Online Supplementary Information for "Dynamic Stochastic Blockmodel Regression for Network Data: Application to International Militarized Conflicts."

October 20, 2021

A Marginalizing the membership vectors and the transition probabilities

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

$$\int \cdots \int \prod_{t=1}^{T} \prod_{p \in V_t} \left[\prod_{m=1}^{M} P(\boldsymbol{\pi}_{pt} \mid \boldsymbol{\alpha}_{ptm})^{s_{tm}} \right] \prod_{q \in V_t} P(\mathbf{z}_{p \to q,t} \mid \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{p \leftarrow q,t} | \boldsymbol{\pi}_{pt}) \, \mathrm{d}\boldsymbol{\pi}_1 \ldots \mathrm{d}\boldsymbol{\pi}_{N_t}$$

$$= \prod_{t=1}^{T} \prod_{p \in V_t} \int \prod_{m=1}^{M} [P(\boldsymbol{\pi}_{pt} \mid \boldsymbol{\alpha}_{ptm})]^{s_{tm}} \prod_{q \in V_t} P(\mathbf{z}_{p \to q,t} \mid \boldsymbol{\pi}_{pt}) P(\mathbf{w}_{p \leftarrow q,t} | \boldsymbol{\pi}_{pt}) \, \mathrm{d}\boldsymbol{\pi}_{pt}$$

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

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

$$\times \int \prod_{k=1}^{K} \pi_{ptk}^{\sum_{m=1}^{M} s_{tm} \alpha_{ptmk} - 1} \prod_{q \in V_t} \prod_{k=1}^{K} \pi_{ptk}^{z_{p \to q,t,k}} \pi_{ptk}^{w_{p \leftarrow q,t,k}} \, \mathrm{d}\boldsymbol{\pi}_{pt}$$

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

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

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

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

where the sum of C_{ptk} over groups k is equal to twice the number of nodes (as nodes must instantiate at least one group in each of interactions, once as a sender and once again as a receiver) in directed networks. A simple reorganization of factors (along with the fact that $s_{t,m}$ is an indicator vector, whereby $\sum_m s_{tm} x = \prod_m x^{s_{tm}}$) yields equation (4) in Section 3.2.

B Details of the Collapsed Variational Algorithm

B.1 Expectation Steps

E step 1: Z and W

To obtain the updates of the $\boldsymbol{\phi}_{p \to q,t}$ variational parameters, we begin by restricting equation (4) 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 $\log \Gamma(\alpha_{ptmk} + C_{ptk}) = z_{p\to q,t,k} \log(\alpha_{ptmk} + C'_{ptk}) + \log \Gamma(\alpha_{ptmk} + C'_{ptk})$ 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 (unnormalized) parameter vector of a multinomial distribution $\widetilde{Q}_2(\mathbf{z}_{p\to q,t} \mid \boldsymbol{\phi}_{p\to q,t})$. The update for $\mathbf{w}_{q\leftarrow p,t}$ is similarly derived.

E step 2: S

Isolating terms in Equation 4 that are not constant with respect to s_{tm} for a specific $t \neq 1$ and m, and rolling all other terms into a const., we have

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

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

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

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 $x \in \{0, 1\}$, $\Gamma(y + x) = y^x \Gamma(y)$:

$$\Gamma \left(M\eta + s_{tm} + U'_m \right)^{-1} \Gamma(\eta + s_{t+1,m} s_{tm} + s_{t-1,m} s_{tm} + U'_{mm})$$

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

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

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

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

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

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

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

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

$$
\hat{\kappa}_{Tm} \propto \exp \left[- \mathbb{E}_{\tilde{Q}_1}[\log(M\eta + U'_m)] \right] \prod_{n=1}^{M} \exp \left[\kappa_{T-1,m} \, \mathbb{E}_{\widetilde{Q}_1}[\log(\eta + U'_{nm})] \right]
$$

$$\times \prod_{p \in V_T} \left[\frac{\Gamma(\xi_{ptm})}{\Gamma(\xi_{ptm} + 2N_t)} \prod_{k=1}^{K} \frac{\mathbb{E}_{\widetilde{Q}_1}[\Gamma(\alpha_{pTmk} + C_{pTk})]}{\Gamma(\alpha_{pTmk})} \right]$$

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

B.2 Maximization steps

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C A Simulation Study

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

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

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

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

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

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

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

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

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

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

where $\gamma_1 = 0.1$.

To explore the conditions under which the model performs best, as well as those under which learning the model's various parameters can be particularly challenging, we refine this data-generating process by defining three sets of values for $\boldsymbol{B}$ and $\boldsymbol{\beta}$ designed to generate easy, medium, and hard learning scenarios. They differ in the extent to which memberships are truly mixed (with more clearly defined memberships being easier to learn), and with respect to the extent to which the blockmodels generate distinct equivalence classes of nodes (with more clearly defined block structures being easier to learn). Accordingly, each scenario's DGP is completed using the parameters in presented in Table S1.

	Easy	**Medium**	**Hard**
$g^{-1}(\mathbf{B}) =$	$\begin{bmatrix} 0.85 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$	$\begin{bmatrix} 0.65 & 0.35 \\ 0.20 & 0.75 \end{bmatrix}$	$\begin{bmatrix} 0.65 & 0.40 \\ 0.50 & 0.45 \end{bmatrix}$
$\boldsymbol{\beta}_1 =$	$\begin{bmatrix} -4.5 & -4.5 \\ 0.0 & 0.0 \end{bmatrix}$	$\begin{bmatrix} 0.05 & 0.75 \\ -0.75 & -1.0 \end{bmatrix}$	$\begin{bmatrix} 0.0 & 0.0 \\ -0.75 & -1.0 \end{bmatrix}$
$\boldsymbol{\beta}_2 =$	$\begin{bmatrix} -4.5 & -4.5 \\ 0.0 & 0.0 \end{bmatrix}$	$\begin{bmatrix} -0.05 & 0.55 \\ -0.75 & 0.75 \end{bmatrix}$	$\begin{bmatrix} 0.0 & 0.0 \\ -0.75 & 0.75 \end{bmatrix}$

Table S1: **Parameters in three different dynamic network DGPs**. The three columns correspond to three types of networks, varying in terms of inferential complexity. In turn, the rows contain the corresponding values of the blockmodel $\mathbf{B}$ and the regression coefficient vectors $\boldsymbol{\beta}$, one for each state of the HMM.

Generating a single, 9-period network under each of these scenarios results in the mixed memberships depicted in Figure S1, which shows the density of membership into the first of two groups across all nodes and time periods. While the 'easy' scenario has very clearly defined memberships of most nodes into one of the underlying groups, the 'hard' scenario has a substantial number of nodes whose membership is decidedly more mixed. The medium, more 'realistic' scenario has a non-negligible number of nodes whose membership is mixed, and a distinct group imbalance in favor of the second group.

C.1 Accuracy of estimation: mixed-memberships and blockmodels

Overall, and as expected, the accuracy with which dynMMSBM can retrieve the true mixed-membership vectors depends on the problem's complexity. The top panel of Figure S2 shows the estimated mixed-membership values against their known, true values,

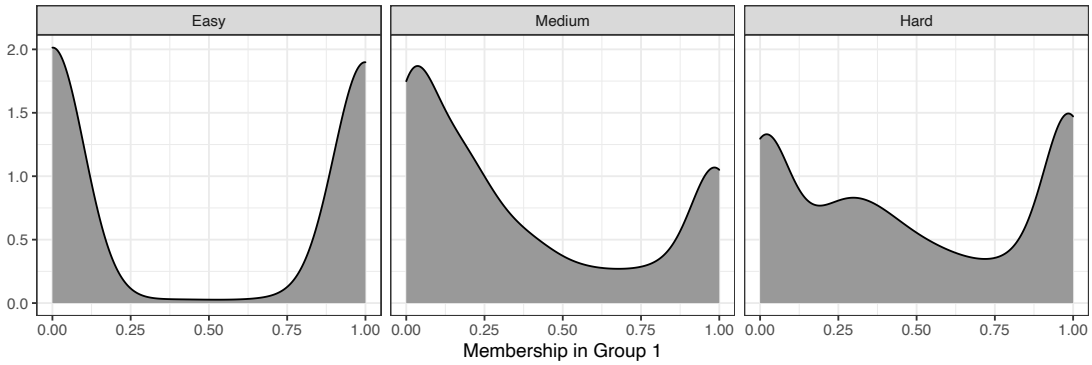

Figure S1: **Simulated mixed-memberships in synthetic networks**. The plots depict the mixed-membership vectors of nodes in three simulated networks, each with 100 nodes observed over 9 time periods. It shows the memberships of nodes in networks generated under an 'easy' DGP (i.e. one where memberships are not mixed, and in which the block structure is clear), 'hard' (i.e. one where memberships are extremely mixed, and no block structure is apparent in the network) and 'medium' (i.e. where some nodes display a mixture of group memberships, and a block structure is somewhat apparent in the network) on the left, right, and central panels, respectively.

evidencing a decrease in estimation accuracy as we move from an easy to a hard inferential task. Despite the clear deterioration, dynMMSBM is still able to produce good quality estimates even under hard inferential situations, with estimates that have a 0.94 correlation with their true values.

The model is also able to accurately estimate the blockmodel structure, as the bottom row of Figure S2 reveals. For each cell of the blockmodels, the true probability of an edge between members of any two groups is shown in white letters, while the cell itself is colored in accordance to the corresponding estimated values. Once again, and although the quality of these estimates (predictably) decreases as the inferential complexity of the scenario increases, the estimation error remains low.

C.2 Estimation accuracy: regression coefficients

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

Figure S3 shows, for each time period, the distribution of estimated effect sizes of the monadic predictor and intercepts for the regression of membership into the second latent group (as boxplots), along with the true parameter values (shown as a red "x"). We obtain estimates for each time period by computing the weighted average of estimated parameters in the two hidden Markov states, using estimated marginal probabilities over states in each time period as our weights. The model is typically able to identify the underlying

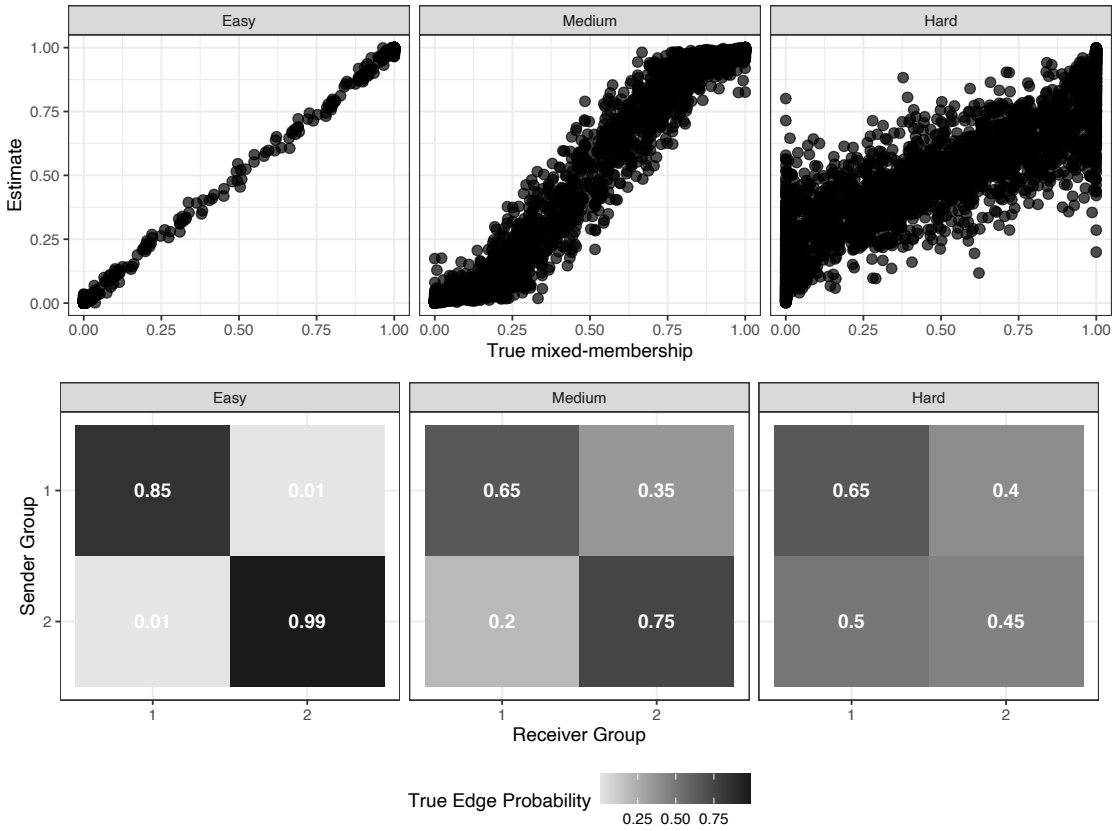

Figure S2: **Estimation accuracy**. For each DGP scenario, the figure shows the estimated mixed-membership vectors (top row) and the estimated blockmodels (bottom row) against their known values (indicated by the white numbers in each cell of the blockmodel for the bottom row). Overall, accuracy of retrieval both sets of parameters depends on the complexity of the learning problem, although recovery is generally very good, even under 'hard' inferential conditions.

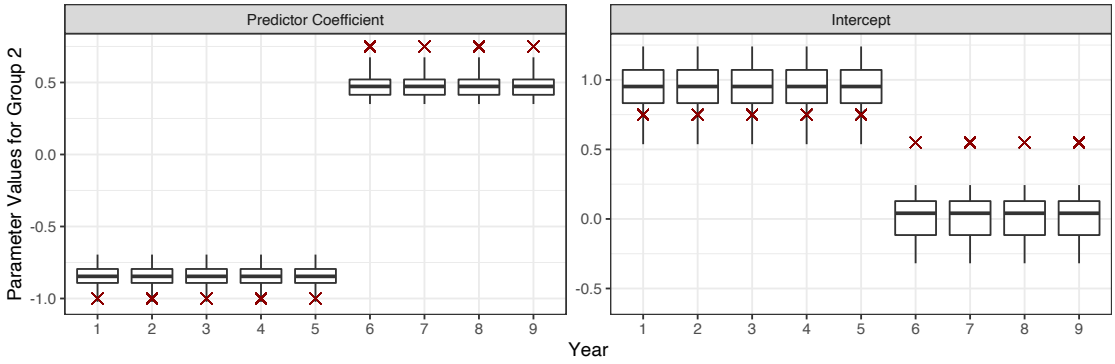

Figure S3: **Estimated parameters of block membership regression**. The figure shows, for each time period, the HMM-weighted effect of a continuous predictor on the probability of instantiating latent group 2 (left panel), and the HMM-weighted intercept of the corresponding regression line (right panel), estimated on 100 networks generated according to our medium DGP. In each instance, the red "x" indicates the true parameter value for that time period, given a known HMM state.

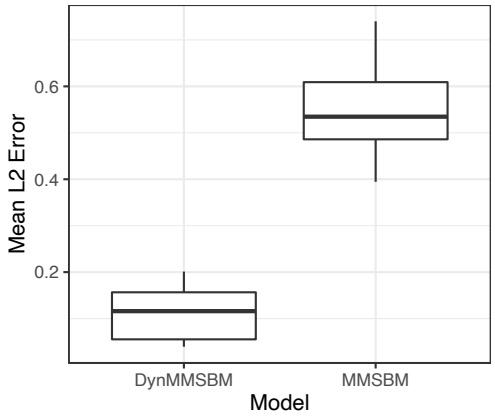

Figure S4: **Error for estimated mixed membership vectors**. The figure shows average L_2 distances between estimated and true mixed-membership vectors for all nodes in each of 100 replicated dynamic networks. On the left, estimates are generated using dynMMSBM. On the right, estimates are generated using the canonical MMSBM, fit separately to the nine time periods in each simulated network.

Markov state that generated the networks, which in turn translates into correctly estimated (albeit regularized) effects of the monadic covariate on membership probabilities. Quality of recovery for regression parameters associated with a given block depends heavily on the extent to which that block is commonly instantiated in the network. And although changes in intercepts across time periods are also correctly recovered, the intercepts themselves tend to be overestimated. This phenomenon, which we found to be common in all our simulations, is likely the result of the difficulty in pinning down the precision of the latent membership vectors. Despite these issues, the mean of the memberships is correctly recovered (as shown earlier in Figure S2).

C.3 Comparison to non-dynamic MMSBM

Finally, and to further evaluate the benefits of modeling the dynamic nature of the network, we estimate a separate MMSBM model to the networks in each time period, and compare their estimated mixed-memberships to those of a single dynMMSBM estimated on the full set of networks. In both cases, we omit all covariates, but estimate the $\boldsymbol{\alpha}_{ptm}$ parameters associated with the mixed-membership vectors. After estimating both sets of models on each of the 100 replications of the "medium" networks, we compute the average L_2 error in estimated mixed-memberships across nodes. The results are presented in Figure S4.

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

thus contributing to the correct estimation of the latent memberships across simulations.

D Additional Empirical Results

D.1 Model forcast accuracy results, with different numbers of latent groups

Table S2 presents out-of-sample (forcast) errors for models with different numbers of latent groups. We estimate models for networks observed between 1816 and 2010, and compute the area under the receiver operating characteristic curve for forecast networks in 2009 and 2010. In contrast, the results reported in the main text all use the entire time span (1816-2010).

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

Table S2: **Out of Sample Prediction, Different Latent Groups**. The table displays the area under the ROC curve (AUROC) and associated standard error for specifications with 2-7 Latent Groups. Each model is fit on data from 1816-2008 and used to forecast conflict in the period 2009-2010.

D.2 Model including geographic region indicators

As an alternative way of capturing geographic determinants of militarized disputes, we estimate a model that includes regional indicators as predictors.Figure S5 presents the estimated blockmodel when using this alternative specification. In turn, Figure S6 shows the evolution of memberships into estimated latent groups under this specification. Finally, Table S3 shows countries with the highest average estimated membership in each of the groups during the Cold War for this alternative specification. Many group assignments are similar to the main specification. Nodes with high membership in Groups 4-6, for example, reflect many of the same countries as in the specification reported in the main text. The most notable difference in this specification is the grouping of Western and some Eastern bloc countries into a single, more belligerent latent group (Group 3): the United States, United Kingdom, and West Germany are part of this group along with Russia and Poland (not shown).

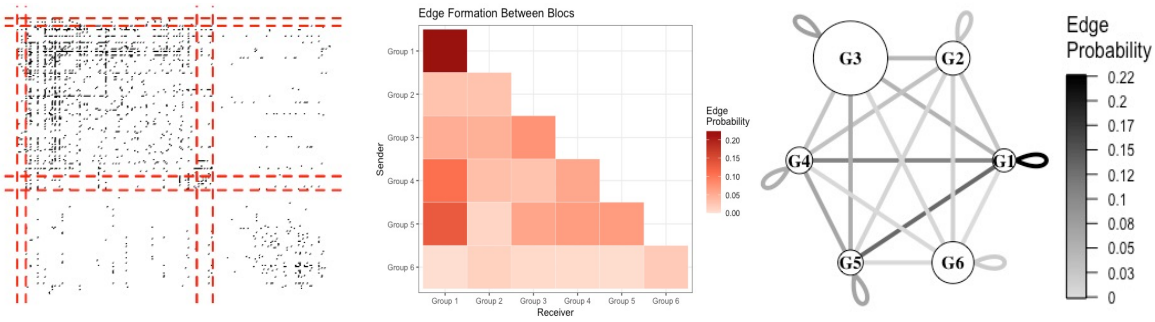

Figure S5: **Estimated blockmodel in the conflict network (regional model)**. Blockmodel visualizations for a specification including an indicator for state region. The left panel displays the adjacency matrix of militarized disputes. The middle panel displays the estimated probability of conflict between groups as a heat map. The right panel is a network graph summarizing the estimated blockmodel. The blockmodel in this specification is moderately correlated (0.104) with the primary model.

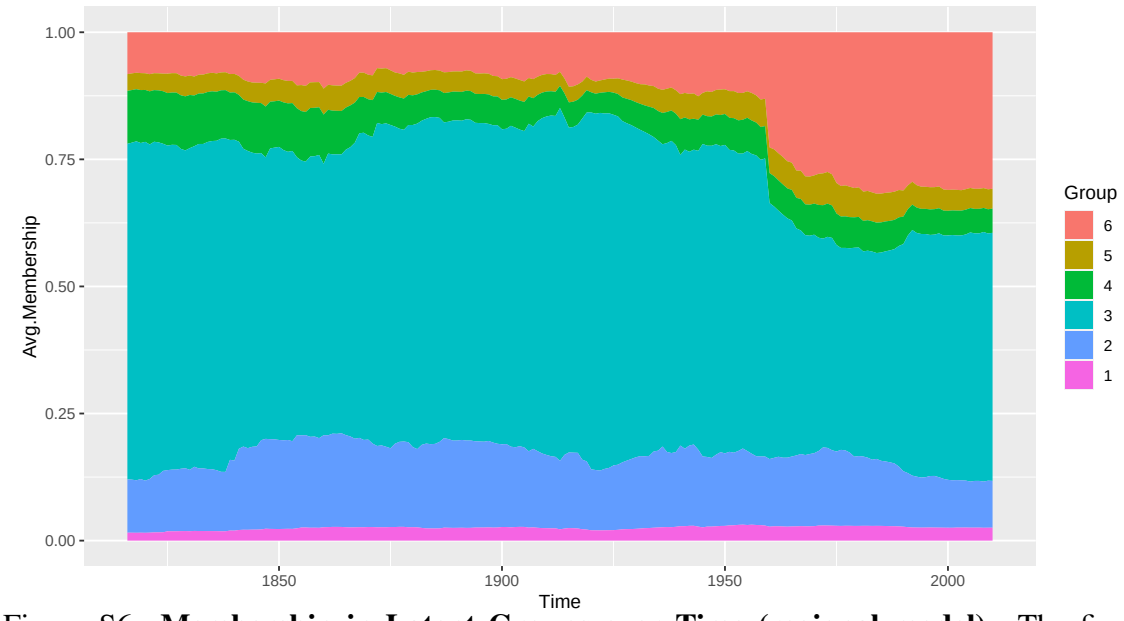

Figure S6: **Membership in Latent Groups over Time (regional model)**. The figure shows the average proportion of membership in six latent groups for each year from 1816–2010. The estimated evolution of membership in this specification is similar in some respects to the primary specification (e.g., the sizeable increase in Group 6 membership in later decades), but differs in others (e.g., Group 2 membership is smaller in earlier years). Notably, this specification does not experience transitions in the hidden Markov state. This may be attributable to the addition of indicators for state region, which are static over time.

Group 1	**Group 2**	**Group 3**
0.083 Lebanon	0.571 Paraguay	0.998 USA
0.079 North Yemen	0.46 Bolivia	0.996 India
0.075 Yemen	0.377 Argentina	0.992 Canada
0.074 South Yemen	0.357 Ecuador	0.976 West Germany
0.074 Bhutan	0.348 Uruguay	0.975 UK
0.073 Tunisia	0.345 Chile	0.974 Germany
0.073 Libyan	0.332 Peru	0.966 Italy
0.073 Jordan	0.326 Mongolia	0.963 Japan
0.072 Maldives	0.324 Korea North	0.959 France
0.072 Kuwait	0.324 Albania	0.955 Pakistan
0.072 United Arab Emirates	0.303 Cambodia	0.947 Sri Lanka
0.07 Morocco	0.289 Taiwan	0.944 Turkey
0.07 Syria	0.267 Laos	0.94 Russia
0.069 Algeria	0.252 Vietnam	0.937 Netherlands
0.069 Iraq	0.246 Myanmar	0.937 Belgium
Group 4	**Group 5**	**Group 6**
0.212 Bhutan	0.282 Bahrain	0.948 Seychelles
0.158 Maldives	0.282 Oman	0.941 Sao Tome Principe
0.155 Bahrain	0.281 Qatar	0.927 Zanzibar
0.154 Oman	0.269 Jordan	0.886 Liechtenstein
0.154 Tunisia	0.268 Saudi Arabia	0.878 Comoros
0.153 Qatar	0.266 United Arab Emirates	0.875 Gambia
0.153 Jordan	0.264 Tunisia	0.863 Cape Verde
0.153 Kuwait	0.264 Kuwait	0.842 Equatorial Guinea
0.153 South Yemen	0.263 South Yemen	0.841 Swaziland
0.152 United Arab Emirates	0.259 Libyan	0.834 Botswana
0.15 Libyan	0.258 Algeria	0.823 St Kitts-Nevis
0.149 North Yemen	0.249 Morocco	0.82 Lesotho
0.149 Saudi Arabia	0.247 Djibouti	0.798 Gabon
0.147 Algeria	0.246 North Yemen	0.797 Dominica
0.145 Morocco	0.236 Syria	0.796 Antigua-Barbuda

Table S3: **States with Highest Membership in Latent Groups, Cold War period (regional model)**. Average group membership in the years 1950-1990 is reported beside the state name for the top 15 states in each latent group.

D.3 Models estimated on entire dataset, with different numbers of latent groups

Figures S7 and S8 show the estimated blockmodels of models defined with 5 and 7 latent groups, respectively. The blockmodels mostly reflect similar relationships and memberships to those of the model reported in the main text, which uses 6 latent groups.

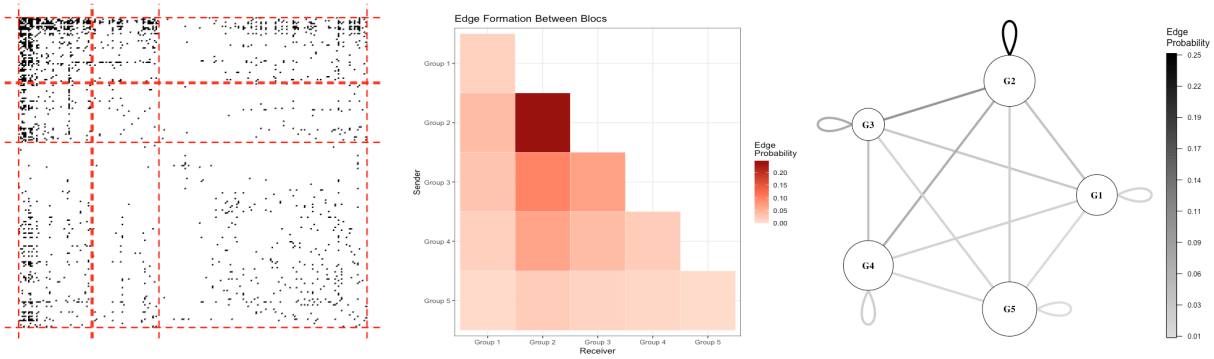

Figure S7: **Estimated blockmodel in the conflict network (5-group specification)**. The left panel displays the adjacency matrix of militarized disputes between 216 states. The middle panel displays the estimated probability of conflict between groups as a heat map. The right panel is a network graph summarizing the estimated blockmodel.

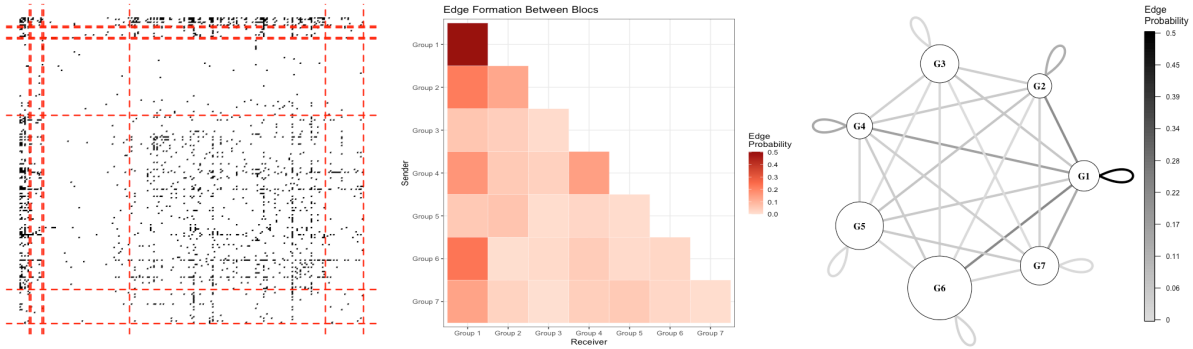

Figure S8: **Estimated blockmodel in the conflict network (7-group specification)**. The left panel displays the adjacency matrix of militarized disputes between 216 states. Dotted lines separate states by estimated group membership; some groups are not visible in the adjacency matrix since they have very low membership. The middle panel displays the estimated probability of conflict between groups as a heat map. The right panel is a network graph summarizing the estimated blockmodel.

D.4 Estimated blockmodel, in table format

Table S4 presents the estimated blockmodel displayed in graphical form in Figure 1 of the main text.

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

Table S4: **Group-Level Edge Formation Probabilities**. The table displays the probability of interstate conflict between nodes that instantiate membership in each of six latent groups. The diagonal shows rates of intra-group conflict and off-diagonal shows rates of conflict between groups.

D.5 Countries with highest membership probabilities

Table S5 shows, for each of the 6 estimated latent groups in the main text, the countries with the highest average membership probabilities during the 1950-1990 time period. The group assignments are consistent with known geopolitical coalitions in the Cold War, with Western allies in Group 1, Eastern bloc countries clustered in Group 2, Western-leaning neutral states in Group 3, and states engulfed in proxy conflicts in Group 4.

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

Table S5: **States with Highest Membership in Latent Groups, Cold War period**. To identify the states with highest membership in each latent group, we average over each states' latent membership probabilities in the years 1950-1990. Average group membership is reported beside the state name for the top 15 states in each latent group.

D.6 Estimated coefficients associated with hidden Markov state 2

The main text focuses on effects during time periods associated with hidden Markov state 1. For completeness, we present the results associated with hidden Markov state 2.

Predictor	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
INTERCEPT	10.420	16.239	12.357	11.622	4.457	4.549
	(1.021)	(1.021)	(1.021)	(1.021)	(1.060)	(1.064)
POLITY	−0.005	−0.137	0.209	0.052	−0.201	−0.157
	(0.914)	(0.913)	(0.913)	(0.913)	(1.047)	(1.062)
MILITARY	0.363	1.017	0.237	0.163	−0.443	−0.556
CAPABILITY	(1.063)	(1.062)	(1.062)	(1.061)	(1.062)	(1.064)
N nodes: 216; N dyad-years: $842,685$; N time periods: 195						
Lower bound at convergence: $-527,587.7$						

Table S6: **Estimated Coefficients and their Standard Errors, Markov State 2**. The table shows the estimated coefficients (and standard errors) of the two monadic predictors for each of six latent groups in the second Markov state. The estimated coefficients for cubic splines and indicators for variable missingness are not shown.

D.7 Results of model estimated using online estimation approach

The main text uses the entire set of observed networks, from 1816 to 2010, to estimate the model's parameters. As all time periods are treated as observed at the time of analysis, this can be thought of as performing *batch* data analysis. An alternative approach considers data streaming in in sequence — either year by year or (as we do illustrate here) after a mini-batch of years has been observed. 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 the model estimated by adding the next window (1821-1825). We repeat until all years are included.

Figure S9 we show how estimated memberships evolve over the entire period under study using this online estimation strategy. Membership patterns are positively correlated with the model reported in the main text (0.516), but the evolution of membership differs in several ways. Group 2 is significantly larger throughout the period, and the late increase in Group 6 (beginning in 2005) is more pronounced. Figure S10 shows the same evolution, disaggregated by some of the countries we used in the main text. Many previously observed structural breaks are apparent in these estimates (e.g., Russia at the end of the Cold War, Japan in 1945), while others are attenuated (Cuba in the 1950s) or absent (Iraq in 1991). Finally, Table S7 reports estimated coefficient (and associated standard errors) for the final time window in the online estimation approach (viz. years 2006-2010).

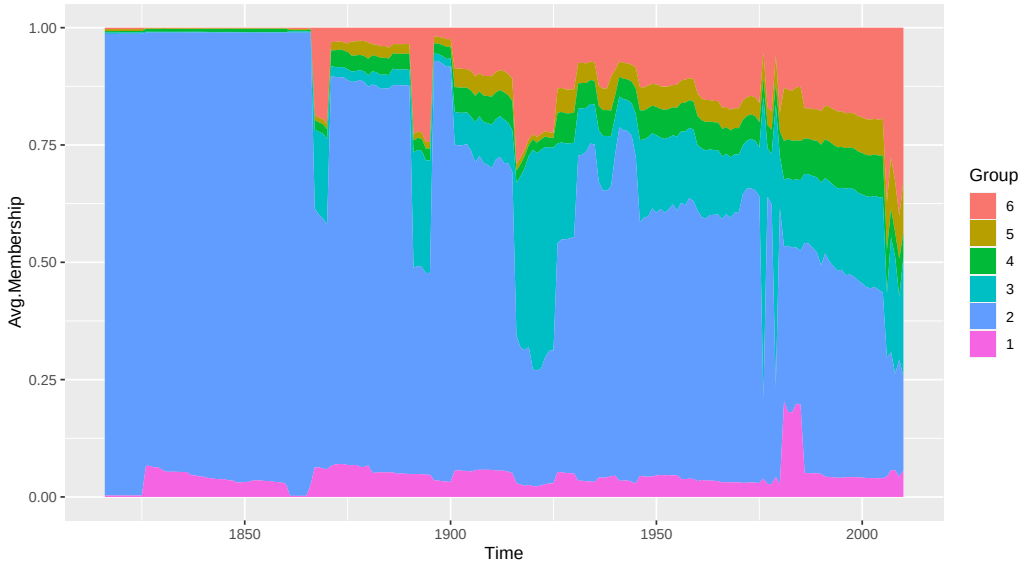

Figure S9: **Membership in Latent Groups over Time (online update model)**. The figure shows the average proportion of membership in six latent groups for each year from 1816–2010. Estimates are derived from specifications using expanding five-year windows.

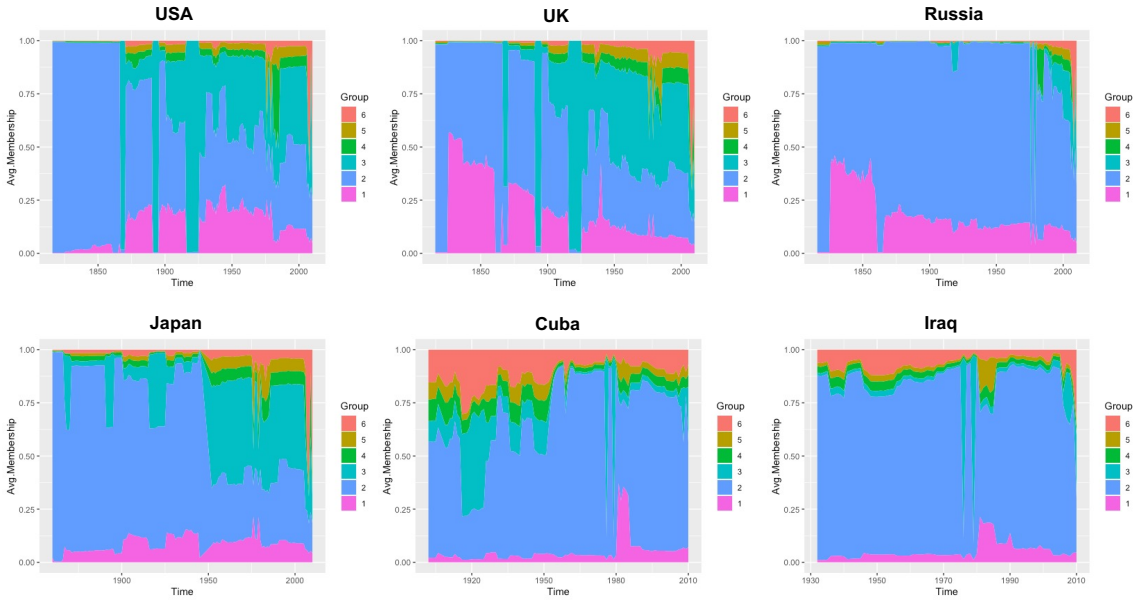

Figure S10: **Average Node Membership over Time, Select States (online update model)**. The figure shows, for six states, the average rate of membership in six latent groups in each year the state is present in the network. Estimates are derived from specifications using expanding five-year windows.

Predictor	Dyadic	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
INTERCEPT		−0.997	1.747	−1.354	−1.817	−2.602	−2.384
		(0.101)	(0.101)	(0.100)	(0.102)	(0.102)	(0107)
POLITY		−0.009	−0.081	0.104	0.024	0.027	0.063
		(0.102)	(0.101)	(0.089)	(0.105)	(0.110)	(0.151)
MILITARY CAPABILITY		−0.036	0.178	−0.170	−0.171	−0.273	−0.332
		(0.103)	(0.101)	(0.109)	(0.106)	(0.101)	(0.123)
BORDERS	2.374						
	(0.003)						
DISTANCE	-0.0001						
	(0.000)						
ALLIANCE	0.082						
	(0.003)						
IO CO-MEMBERS	0.002						
	(0.000)						
PEACE YRS	-0.019						
	(0.000)						
N nodes: 216; N dyad-years: 842, 685; N time periods: 195							
Lower bound at convergence: −527, 686.5							

Table S7: **Estimated Coefficients and their Standard Errors (online update model)**. The table shows the estimated coefficients (and standard errors) of the two monadic predictors for each of six latent groups, as well as those of the dyadic predictors for edge formation. Estimates are derived from specifications using expanding five-year windows. We report coefficient estimates for the final model (2006-2010).

D.8 Effects of decreasing POLITY

In the main text, we focus on the effects of increasing POLITY scores (i.e. making countries more democratic). The proportion of states with the maximum polity score generally increases over time (approximately 17.5% of states in 2010), so the effects we see may be driven by a sort of celing effect, as these countries are as democratic as they can be. To ensure the estimated effect of POLITY is not driven by a ceiling effect, we analyze the effects of decreasing POLITY scores. The results are substantively identical to those reported in the main text, mirroring the pattern shown in Figures 5 and 6.

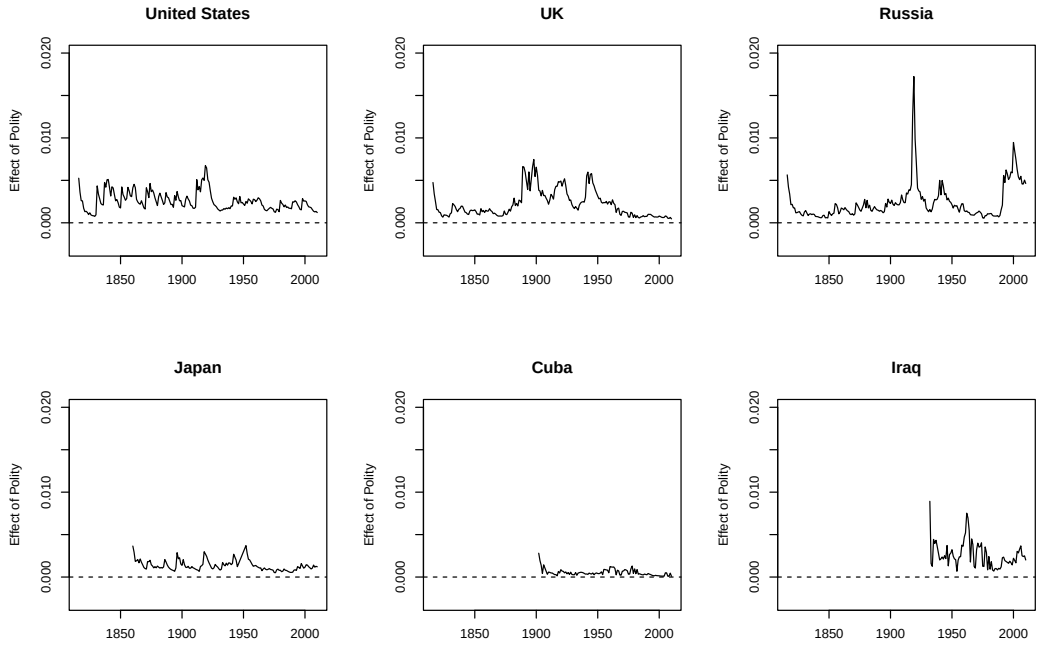

Figure S11: **Effect of Decrease in Polity over Time, Select States**. The figure shows the estimated change in the probability of interstate conflict over time if a country's `POLITY` score is decreased by one standard deviation (6.78) from its observed value (down to a minimum of -10).

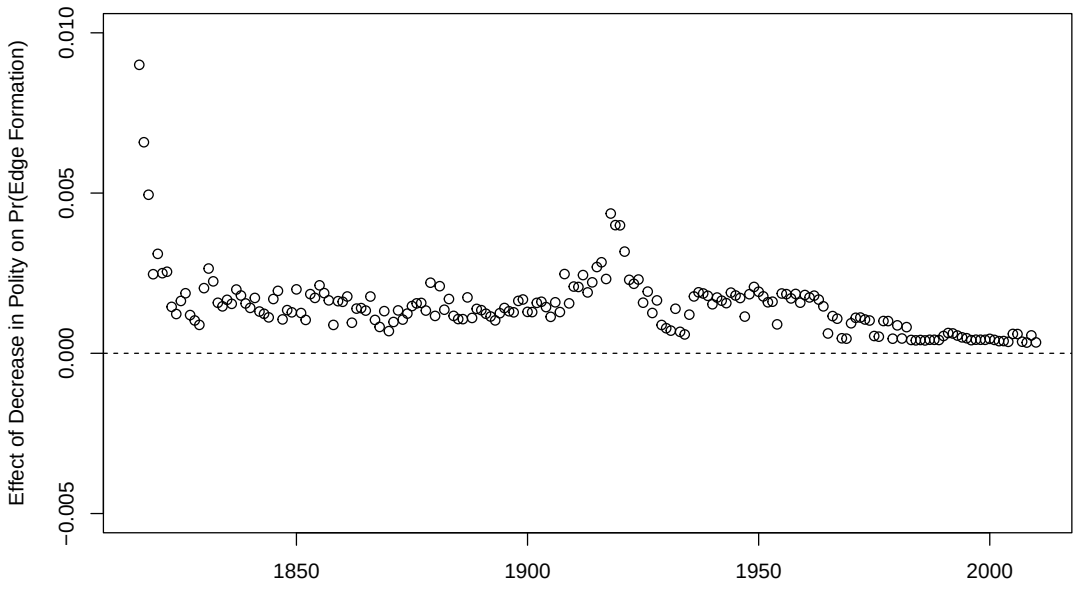

Figure S12: **Estimated Aggregate Effect of Decrease in Polity over Time**. The figure shows the estimated average change in the probability of interstate conflict when countries' POLITY scores are decreased by one standard deviation (6.78) down to the minimum POLITY score.

D.9 Comparison with Logistic Regression

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

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

We evaluate the predictive accuracy of both models by comparing their predictions with the observed pattern of conflict in 2009–2010. First, we conduct a Diebold-Mariano test of comparative forecasting accuracy (Diebold and Mariano, 1995; Harvey, Leybourne, and Newbold, 1997). The test, which compares mean-squared forecasting error of the two methods, indicates that the dynMMSBM significantly outperforms the logit model in dispute prediction (DM statistic = -2.12, $p = 0.034$).

Second, we compare the receiver operating characteristic curves (ROCs) for each model. We display the ROC curves in Figure S13 and show the area under the ROC curves in S8. By this criterion, the dynMMSBM continues to outperform the logit model but only marginally. The dynMMSBM has a larger area under the ROC curve, though the difference is not statistically significant.

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

Table S8: **Out of Sample Prediction, dynMMSBM vs. Logit**. The table displays the area under the ROC curve (AUROC) and associated standard error for the two models. Each model is fit on data from 1816-2008 and used to forecast conflict in the period 2009-2010.

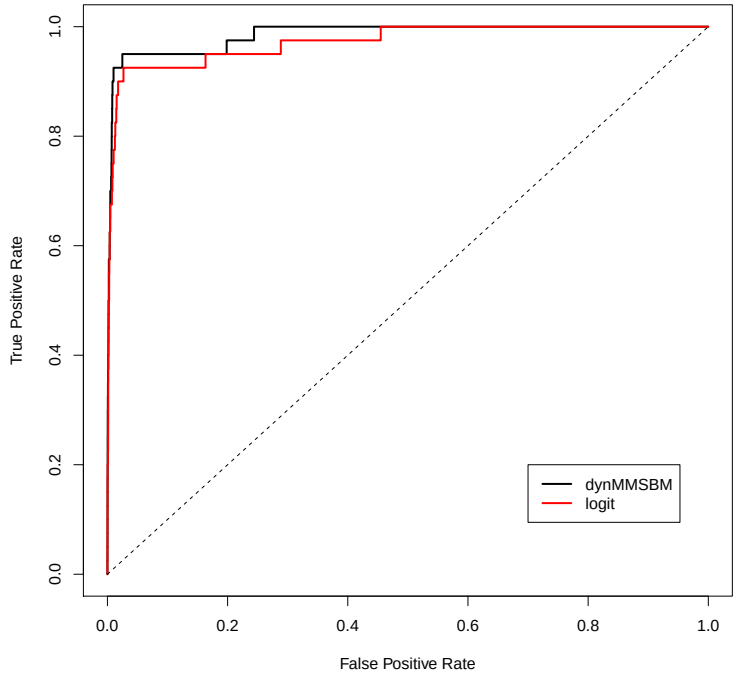

Figure S13: **ROC Curve: Logit, Dynamic Mixed-membership SBM Models**. To perform the forecast, we exclude the final two years (2009-2010) from the dataset and estimate each model on the preceding years (1816-2008). Then we predict the missing years based solely on the covariate data.

References

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

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

Goldstone, Jack A et al. (2010). "A Global Model for Forecasting Political Instability". *American Journal of Political Science* 54.1, pp. 190–208.

Harvey, D., S. Leybourne, and P. Newbold (1997). "Testing the equality of prediction mean squared errors. International Journal of forecasting". *International Journal of forecasting* 13.2, pp. 253–263.