Dynamic Bipartite Stochastic Blockmodel Regression for Network Data

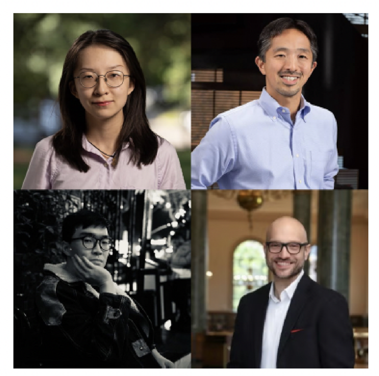
Application to State and Intergovernmental Organization Networks

Qi Liu [1] Ruofan Ma [1] Santiago Olivella [3] Kosuke Imai [1,2]

[1]Department of Government, Harvard University [2]Department of Statistics, Harvard University [3]Department of Political Science, UNC

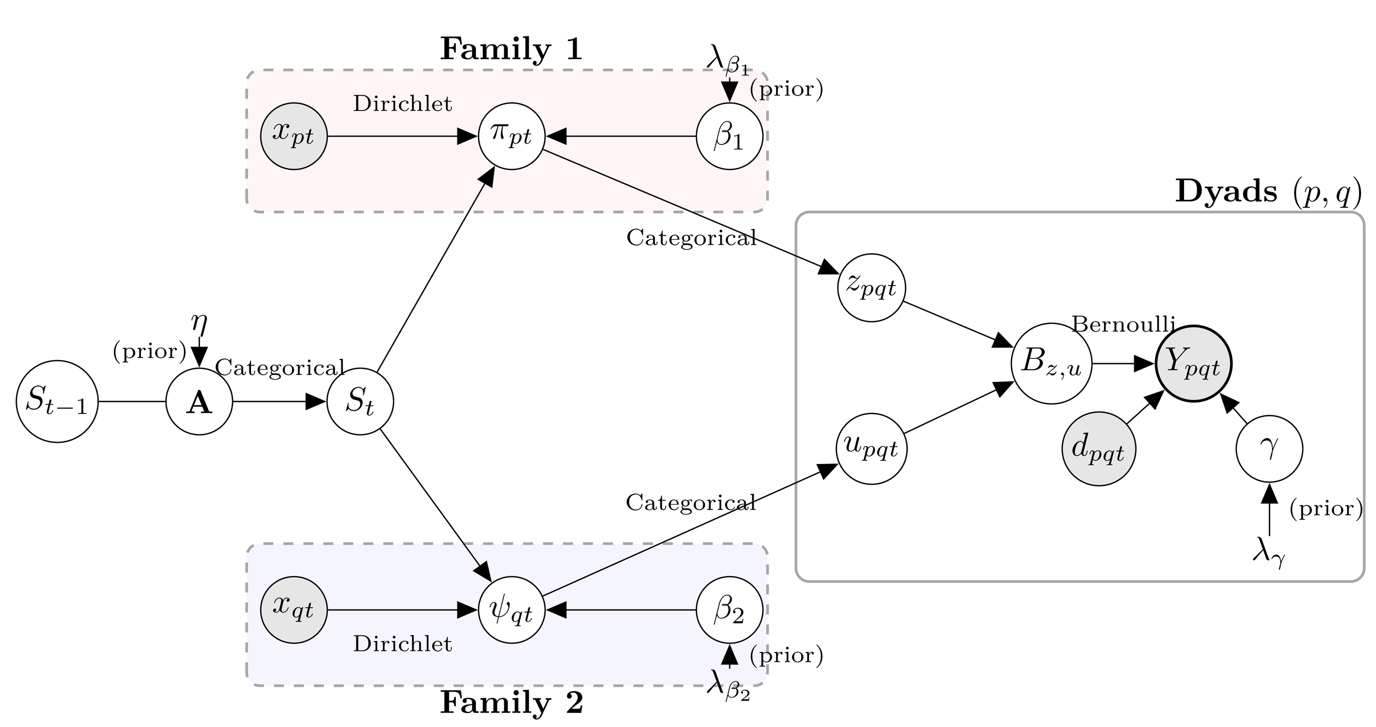
Motivation

- **Most political networks are:**
  - **Bipartite** — two node types, ties only across types *(e.g., states–treaties, legislatures–bills, lobbyists-politicians)*
  - **Dynamic** — ties evolve over time
- **Problem:** Use static/ projected unipartite models $\rightsquigarrow$ bias and spurious clustering
- **Our contribution: Dynamic Bipartite MMSBM**

	Dynamic: No	Dynamic: Yes
Bipartite: No	Box-Steffensmeier et al. (2019); Siegel & Badaan (2020); Weschle (2018); Naidu et al. (2021); Cruz et al. (2020); Weschle (2019); Bouchcer & Thies (2019); Bodea & Hicks (2015); Jiang & Zeng (2020); Shaffer (2022); Franzese et al. (2012); Lo et al. (2023); Goddard (2018); Edgerton (2024); Battaglini et al. (2020); Nyhan & Montgomery (2015); Box-Steffensmeier et al. (2020); Oklobdzija (2024); Aarøe & Peterson (2020); Kim et al. (2019); Blair et al. (2022); Cho & Fowler (2010); Cranmer & Desmarais (2011); Fishman and Davis (2022); Abi-Hassan et al. (2023)	Kim et al. (2020); Gilardi et al. (2020); Nieman et al. (2021); Harden et al. (2023); Uppala and Desmarais (2023)
Bipartite: Yes	Sweet (2021); Kim & Kunisky (2020)	

Model Setup

- **Dynamic bipartite graph** $G_t = (V_{1,t}, V_{2,t}, Y_t)$
  - Nodes $p \in V_{1,t}$ and $q \in V_{2,t}$. $Y_{pqt} = 1$ if an edge from p to q exists at time t, $Y_{pqt} = 0$ otherwise
  - s_t: latent state; A: transition matrix
  - π_{pt}, ψ_{qt}: mixed membership; $z_{pq,t}, u_{pq,t}$: (p, q) interaction specific group indicators
  - $\mathbf{B}$: $K_1 \times K_2$ block matrix. B_{gh}: log-odds of an edge forming between latent groups g and h
  - $\mathbf{x}_{pt}, \mathbf{x}_{qt}$: monadic covariates (coefficient: $\boldsymbol{\beta}_{1gm}, \boldsymbol{\beta}_{2hm}$); $\mathbf{d}_{pqt}$: dyadic covariates (coefficient: γ)



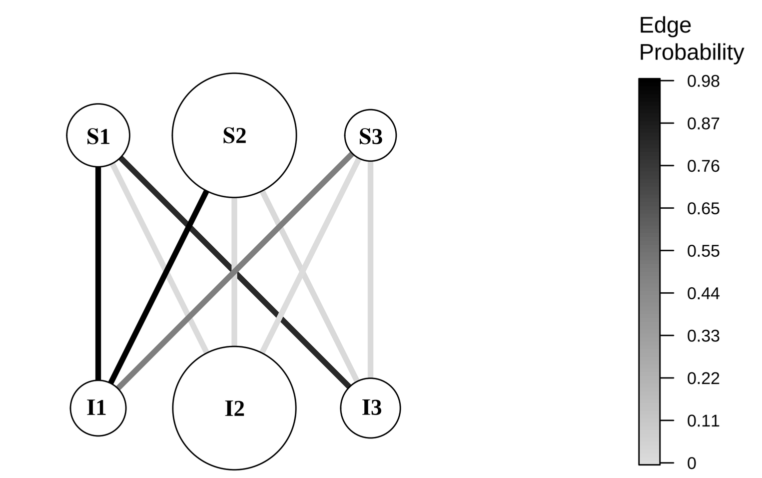
Simulation

Dynamic bipartite networks over 50 periods, 100 nodes per family:

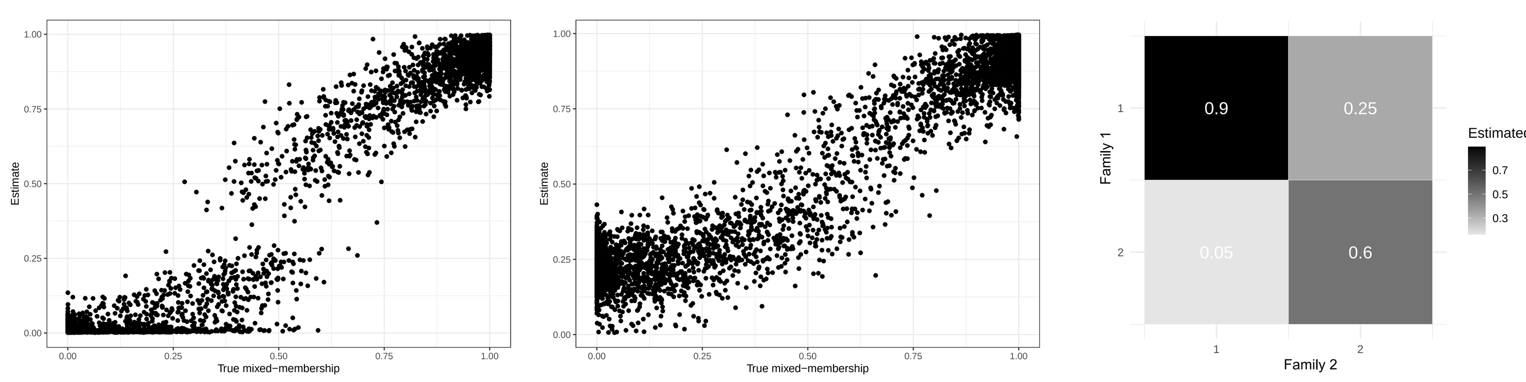
- **Latent states (HMM):** Periods 1–25 in state 1; 26–50 in state 2
- **Covariates:**
  - 1 monadic covariate per family: $x_{pt}, x_{qt} \sim 0.5 \cdot \mathcal{N}(-1.25, 0.09) + 0.5 \cdot \mathcal{N}(1.25, 0.09)$
  - 1 dyadic covariate: $\mathbf{d}_{pqt} = \mathbf{d}_{pq,1} + \epsilon_{dt}$, where $\epsilon_{dt} \sim \mathcal{N}(0, 1)$
- **Difficulty levels:** *Easy, Medium, Hard* — vary B and β
- **Model recovers:**
  - **Mixed-membership vectors, group structure, regression parameters**
  - **Good performance across difficulty levels**

	Easy	Medium	Hard
logit$^{-1}(\mathbf{B})$	$\begin{bmatrix} 0.90 & 0.01 \\ 0.10 & 0.60 \end{bmatrix}$	$\begin{bmatrix} 0.90 & 0.05 \\ 0.25 & 0.60 \end{bmatrix}$	$\begin{bmatrix} 0.90 & 0.10 \\ 0.40 & 0.60 \end{bmatrix}$
β_1	$\begin{bmatrix} 1.50 & -0.50 \\ 1.25 & -1.50 \end{bmatrix}$	$\begin{bmatrix} 1.00 & -0.25 \\ 0.75 & -1.00 \end{bmatrix}$	$\begin{bmatrix} 0.60 & -0.05 \\ 0.50 & -0.55 \end{bmatrix}$
β_2	$\begin{bmatrix} -3.00 & 1.00 \\ 7.25 & -4.25 \end{bmatrix}$	$\begin{bmatrix} -0.50 & 0.25 \\ 1.25 & -1.25 \end{bmatrix}$	$\begin{bmatrix} -1.05 & -0.55 \\ 1.55 & -0.25 \end{bmatrix}$

Easy, medium, to hard DGPs: more similar entries in B and more mixed memberships

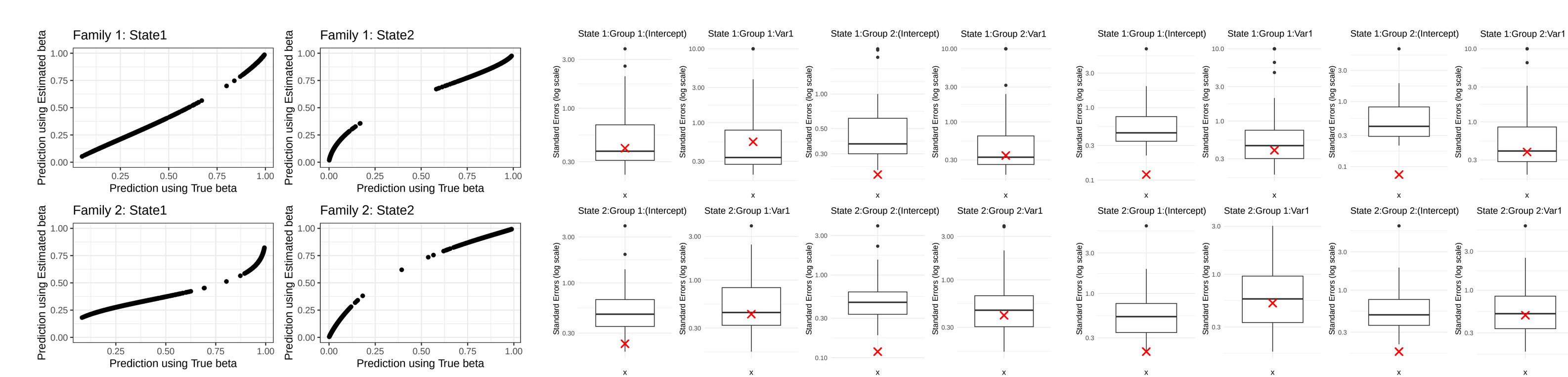


Medium Case Results:



a) Mixed-membership (family 1) b) Mixed-membership (family 2) c) Blockmodel

a) and b): estimated mixed-membership vectors align with known values;
c): estimated blockmodel match known values (white numbers)



a) Mixed-membership prediction with $\hat{\beta}$ b) SE of $\hat{\beta}$ (family 1) c) SE of $\hat{\beta}$ (family 2)

a): mixed-membership predicted using estimated coefficients match prediction using true coefficients.
b) and c): SEs across 100 simulated networks covers sd($\hat{\beta}$) (red crosses) well.

Application: State-IO Network, 1965-2014

- **Data**: Yearly state-IGO membership data between 1965-2014, covering 200 countries and regions and 471 IGOs (Pevehouse et al., 2020; Davis & Pratt, 2021).
- **Parameters**: Three groups for both the state (**S**) and IGO (**I**) families. One latent state.

Group Labels Based on Estimated Block Model

- **S1 (Internationalists)**: Most likely to instantiate links to a large number of IGOs in I1 and I3.
- **S2 (Opportunists)**: Most populus group in the state family. Very likely to instantiate links to IGOs in I1, unlikely to interact with I2 and I3.
- **S3 (Isolationists)**: A small number of states that has a low likelihood to interact with IGOs across groups.

- **I1 (Universal IGOs)**: Attract a large number of states across groups.
- **I2 (Trivial IGOs)**: Large number of IGOs that are unlikely to instantiate links with any groups of state.
- **I3 (Exclusive IGOs)**: A small number of IGOs that only interact with states in S1.

Validating Group Labels

The **covariate effects align with our group labels**. For instance:

- Rich, democratic states that are geo-politically aligned with the US are more likely to have larger share of mixed-membership in S1.
- IGOs whose average members are more US-aligned and cross-regional are more likely to have a larger share of mixed-membership in I3.

Predictor	S1	S2	S3	I1	I2	I3
UN IP	-0.59	-2.26	-0.83			
	(0.29)	(0.26)	(0.25)			
V-Dem	2.95	3.97	0.68			
	(0.63)	(0.72)	(0.71)			
GDPpc	0.25	0.02	-0.05			
	(0.07)	(0.06)	(0.08)			
Europe	1.33	-2.61	0.37			
	(0.50)	(0.47)	(0.46)			
Ideal point (lagged)				-2.70	-2.73	-0.39
				(0.29)	(0.28)	(0.25)
Regional IO				0.72	0.80	-0.05
				(0.51)	(0.57)	(0.53)
Mem. Size (lagged)				0.00	-0.05	-0.01
				(0.01)	(0.01)	(0.01)
Salient IO				-0.16	-0.12	-0.23
				(0.57)	(0.57)	(0.63)
Number of Dyads	1,833,000					

Dynamic changes in mixed-membership align with related political events: