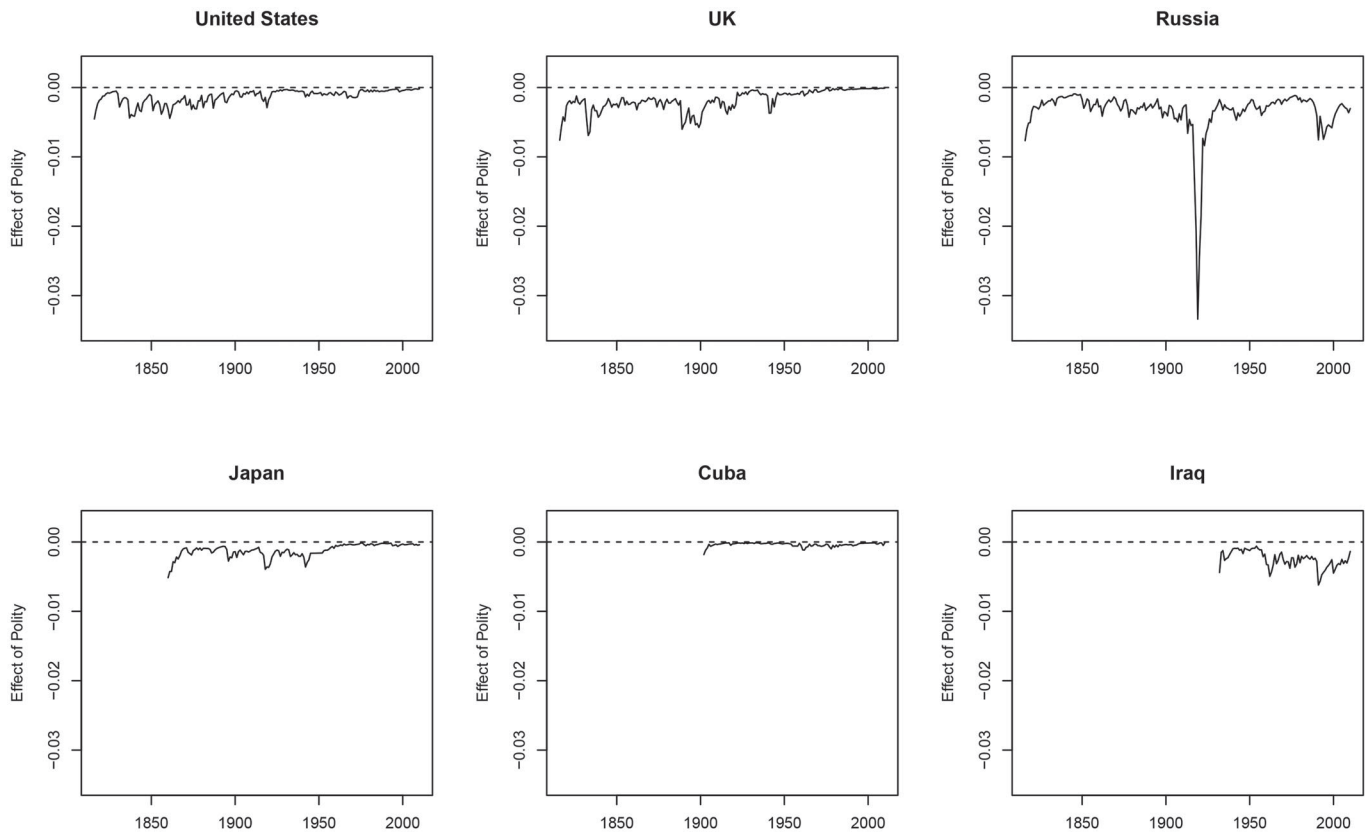


Figure 5. Effect of shift in polity over time, select states. The figure shows the estimated change in the probability of interstate conflict over time if a country's POLITY score is increased by one standard deviation (6.78) from its observed value (up to a maximum of 10).

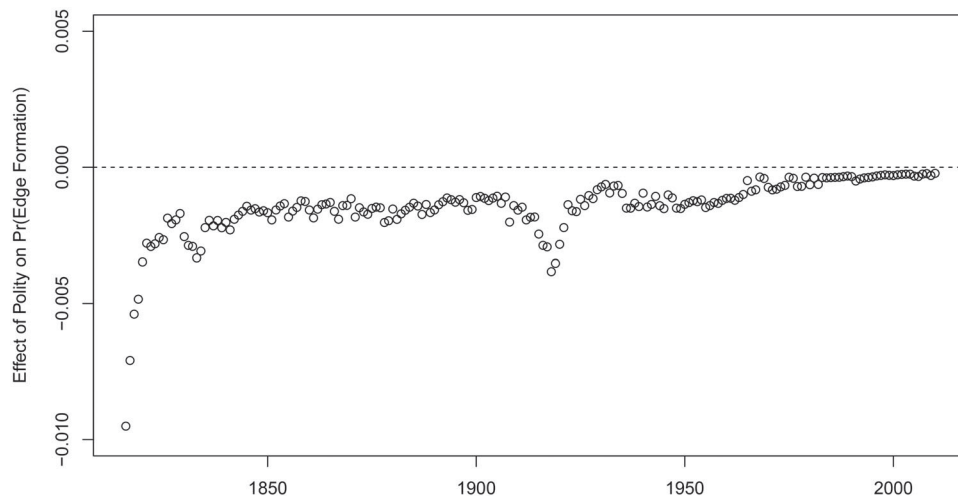


Figure 6. Estimated aggregate effect of shift in polity over time. The figure shows the estimated average change in the probability of interstate conflict when countries' POLITY scores are increased by one standard deviation (6.78) up to the maximum POLITY score.

elevated rate of conflict between Group 1 (United States, United Kingdom, 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.³

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.

4.5. Additional Analyses

In Section D.9 of the supplementary material, we compare the results of our empirical analysis with those of the standard logistic regression model, which assumes all dyad-years are conditionally independent 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 to 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.

Our primary results reflect a batch analysis of the data, taking all years into consideration. In Section D.7 of the supplementary material, we replicate our analysis via “online” updating, where we iteratively expand the time window to update estimates as if the data had been obtained sequentially, rather than in batch. To illustrate this approach, we first fit a model for the years 1816–1820, then use the resulting mixed membership estimates as starting values for a model that incorporates the next window (1821–1825). We repeat until all years are included (see Table S7 and Figures S9 and S10).

5. Conclusion

We have analyzed a network defined by almost 200 years of militarized interstate disputes in the international system, uncovering previously understudied spatial and temporal heterogeneity. While prior substantive research often assumes the decision to engage in international conflict is independent across dyads and static over time, we demonstrate that conflict is driven by states’ evolving membership in geopolitical blocs.

Our findings add important nuance to 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 analysis of conflict patterns reveals two distinct communities of democratic states: one highly peaceful, the other regularly belligerent. The effect of democracy on conflict is conditional on a state’s initial position in the latent group structure and its membership in these democratic blocs, generating heterogeneous effects across states and over time. We also uncover the evolving nature of unobserved geopolitical coalitions, with memberships that conform to theoretical expectations.

In addition to these substantive contributions to the study of conflict, this paper provides applied researchers with the dynMMSBM, a model that can accommodate a variety of theorized relationships for dynamic network outcomes that display some form of stochastic equivalence. The field of international relations is full of such dynamic network outcomes: international organization memberships, the signing of treaties and other agreements, and international cooperation in criminal investigation and prosecution are just a few examples of outcomes that can be studied using the tools we develop here. To this end, we make available the open-source R software package NetMix that implements the methodology we have used to study international conflict. In the future, we plan to further broaden the model’s applicability by considering these and other outcome variable types. In particular, and given their prevalence in social scientific research, we plan to extend the model to accommodate bipartite (or affiliation) networks.

Supplementary Materials

The supplementary materials contain complementary empirical analyses and derivations. Code and data needed to replicate the results presented in the main text, as well as all analyses presented in the supplementary materials, can be found at <https://doi.org/10.7910/DVN/82CULX>.

ORCID

Kosuke Imai  <http://orcid.org/0000-0002-2748-1022>

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. [1,2,6]
- 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*, pp. 27–34. AUAI Press. [5]
- Barbieri, K. (1996), “Economic Interdependence: A Path to Peace or a Source of Interstate Conflict?” *Journal of Peace Research*, 33, 29–49. [2]
- 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. [7]
- Beck, N., King, G., and Zeng, L. (2000), “Improving Quantitative Studies of International Conflict: A Conjecture,” *American Political Science Review*, 94, 21–35. [2]
- 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. [3]
- Cappé, O., and Moulines, E. (2009), “On-line Expectation–Maximization Algorithm for Latent Data Models,” *Journal of the Royal Statistical Society, Series B*, 71, 593–613. [5]

³To ensure these patterns are not a function of ceiling effects—given that the number of states with the maximum polity score of 10 is increasing over the time period—we also calculate the effect of a one standard deviation decrease in polity (see Figures S11 and S12 in the supplementary material). The effects are substantively identical.

- Cederman, L.-E. (2001), "Back to Kant: Reinterpreting the Democratic Peace as a Macrohistorical Learning Process," *American Political Science Review*, 95, 15–31. [3,8]
- Cranmer, S. J., and Desmarais, B. A. (2011), "Inferential Network Analysis with Exponential Random Graph Models," *Political Analysis*, 19, 66–86. [3]
- Dafoe, A., Oneal, J. R., and Russett, B. (2013), "The Democratic Peace: Weighing the Evidence and Cautious Inference," *International Studies Quarterly*, 57, 201–214. [2,3]
- Diebold, F., and Mariano, R. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263. [12]
- Dulac, A., Gaussier, E., and Langeron, C. (2020), "Mixed-Membership Stochastic Block Models for Weighted Networks," in *Conference on Uncertainty in Artificial Intelligence*, pp. 679–688. PMLR. [5]
- 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. [2]
- Farber, H. S., and Gowa, J. (1997), "Common Interests or Common Politics? Reinterpreting the Democratic Peace," *The Journal of Politics*, 59, 393–417. [1,2,3]
- Fearon, J. D. (1994), "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review*, 88, 577–592. [3]
- 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. [5]
- Gartzke, E. (2007), "The Capitalist Peace," *American Journal of Political Science*, 51, 166–191. [2]
- Gibler, D. M. (2009), *International Military Alliances, 1648–2008*, Washington DC: CQ Press. [6]
- Gleditsch, N. P., and Hegre, H. (1997), "Peace and Democracy: Three Levels of Analysis," *Journal of Conflict Resolution*, 41, 283–310. [2,3]
- 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. [1,6]
- Gowa, J. (2011), *Ballots and Bullets: The Elusive Democratic Peace*, Princeton, NJ: Princeton University Press. [2]
- 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*, 170, 301–354. [2]
- Hegre, H. (2008), "Gravitating Toward War: Preponderance May Pacify, but Power Kills," *Journal of Conflict Resolution*, 52, 566–589. [2,6]
- Hegre, H., Metternich, N. W., Nygaard, H. M., and Wucherpfennig, J. (2017), "Introduction: Forecasting in Peace Research," *Journal of Peace Research*, 54, 5–18. [2]
- Ho, Q., and Xing, E. P. (2015), "Analyzing Time-Evolving Networks Using an Evolving Cluster Mixed Membership Blockmodel," in *Handbook of Mixed Membership Models and Their Applications*, eds. E. M. Airlodi, D. M. Blei, E. A. Erosheva, and S. E. Fienberg, pp. 489–525, Boca Raton, FL: CRC Press. [2]
- Hoff, P. D. (2009), "Multiplicative Latent Factor Models for Description and Prediction of Social Networks," *Computational and Mathematical Organization Theory*, 15, 261–272. [1]
- Hoff, P. D., and Ward, M. D. (2004), "Modeling Dependencies in International Relations Networks," *Political Analysis*, 12, 160–175. [3]
- 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. [3]
- Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013), "Stochastic Variational Inference," *The Journal of Machine Learning Research*, 14, 1303–1347. [1,5,6]
- Imai, K., and Lo, J. (2021), "Robustness of Empirical Evidence for the Democratic Peace: A Nonparametric Sensitivity Analysis," *International Organization*, 75, 901–919. [2]
- Jin, J., Ke, Z. T., and Luo, S. (2018), "SCORE+ for Network Community Detection," *CoRR abs/1811.05927*. [6]
- 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. [6]
- 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. [4]
- Kim, M., and Leskovec, J. (2013), "Nonparametric Multi-Group Membership Model for Dynamic Networks," in *Advances in Neural Information Processing Systems*, pp. 1385–1393. [2]
- 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. [2]
- Lorrain, F., and White, H. C. (1971), "Structural Equivalence of Individuals in Social Networks," *The Journal of Mathematical Sociology*, 1, 49–80. [1]
- Lyzinski, V., Fishkind, D. E., and Priebe, C. E. (2014), "Seeded Graph Matching for Correlated Erdős-Rényi Graphs," *Journal of Machine Learning Research*, 15, 3513–3540. [6]
- Maoz, Z., and Russett, B. (1993), "Normative and Structural Causes of Democratic Peace, 1946–1986," *American Political Science Review*, 87, 624–638. [2,3,6]
- 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. [3]
- Marshall, M., Gurr, T. R., and Jagers, K. (2017), "Polity IV Project, Political Regime Characteristics and Transitions, 1800–2016," *Polity IV Project-Dataset Users' Manual*. [6]
- 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*, 79, 1119–1141. [2]
- 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. [1,2,7]
- 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. [2]
- Palmer, G., d'Orazio, V., Kenwick, M., and Lane, M. (2015), "The MID4 Dataset, 2002–2010: Procedures, Coding Rules and Description," *Conflict Management and Peace Science*, 32, 222–242. [6]
- Peceny, M., Beer, C. C., and Sanchez-Terry, S. (2002), "Dictatorial Peace?" *American Political Science Review*, 96, 15–26. [3]
- Salter-Townshend, M., and Murphy, T. B. (2015), "Role Analysis in Networks Using Mixtures of Exponential Random Graph Models," *Journal of Computational and Graphical Statistics*, 24, 520–538. [2]
- Schrodt, P. A. (1991), "Prediction of Interstate Conflict Outcomes Using a Neural Network," *Social Science Computer Review*, 9, 359–380. [2]
- Singer, J. D., Bremer, S., and Stuckey, J. (1972), "Capability Distribution, Uncertainty, and Major Power War, 1820–1965," in *Peace, War, and Numbers*, ed. B. Russett, pp. 19–48, Beverly Hills: Sage. [6]
- 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. [7]
- Sweet, T., Thomas, A., and Junker, B. (2014), "Hierarchical Mixed Membership Stochastic Blockmodels for Multiple Networks and Experimental Interventions," in *Handbook of Mixed Membership Models and their Applications*, eds. E. M. Airlodi, D. M. Blei, E. A. Erosheva, and S. E. Fienberg, pp. 463–488, Boca Raton, FL: CRC Press. [2]
- 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. [1,4]
- Wang, P., and Blunsom, P. (2013), "Collapsed Variational Bayesian Inference for Hidden Markov Models," in *Artificial Intelligence and Statistics*, pp. 599–607. [4]
- Wang, Y. J., and Wong, G. Y. (1987), "Stochastic Blockmodels for Directed Graphs," *Journal of the American Statistical Association*, 82, 8–19. [1]

- 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. [3]
- 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. [2]
- Wasserman, S., and Faust, K. (1994), *Social Network Analysis: Methods and Applications* (Vol. 8), Cambridge, UK: Cambridge University Press. [1]
- White, A., and Murphy, T. B. (2016), “Mixed-Membership of Experts Stochastic Blockmodel,” *Network Science*, 4, 48–80. [2]
- 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. [2]
- 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. [2]