Statistical Inference for Complex Dynamic Networks

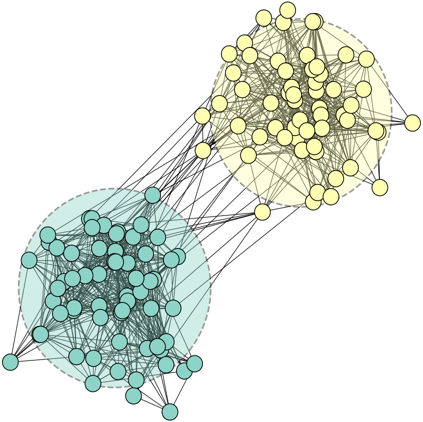
Joshua Daniel Loyal

jloyal2@illinois.edu

Website: joshloyal.github.io

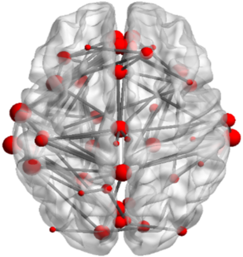
Department of Statistics
University of Illinois at Urbana-Champaign

University of Texas at Dallas
December 6th, 2021



Networks in (Data) Scientific Problems

Social Sciences



Biology

Economics and Political Science

Outline

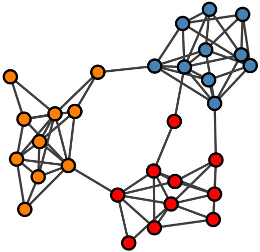
Part I: An Eigenmodel for Dynamic Multilayer Networks

- Background and Motivation
- The Model
- Theoretical Results Establishing Parameter Identifiability
- Variational Inference
- Simulation Studies
- Real Data Applications

Part II: Other Projects and Future Research Directions

Motivation

Complex systems are dynamic and consist of multiple correlated relations.

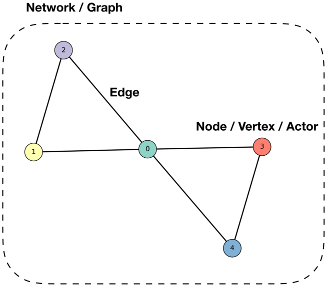


Social Networks: friendship, coworker-ship, mentorship, etc.

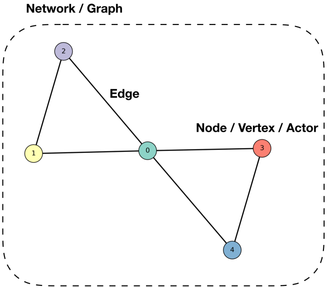
Social Media: Liking, replying to, and re-tweeting users.

International Relations: offering aid, verbal condemnation, military conflict, and others.

Network Data



Network Data



Adjacency Matrix

$$\begin{array}{c|c} & \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} & \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

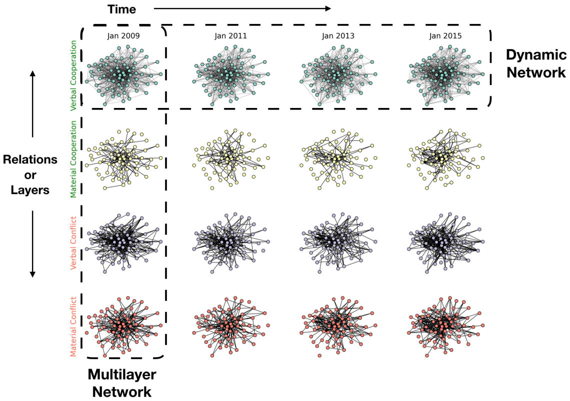
Degree
4
2
2
2
2

Degree: Number of edges connected to a node.

Dynamic Multilayer Networks

Multiple node-aligned graphs that co-evolve over time.

- **Example** (ICEWS): Graphs measuring whether two countries had a {verbal cooperation, material cooperation, verbal conflict, material conflict} on a given month.



Multiple Time-Varying Network Data

Dynamic Multilayer Networks:

A collection of $n \times n$ adjacency matrices $\mathbf{Y}_t^k$ collected over $1 \leq t \leq T$ time periods for each layer $1 \leq k \leq K$. Each $\mathbf{Y}_t^k$ has elements Y_{ijt}^k:

$$Y_{ijt}^k = \begin{cases} 1, & (i, j) \text{ are connected at time } t \text{ in layer } k, \\ 0, & \text{otherwise.} \end{cases}$$

Example:

$Y_{ijt}^k = 1$: country i and country j had a $\{$verbal cooperation, material cooperation, verbal conflict, material conflict$\}$ on the tth month.

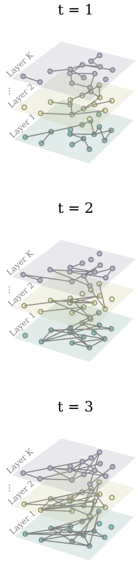
Inference for Dynamic Multilayer Networks

Two main questions:

- Inferences about **individual** network time-series:
  Forecasting future edges and graph properties, smoothing graph statistics, change-point detection
- Inferences about **common** structure:
  Community detection, graph similiarty across layers, e.g., clustering

Challenges in dynamic multilayer network analysis:

- Network heterogeneity
- High dimensionality
- Computational scalabiliity
- Proper uncertainty quantification



Previous Approaches

- **A stochastic actor oriented model** (Snijders et al., 2013). Requires **careful feature engineering** and only quantifies **local structure**.
- **Multilinear tensor autoregression** (Hoff, 2015). Only developed for real-valued networks and **very high-dimensional**.
- **A Bayesian nonparametric model** (Durante et al., 2017). **Lacks interpretability** and **not scalable** to networks with more than a dozen nodes and time-points.

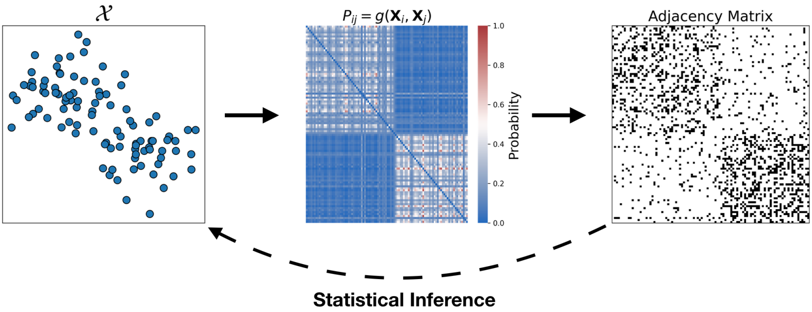
Latent Space Models for Networks (Hoff et al., 2002)

- Nodes are represented with latent positions in $\mathbb{R}^d$

$$\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^{\mathrm{T}} \in \mathbb{R}^{n \times d}.$$

- Edges are conditionally independent given latent positions

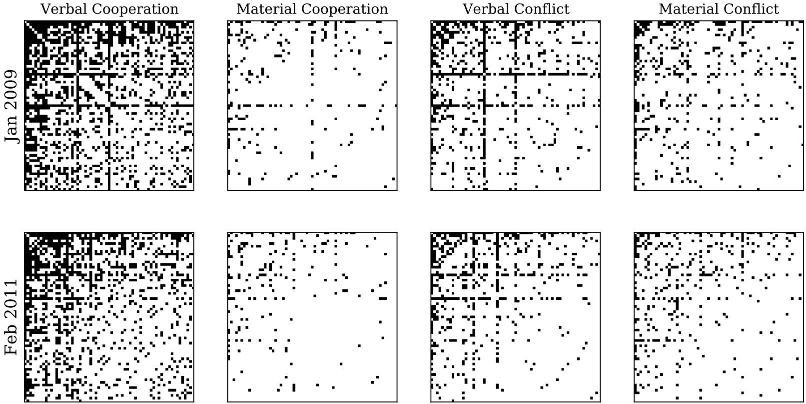
$$Y_{ij} \overset{\text{ind.}}{\sim} \mathrm{Bernoulli}(g(\mathbf{X}_i, \mathbf{X}_j)).$$



International Relations

$Y_{ijt}^k = 1$: country i and country j had a $\{$verbal cooperation, material cooperation, verbal conflict, material conflict$\}$ on the tth month.

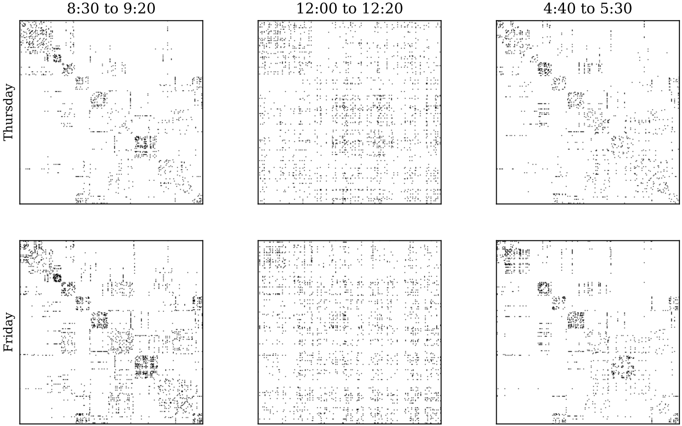
Note: Popular (high degree) nodes vary by time and layer.



School Contact Networks

$Y_{ijt}^{k} = 1$: student i and student j were in contact on $\{$Thursday, Friday$\}$ during the tth time period.

Note: Layers contain a highly-correlated time-varying structure.



Our Contribution

An Eigenmodel for Dynamic Multilayer Networks

$$Y_{ijt}^k \overset{\text{ind.}}{\sim} \mathrm{Bernoulli}(P_{ijt}^k)$$

$$\begin{aligned}\mathrm{logit}(P_{ijt}^k) &= \delta_{k,t}^i + \delta_{k,t}^j + \mathbf{X}_t^{i\mathrm{T}} \mathbf{\Lambda}_k \mathbf{X}_t^j \\ &= \delta_{k,t}^i + \delta_{k,t}^j + \sum_{h=1}^d \lambda_{kh} X_{th}^i X_{th}^j\end{aligned}$$

- Nodes are assigned a scalar-valued sociality, varies by layer and time:

$$\boldsymbol{\delta}_{k,t} = (\delta_{k,t}^1, \ldots, \delta_{k,t}^n) \in \mathbb{R}^n.$$

- Nodes are assigned latent vectors in $\mathbb{R}^d$, shared by layers but time-varying

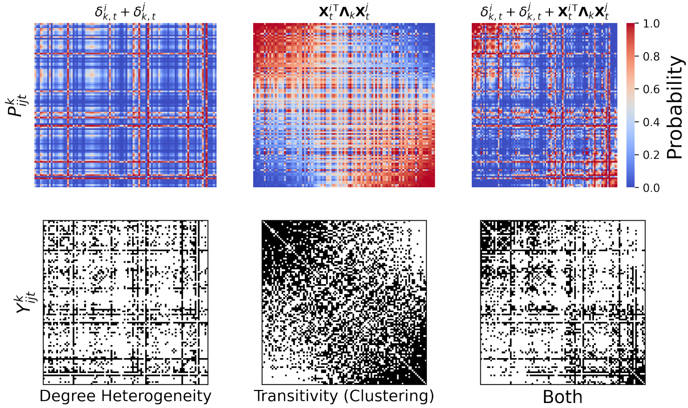
$$\mathcal{X}_t = (\mathbf{X}_t^1, \ldots, \mathbf{X}_t^n)^{\mathrm{T}} \in \mathbb{R}^{n \times d}.$$

- A diagonal homophily matrix, varies by layers.

$$\mathbf{\Lambda}_k = \mathrm{diag}(\boldsymbol{\lambda}_k) \in \mathbb{R}^{d \times d}.$$

Components of the Decomposition

Generate $\delta^i_{k,t} \overset{\text{iid}}{\sim} t_3 - 1$, $\mathbf{X}^i_t \overset{\text{iid}}{\sim} N(0, 2\, I_2)$, and $\boldsymbol{\Lambda}_k = I_2$:



The Role of the Homophily Matrix (Λ_k)

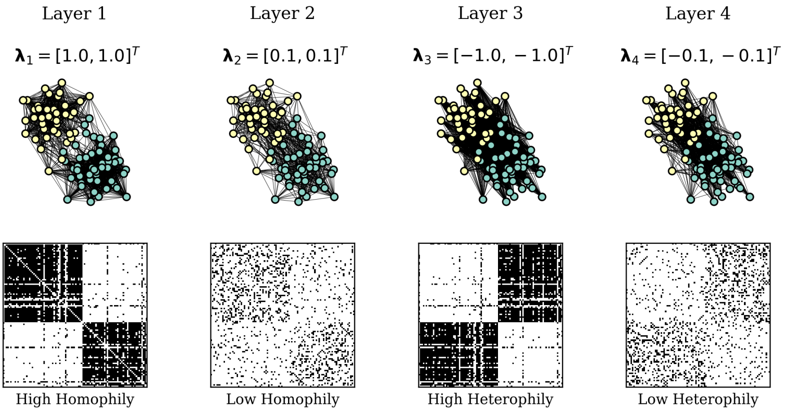
Model: $\mathrm{logit}(P_{ijt}^k) = \mathbf{X}_t^{i\mathrm{T}} \,\mathrm{diag}(\boldsymbol{\lambda}_k)\, \mathbf{X}_t^j = \sum_{h=1}^{d} \lambda_{kh} X_{th}^i X_{th}^j$.

- **Homophily:** Nodes with similar features form edges.
- **Heterophily:** Nodes with different features form edges.

The Role of the Homophily Matrix (Λ_k)

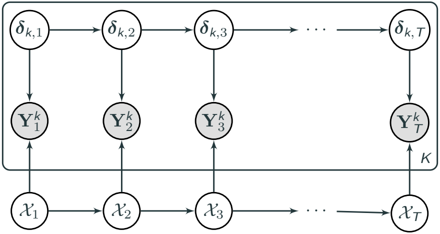
Model: $\mathrm{logit}(P^k_{ijt}) = \mathbf{X}_t^{i\mathrm{T}}\,\mathrm{diag}(\boldsymbol{\lambda}_k)\,\mathbf{X}_t^j = \sum_{h=1}^{d} \lambda_{kh} X^i_{th} X^j_{th}$.

- **Homophily:** Nodes with similar features form edges.
- **Heterophily:** Nodes with different features form edges.



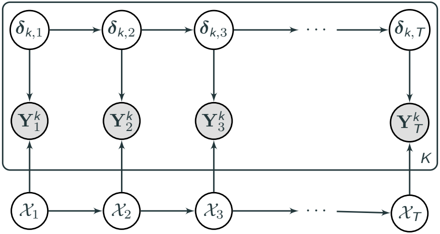
A State-Space Model for Dynamic Multilayer Networks

A network-valued state-space model (Sarkar and Moore, 2005; Sewell and Chen, 2015):



A State-Space Model for Dynamic Multilayer Networks

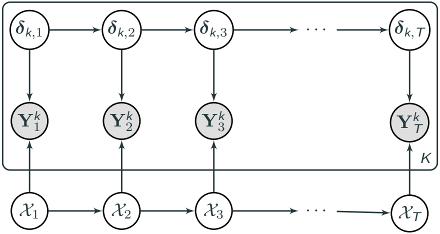
A network-valued state-space model (Sarkar and Moore, 2005; Sewell and Chen, 2015):



Gaussian Random Walk Priors

A State-Space Model for Dynamic Multilayer Networks

A network-valued state-space model (Sarkar and Moore, 2005; Sewell and Chen, 2015):



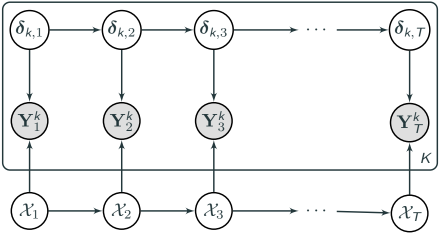
Gaussian Random Walk Priors

Social Trajectory: $\delta^i_{k,1:T} = (\delta^i_{k,1}, \ldots, \delta^i_{k,t})$

$$\delta^i_{k,1} \sim N(0, \tau^2_{\delta}), \qquad \delta^i_{k,t} \sim N(\delta^i_{k,t-1}, \sigma^2_{\delta}), \quad 2 \leq t \leq T.$$

A State-Space Model for Dynamic Multilayer Networks

A network-valued state-space model (Sarkar and Moore, 2005; Sewell and Chen, 2015):



Gaussian Random Walk Priors

Social Trajectory: $\delta^i_{k,1:T} = (\delta^i_{k,1}, \ldots, \delta^i_{k,t})$

$$\delta^i_{k,1} \sim N(0, \tau^2_\delta), \qquad \delta^i_{k,t} \sim N(\delta^i_{k,t-1}, \sigma^2_\delta), \quad 2 \leq t \leq T.$$

Latent Trajectory: $\mathbf{X}^i_{1:T} = (\mathbf{X}^i_1, \ldots, \mathbf{X}^i_T)$

$$\mathbf{X}^i_1 \sim N(0, \tau^2 I_d), \qquad \mathbf{X}^i_t \sim N(\mathbf{X}^i_{t-1}, \sigma^2 I_d), \quad 2 \leq t \leq T.$$

Remaining Priors

The remaining priors are

- $\boldsymbol{\lambda}_k \stackrel{\text{iid}}{\sim} N(0, \sigma_\lambda^2 I_d)$ for $1 \leq k \leq K$,
- $\tau_\delta^2 \sim \Gamma^{-1}(a_{\tau_\delta^2}, b_{\tau_\delta^2})$,
- $\sigma_\delta^2 \sim \Gamma^{-1}(c_{\sigma_\delta^2}, d_{\sigma_\delta^2})$,
- $\tau^2 \sim \Gamma^{-1}(a_{\tau^2}, b_{\tau^2})$,
- $\sigma^2 \sim \Gamma^{-1}(c_{\sigma^2}, d_{\sigma^2})$.

Hyperparameters where to chosen to be uninformative.

Parameter Identifiability

$$Y_{ijt}^{k} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\mathsf{logit}^{-1}\Big[\underbrace{\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\boldsymbol{\Lambda}_k\mathcal{X}_t^{\mathrm{T}}}_{\text{log-odds matrix}}\Big]_{ij}).$$

Sufficient Conditions for Latent Space Identifiability

Parameter Identifiability

$$Y_{ijt}^k \overset{\text{ind.}}{\sim} \text{Bernoulli}(\text{logit}^{-1}\Big[\underbrace{\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\boldsymbol{\Lambda}_k\mathcal{X}_t^{\mathrm{T}}}_{\text{log-odds matrix}}\Big]_{ij}).$$

Sufficient Conditions for Latent Space Identifiability

A1. *Centering:* $J_n\mathcal{X}_t = \mathcal{X}_t$, where $J_n = I_n - (1/n)\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$, for $1 \leq t \leq T$.

Parameter Identifiability

$$Y^k_{ijt} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\text{logit}^{-1}\Big[\underbrace{\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\boldsymbol{\Lambda}_k\mathcal{X}_t^{\mathrm{T}}}_{\text{log-odds matrix}}\Big]_{ij}).$$

Sufficient Conditions for Latent Space Identifiability

A1. *Centering:* $J_n\mathcal{X}_t = \mathcal{X}_t$, where $J_n = I_n - (1/n)\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$, for $1 \leq t \leq T$.

A2. *Full Rank:* $\text{rank}(\mathcal{X}_t) = d$ for $1 \leq t \leq T$.

Parameter Identifiability

$$Y_{ijt}^{k} \overset{\text{ind.}}{\sim} \mathrm{Bernoulli}(\mathrm{logit}^{-1}\Big[\underbrace{\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\boldsymbol{\Lambda}_k\mathcal{X}_t^{\mathrm{T}}}_{\text{log-odds matrix}}\Big]_{ij}).$$

Sufficient Conditions for Latent Space Identifiability

A1. *Centering:* $J_n\mathcal{X}_t = \mathcal{X}_t$, where $J_n = I_n - (1/n)\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$, for $1 \leq t \leq T$.

A2. *Full Rank:* $\mathrm{rank}(\mathcal{X}_t) = d$ for $1 \leq t \leq T$.

A3. *Reference Layer:* $\boldsymbol{\Lambda}_r = I_{p,q} = \mathrm{diag}(1, \dots, 1, -1, \dots, -1)$ for at least one $1 \leq r \leq K$.

Parameter Identifiability

$$Y_{ijt}^{k} \overset{\text{ind.}}{\sim} \mathrm{Bernoulli}(\mathsf{logit}^{-1}\Big[\underbrace{\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\boldsymbol{\Lambda}_k\mathcal{X}_t^{\mathrm{T}}}_{\text{log-odds matrix}}\Big]_{ij}).$$

Sufficient Conditions for Latent Space Identifiability

A1. *Centering:* $J_n\mathcal{X}_t = \mathcal{X}_t$, where $J_n = I_n - (1/n)\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$, for $1 \leq t \leq T$.

A2. *Full Rank:* $\mathrm{rank}(\mathcal{X}_t) = d$ for $1 \leq t \leq T$.

A3. *Reference Layer:* $\boldsymbol{\Lambda}_r = I_{p,q} = \mathrm{diag}(1, \ldots, 1, -1, \ldots, -1)$ for at least one $1 \leq r \leq K$.

A4. *Distinct Layers:* For at least on layer $k \neq r$, $\mathrm{rank}(\boldsymbol{\Lambda}_k) = d$ and $\boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_r$ has distinct diagonal elements, e.g., $\boldsymbol{\Lambda}_k \neq \alpha I_{p,q}$ for $\alpha \in \mathbb{R}$.

Parameter Identifiability with a Pair of Distinct Layers

Proposition 1

Suppose two sets of parameters $\{\boldsymbol{\delta}_{1:K,1:T}, \mathcal{X}_{1:T}, \boldsymbol{\Lambda}_{1:K}\}$ and $\{\tilde{\boldsymbol{\delta}}_{1:K,1:T}, \tilde{\mathcal{X}}_{1:T}, \tilde{\boldsymbol{\Lambda}}_{1:K}\}$ satisfy conditions **A1** – **A4** with $\boldsymbol{\Lambda}_r = I_{p,q}$ and $\tilde{\boldsymbol{\Lambda}}_r = I_{p',q'}$ and their log-odds matrices are equal, then the parameters are equal up to a signed permutation of the latent space.

That is, for all $1 \leq k \leq K$ and $1 \leq t \leq T$, we have that

$$\tilde{\boldsymbol{\delta}}_{k,t} = \boldsymbol{\delta}_{k,t}, \quad \tilde{\mathcal{X}}_t = \mathcal{X}_t \mathbf{M}_t, \quad \tilde{\boldsymbol{\Lambda}}_k = \mathbf{M}_t^{\mathrm{T}} \boldsymbol{\Lambda}_k \mathbf{M}_t,$$

where $\mathbf{M}_t = P \operatorname{diag}(\mathbf{s})$, $\mathbf{s} \in \{\pm 1\}^d$, and P is a $d \times d$ permutation matrix.

Remark: Most latent space models are identifiable up to a rotation.

Enforcing Identifiability

- **A2** (full rank) and **A4** (distinct layers):

  Holds with probability 1 under our priors.

- **A3** (reference layer):

  Re-parameterize the reference layer as follows

$$\lambda_{rh} = 2u_h - 1, \quad u_h \overset{\text{iid}}{\sim} \text{Bernoulli}(\rho), \quad h = 1, \ldots, d.$$

- **A1** (centering):

  Not enforced during inference, we center estimates upon convergence.

Parameter Identifiability without a Pair of Distinct Layers

Proposition 2

Remove Assumption **A4** from Proposition 1 and assume $d \leq 3$, then the parameters are identifiable up to a indefinite orthogonal transformation.

That is, for all $1 \leq k \leq K$ and $1 \leq t \leq T$, we have that

$$\tilde{\boldsymbol{\delta}}_{k,t} = \boldsymbol{\delta}_{k,t}, \quad \tilde{\mathcal{X}}_t = \mathcal{X}_t \mathbf{M}_t, \quad \tilde{\boldsymbol{\Lambda}}_k = \mathbf{M}_t^{\mathrm{T}} \boldsymbol{\Lambda}_k \mathbf{M}_t,$$

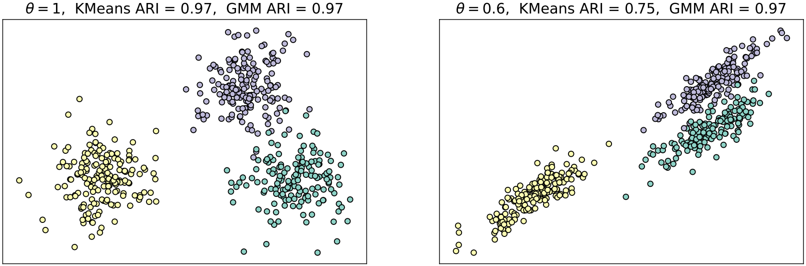
where $\mathbf{M}_t \in \mathbb{R}^{d \times d}$ satisfies $\mathbf{M}_t I_{p,q} \mathbf{M}_t^{\mathrm{T}} = I_{p,q}$.

Remark: When $d > 3$, $\mathbf{M}_t$ satisfies $\mathbf{M}_t I_{p',q'} \mathbf{M}_t^{\mathrm{T}} = I_{p,q}$.

The Indefinite Orthogonal Group and Community Detection

The set of matrices $\mathbf{M} I_{1,1} \mathbf{M}^{\mathrm{T}} = I_{1,1}$ contains hyperbolic rotations:

$$\mathbf{M} = \begin{pmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{pmatrix}.$$



Note: When using the latent space for community detection, do not assume spherical clusters!

Posterior Inference: Given observed networks $\{\mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T}\}$, we want to infer the latent parameters

$$\boldsymbol{\theta} = \{\boldsymbol{\delta}_{1:K,1:T}, \mathcal{X}_{1:T}, \boldsymbol{\Lambda}_{1:K}, \tau^2, \sigma^2, \tau^2_\delta, \sigma^2_\delta\}$$

based on the posterior:

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})}_{\substack{\textbf{Posterior:} \\ \text{coherent point estimation} \\ \text{\& uncertainty quantification}}} =$$

Bayesian Inference

Posterior Inference: Given observed networks $\{\mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T}\}$, we want to infer the latent parameters

$$\boldsymbol{\theta} = \{\boldsymbol{\delta}_{1:K,1:T}, \mathcal{X}_{1:T}, \boldsymbol{\Lambda}_{1:K}, \tau^2, \sigma^2, \tau^2_\delta, \sigma^2_\delta\}$$

based on the posterior:

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})}_{\substack{\textbf{Posterior:} \\ \text{coherent point estimation} \\ \text{\& uncertainty quantification}}} = \frac{1}{\underbrace{p(\mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})}_{\text{Normalizing Constant}}} \times \underbrace{\prod_{t=1}^{T} \prod_{k=1}^{K} \prod_{i<j} p(Y^k_{ijt} \mid \boldsymbol{\theta})}_{\text{Likelihood}} \times \underbrace{p(\boldsymbol{\theta})}_{\text{Prior}}.$$

Bayesian Inference

Posterior Inference: Given observed networks $\{\mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T}\}$, we want to infer the latent parameters

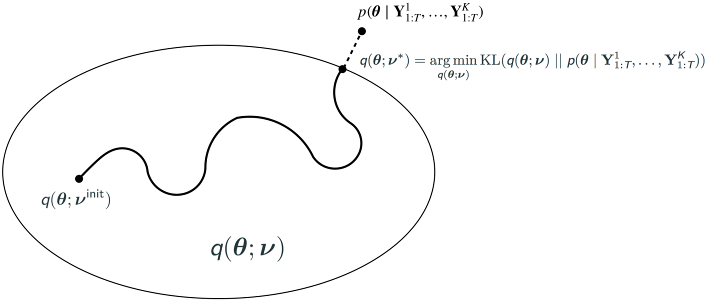
$$\boldsymbol{\theta} = \{\boldsymbol{\delta}_{1:K,1:T}, \mathcal{X}_{1:T}, \boldsymbol{\Lambda}_{1:K}, \tau^2, \sigma^2, \tau^2_\delta, \sigma^2_\delta\}$$

based on the posterior:

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})}_{\substack{\textbf{Posterior:} \\ \text{coherent point estimation} \\ \text{\& uncertainty quantification}}} = \frac{1}{\underbrace{p(\mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})}_{\text{Normalizing Constant}}} \times \underbrace{\prod_{t=1}^{T} \prod_{k=1}^{K} \prod_{i<j} p(Y^k_{ijt} \mid \boldsymbol{\theta})}_{\text{Likelihood}} \times \underbrace{p(\boldsymbol{\theta})}_{\text{Prior}}.$$

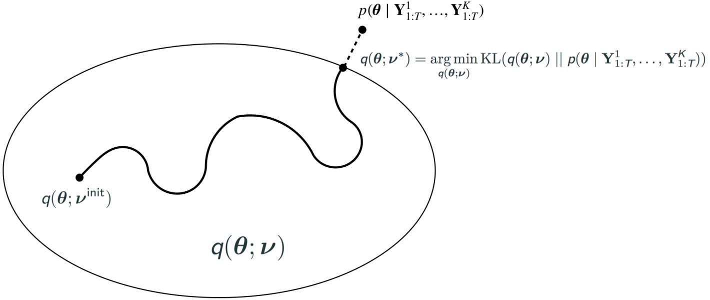
Challenge: The posterior is analytically intractable and expensive to sample from using MCMC even for small networks.

Solution: Approximate the intractable posterior $p(\boldsymbol{\theta} \mid \mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})$ with a parametric distribution $q(\boldsymbol{\theta}; \boldsymbol{\nu})$ with estimable parameters $\boldsymbol{\nu}$.



Variational Inference

Solution: Approximate the intractable posterior $p(\boldsymbol{\theta} \mid \mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})$ with a parametric distribution $q(\boldsymbol{\theta}; \boldsymbol{\nu})$ with estimable parameters $\boldsymbol{\nu}$.



New Challenge: Choose $q(\boldsymbol{\theta}; \boldsymbol{\nu})$ for an accurate approximation.

1. How should we factor $q(\boldsymbol{\theta}, \boldsymbol{\nu}) = \prod_{j=1}^{m} q(\boldsymbol{\theta}_j; \boldsymbol{\nu}_j)$?
2. What parametric form should each $q(\boldsymbol{\theta}_j; \boldsymbol{\nu}_j)$ take?

How Should We Factor $q(\theta; \nu)$?

General Rule: Maintain the posterior's strongest dependencies.

How Should We Factor $q(\theta; \nu)$?

General Rule: Maintain the posterior's strongest dependencies.

Example:

Generate $\mathbf{y}_i \overset{\text{iid}}{\sim} N_2(\boldsymbol{\mu}, \Sigma)$ for $i = 1, \ldots, N$ with a known, highly-correlated, Σ. Place a $N_2(\boldsymbol{\mu}_0, \Sigma_0)$ prior on $\boldsymbol{\mu}$. Goal is to approximate $p(\boldsymbol{\mu} \mid \mathbf{y}_{1:N}, \Sigma)$.

How Should We Factor $q(\boldsymbol{\theta}; \boldsymbol{\nu})$?

General Rule: Maintain the posterior's strongest dependencies.

Example:

Generate $\mathbf{y}_i \stackrel{\text{iid}}{\sim} N_2(\boldsymbol{\mu}, \Sigma)$ for $i = 1, \ldots, N$ with a known, highly-correlated, Σ. Place a $N_2(\boldsymbol{\mu}_0, \Sigma_0)$ prior on $\boldsymbol{\mu}$. Goal is to approximate $p(\boldsymbol{\mu} \mid \mathbf{y}_{1:N}, \Sigma)$.

$$q(\boldsymbol{\mu}; \boldsymbol{\nu}) = \underbrace{N_1(\tilde{\mu}_1, \tilde{\sigma}_1^2) \cdot N_1(\tilde{\mu}_2, \tilde{\sigma}_2^2)}_{\text{Mean Field}} \quad \text{or} \quad \underbrace{N_2(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})}_{\text{Full Rank}}$$

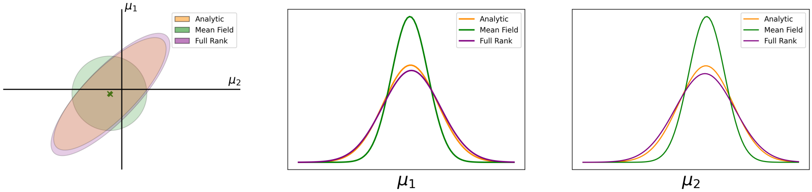
How Should We Factor $q(\boldsymbol{\theta}; \boldsymbol{\nu})$?

General Rule: Maintain the posterior's strongest dependencies.

Example:

Generate $\mathbf{y}_i \overset{\text{iid}}{\sim} N_2(\boldsymbol{\mu}, \Sigma)$ for $i = 1, \ldots, N$ with a known, highly-correlated, Σ. Place a $N_2(\boldsymbol{\mu}_0, \Sigma_0)$ prior on $\boldsymbol{\mu}$. Goal is to approximate $p(\boldsymbol{\mu} \mid \mathbf{y}_{1:N}, \Sigma)$.

$$q(\boldsymbol{\mu}; \boldsymbol{\nu}) = \underbrace{N_1(\tilde{\mu}_1, \tilde{\sigma}_1^2) \cdot N_1(\tilde{\mu}_2, \tilde{\sigma}_2^2)}_{\text{Mean Field}} \quad \text{or} \quad \underbrace{N_2(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})}_{\text{Full Rank}}$$



A Structured Approximation of the Eigenmodel's Posterior

Structured Variational Approximation

$$q(\boldsymbol{\theta}; \boldsymbol{\nu}) = \left[\prod_{h=1}^{d} q(\lambda_{1h})\right] \left[\prod_{k=2}^{K} q(\boldsymbol{\lambda}_k)\right] \left[\prod_{k=1}^{K} \prod_{i=1}^{n} q(\delta^i_{k,1:T})\right] \left[\prod_{i=1}^{n} q(\mathbf{X}^i_{1:T})\right] \\ \times q(\tau^2) q(\sigma^2) q(\tau^2_\delta) q(\sigma^2_\delta).$$

Unlike previous approaches (Liu and Chen, 2021), we

- Maintain the latent variable's strong temporal dependence.
- Use optimal distributions under this factorization.

Optimizing the Variational Objective

Optimal distributions are computed by iterating the following updates:

Coordinate Ascent Variational Inference (CAVI)

Cycle through $j \in \{1, \ldots, m\}$ until convergence:

$$\log q(\boldsymbol{\theta}_j; \boldsymbol{\nu}_j) = \mathbb{E}_{-j}\left[\log p(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{-j}, \mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T})\right] + c.$$

If the full conditionals are in the exponential family, then these updates are available in closed-form.

Problem: The Eigenmodel does not have this property!

Restoring Conditional Conjugacy through Data Augmentation

Pólya-gamma augmentation (Polson et al., 2013): For each dyad, we introduce Pólya-gamma latent variables $\omega_{ijt}^{k} \overset{\text{iid}}{\sim} \mathrm{PG}(0, 1)$.

Restoring Conditional Conjugacy through Data Augmentation

Pólya-gamma augmentation (Polson et al., 2013): For each dyad, we introduce Pólya-gamma latent variables $\omega_{ijt}^{k} \overset{\text{iid}}{\sim} \mathrm{PG}(0, 1)$.

The joint distribution is now

$$p(\mathbf{Y}_{1:T}^{1}, \ldots, \mathbf{Y}_{1:T}^{K}, \boldsymbol{\theta}, \boldsymbol{\omega}) = \underbrace{p(\mathbf{Y}_{1:T}^{1}, \ldots, \mathbf{Y}_{1:T}^{K} \mid \boldsymbol{\theta}, \boldsymbol{\omega})}_{\text{Augmented Likelihood}} \times \underbrace{p(\boldsymbol{\theta})p(\boldsymbol{\omega})}_{\text{Priors}}$$

Note: Marginalizing over $\boldsymbol{\omega}$, we recover the original joint distribution.

Restoring Conditional Conjugacy through Data Augmentation

Pólya-gamma augmentation (Polson et al., 2013): For each dyad, we introduce Pólya-gamma latent variables $\omega^k_{ijt} \overset{\text{iid}}{\sim} \mathrm{PG}(0,1)$.

The joint distribution is now

$$p(\mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T}, \boldsymbol{\theta}, \boldsymbol{\omega}) = \underbrace{p(\mathbf{Y}^1_{1:T}, \ldots, \mathbf{Y}^K_{1:T} \mid \boldsymbol{\theta}, \boldsymbol{\omega})}_{\text{Augmented Likelihood}} \times \underbrace{p(\boldsymbol{\theta})p(\boldsymbol{\omega})}_{\text{Priors}}$$

$$\propto \underbrace{\prod_{k=1}^{K} \prod_{t=1}^{T} \prod_{i<j} \exp\{z^k_{ijt}\,\psi^k_{ijt} - \omega^k_{ijt}\,(\psi^k_{ijt})^2/2\}}_{\text{Quadratic Likelihood}} \times \underbrace{p(\boldsymbol{\theta})p(\boldsymbol{\omega})}_{\text{Priors}},$$

where $z^k_{ijt} = Y^k_{ijt} - 1/2$ and $\psi^k_{ijt} = \delta^i_{k,t} + \delta^j_{k,t} + \mathbf{X}^{i\mathrm{T}}_t \boldsymbol{\Lambda}_k \mathbf{X}^j_t$.

Note: Marginalizing over $\boldsymbol{\omega}$, we recover the original joint distribution.

Variational Kalman Smoothing

The optimal latent (social) trajectories' variational distributions are Gaussian state-space models, e.g.,

$$\log q(\mathbf{X}^i_{1:T}) = \log h(\mathbf{X}^i_1) + \sum_{t=2}^{T} \log h(\mathbf{X}^i_t \mid \mathbf{X}^i_{t-1}) + \sum_{t=1}^{T} \log h(\mathbf{z}^i_t \mid \mathbf{X}^i_t) + c,$$

where

$$\log h(\mathbf{X}^i_1) = \mathbb{E}_{q(\tau^2)}\left[\log N(\mathbf{X}^i_1 \mid 0, \tau^2)\right],$$

$$\log h(\mathbf{X}^i_t \mid \mathbf{X}^i_{t-1}) = \mathbb{E}_{q(\sigma^2)}\left[\log N(\mathbf{X}^i_t \mid \mathbf{X}^i_{t-1}, \sigma^2)\right],$$

$$\log h(\mathbf{z}^i_t \mid \mathbf{X}^i_t) = \mathbb{E}_{-q(\mathbf{X}^i_{1:T})}\left[\sum_{k=1}^{K} \sum_{j \neq i} \log N(z^k_{ijt} \mid \omega^k_{ijt}(\delta^i_{k,t} + \delta^j_{k,t} + \mathbf{X}^{j\mathrm{T}}_t \mathbf{\Lambda}_k \mathbf{X}^i_t),\ \omega^k_{ijt})\right].$$

We then derive a novel Kalman smoothing type algorithm that calculates the moments of this variational distribution in closed-form.

CAVI for the Eigenmodel for Dynamic Multilayer Networks

Iterate the following steps until convergence:

1. Update each $q(\omega_{ijt}^{k}) = \mathrm{PG}(1, c_{ijt}^{k})$.
2. Update

   $q(\delta_{k,1:T}^{i})$: a Gaussian state space model for $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, K\}$,
   $q(\tau_{\delta}^{2}) = \Gamma^{-1}(\bar{a}_{\tau_{\delta}^{2}}/2, \bar{b}_{\tau_{\delta}^{2}}/2)$,
   $q(\sigma_{\delta}^{2}) = \Gamma^{-1}(\bar{c}_{\sigma_{\delta}^{2}}/2, \bar{d}_{\sigma_{\delta}^{2}}/2)$,

   using a variational Kalman smoother.
3. Update

   $q(\mathbf{X}_{1:T}^{i})$: a Gaussian state space model for $i \in \{1, \ldots, n\}$,
   $q(\tau^{2}) = \Gamma^{-1}(\bar{a}_{\tau^{2}}/2, \bar{b}_{\tau^{2}}/2)$,
   $q(\sigma^{2}) = \Gamma^{-1}(\bar{c}_{\sigma^{2}}/2, \bar{d}_{\sigma^{2}}/2)$,

   using a variational Kalman smoother.
4. Update $q(\lambda_{1h}) = p_{\lambda_{1h}}^{\mathbb{1}\{\lambda_{1h}=1\}}\,(1 - p_{\lambda_{1h}})^{\mathbb{1}\{\lambda_{1h}=-1\}}$ for $h \in \{1, \ldots, d\}$.
5. Update $q(\boldsymbol{\lambda}_{k}) = N(\boldsymbol{\mu}_{\boldsymbol{\lambda}_{k}}, \Sigma_{\boldsymbol{\lambda}_{k}})$ for $k \in \{2, \ldots, K\}$.

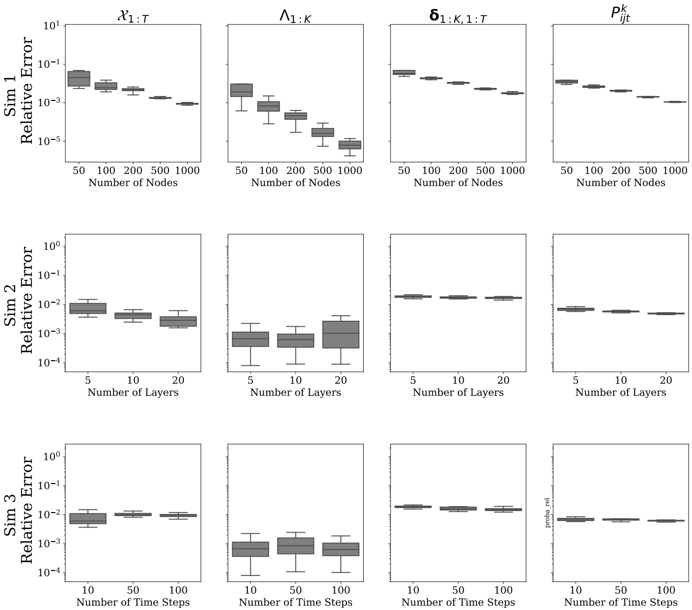
Simulation Study

We conducted simulations to see how estimation scaled with network size.

- **Simulation 1**: An increase in nodes $(n, K, T) \in \{50, 100, 200, 500, 1000\} \times \{5\} \times \{10\}$.
- **Simulation 2**: An increase in layers $(n, K, T) \in \{100\} \times \{5, 10, 20\} \times \{10\}$.
- **Simulation 3**: An increase in time points $(n, K, T) \in \{100\} \times \{5\} \times \{10, 50, 100\}$.

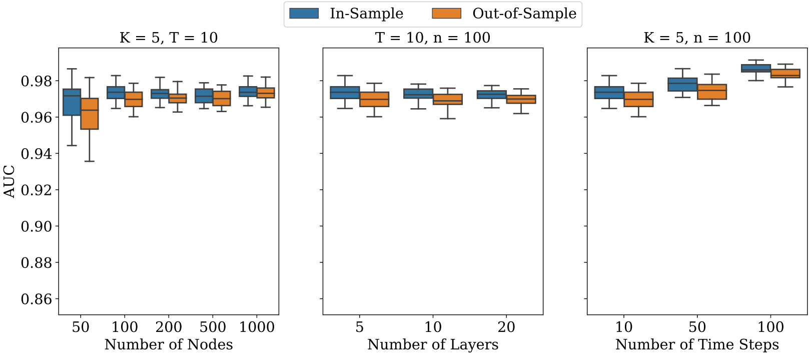
Estimation Error: The relative Frobenius norm $\|A - \hat{A}\|_F^2 / \|A\|_F^2$.

Estimation Error



Predictive Performance

AUC for in-sample and out-of-sample dyads based on our model's predictions. We removed 20% of the dyads randomly from each layer and time-step.

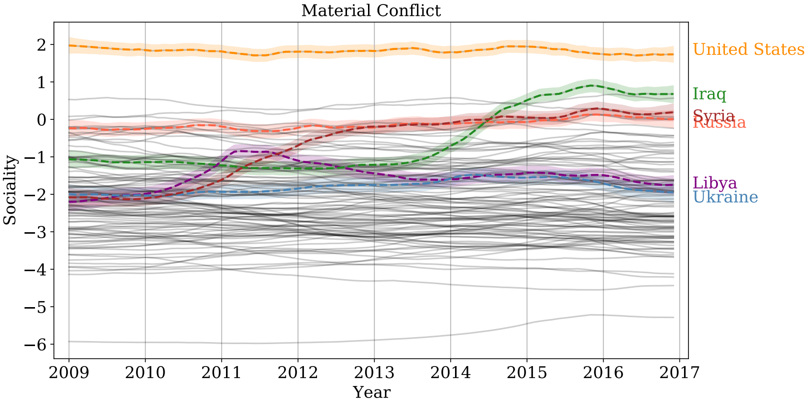


International Relation Networks, 2009–2017

Eight years during the Obama administration of monthly relational data taken from ICEWS (Boschee et al., 2015).

- $Y^k_{ijt} = 1$: country i and country j had a $\{$verbal cooperation, material cooperation, verbal conflict, material conflict$\}$ on the tth month.
- The verbal conflict relation was taken as the reference layer because it had the highest density.
- $n = 100$ countries, $T = 96$ months, $K = 4$ relations.
- Estimated a model with $d = 2$ for visualization.

Social Trajectories Reveal Global Events

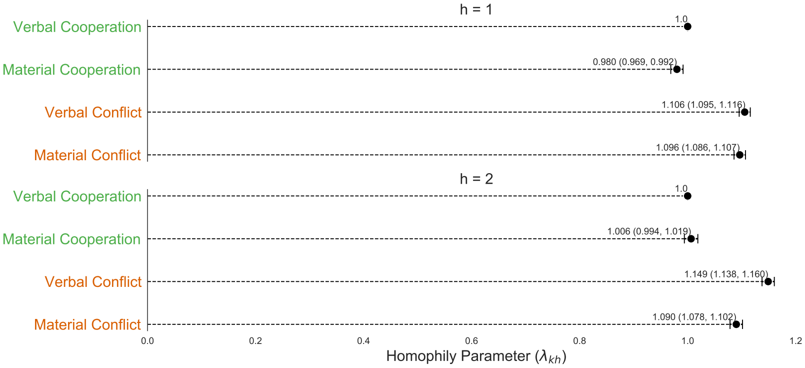


Reveals global conflicts: Arab Spring in Libya (2011), Rise of ISIL in Syria (2013 – present), The American-led intervention in Iraq (2014).

Does not indicate the Crimean Crisis between Russia and Ukraine (2014).

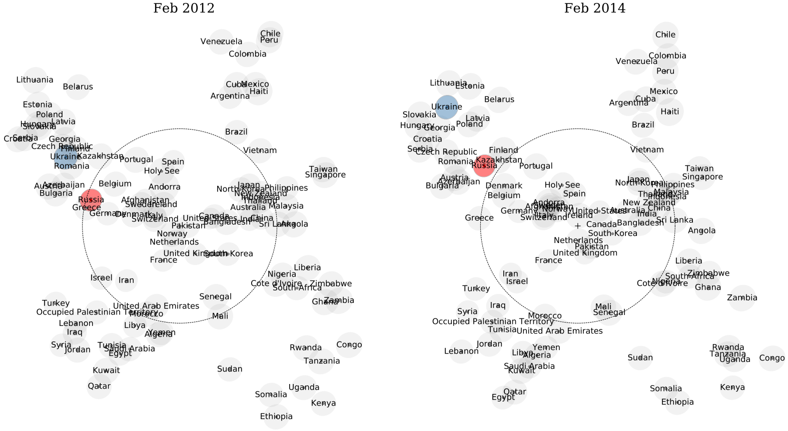
Homophily Varies between Cooperation and Conflict

Conflict layers utilize the latent space more when predicting a link:



Latent Positions Correlate with Geographic Location

The latent space reflects a nation's geographic location:



Latent Space Dynamics Reveal Regional Events

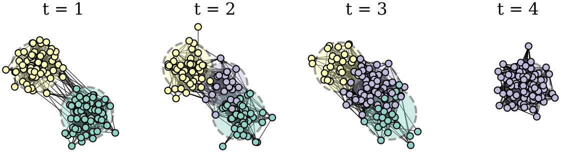
Dynamics reflect the 2014 Crimean Crisis between Ukraine and Russia:

<latent space movie>

Conclusion

- The Eigenmodel for Dynamic Multilayer Networks is a tractable model for multiple time-varying network data.
- Unlike previous methods, it's parameters are interpretable and identifiable.
- A novel variational inference algorithm provides meaningful uncertainty quantification and scales to large networks.
- Applications in international relations, epidemiology, and other fields.

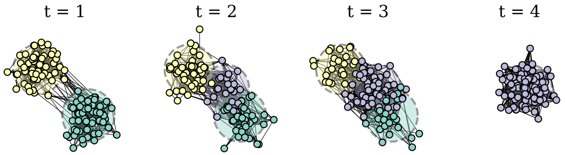
1. **Time-varying community structure in dynamic networks.**
   - A new Bayesian nonparametric prior that infers discrete changes in community structure in dynamic networks.



(*Bayesian Analysis*, in press.)

Other Projects

1. **Time-varying community structure in dynamic networks.**
   - A new Bayesian nonparametric prior that infers discrete changes in community structure in dynamic networks.



(*Bayesian Analysis*, in press.)

2. **Local variable importance for random forests.**
   - Adapt the random forest kernel to local structures using a new linear combination splitting rule.

(*Journal of Computational and Graphical Statistics*, in revision.)

3. **Statistical network analysis for the COVID-19 pandemic.**
   - Combined statistical network models with network compartmental models to study disease progression.

   (*International Statistical Review*, 2020)

Other Projects

3. **Statistical network analysis for the COVID-19 pandemic.**
   - Combined statistical network models with network compartmental models to study disease progression.

   (*International Statistical Review*, 2020)

4. **Bayesian modeling averaging for dynamic LSMs.**
   - Infer $p(d \mid Y_{1:T})$ using a state-space model on the Steifel manifold with sparsity inducing priors.

   (in preparation.)

Future Directions

Statistical models for complex and higher-order networks.

- Testing for layerwise correlation in multilayer networks using network random-effects.
- Community detection in dynamic multilayer networks.

Scalable Bayesian inference for network data.

- Bayesian coresets, consensus Monte Carlo, stochastic gradient MCMC, stochastic variational inference, variational graph auto-encoders.

Future Directions

Quantifying homophily in networks.

- Networks often contain node-level and edge-level covariates.
- Combine machine learning models (e.g., random forests) to better understand how these covariates contribute to network formation.

Regression with network-valued covariates

- Regression of a univariate response on network-valued covariates:

$$y_i = g(\underbrace{A_i}_{\text{network}}) + \epsilon_i, \quad i = 1, \ldots n.$$

- Use sufficient dimension reduction to infer the network's contribution.

Applied scientific problems

- At UIUC, I collaborated heavily with scientists from Sandia National Laboratories and the Environmental Science department.
- Leveraging my expertise in Bayesian inference and computational statistics, I hope to develop new interdisciplinary collaborations.

Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J., and Ward, M. (2015). ICEWS Coded Event Data.

Durante, D., Mukherjee, N., and Steorts, R. C. (2017). Bayesian learning of dynamic multilayer networks. *Journal of Machine Learning Research*, 18(43):1–29.

Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *Annals of Applied Statistics*, 9(3):1169–1193.

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(460):1090–1098.

Liu, Y. and Chen, Y. (2021). Variational inference for latent space models for dynamic networks. in press.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference of logistic models using Pólya-gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–13349.

Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models. *SIGKDD Explorations*, 7(2):31–40.

Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.

Snijders, T. A., Lomi, A., and Torió, V. J. (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks*, 35(2):265–276.

Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaggiotto, M., den Broeck, W. V., Régis, C., Lina, B., and Vanhems, P. (2011). High-resolution measurements of face-to-face contact patterns in primary school. *PLoS ONE*, 6(8):e23176.

Extensions and Future Directions

Simple Extensions:

- Incorporation of dyad-wise covariates $\{\mathbf{Z}^k_{ijt}\}$ through a linear term

$$\text{logit}(P^k_{ijt}) = \beta^{\mathrm{T}}\mathbf{Z}^k_{ijt} + \delta^i_{t,k} + \delta^j_{t,k} + \mathbf{X}^{i\mathrm{T}}_t \Lambda_k \mathbf{X}^j_t.$$

- Extension to directed networks

$$\text{logit}(P^k_{ijt}) = \delta^i_{t,k} + \gamma^j_{t,k} + \mathbf{U}^{i\mathrm{T}}_t \Lambda_k \mathbf{V}^j_t.$$

Future Research:

- Allow inference on larger networks through stochastic optimization.
- Explore non-stationary state-space models.
- Extend the VB algorithm beyond binary or real-valued dyads, e.g., count data.

Correcting for Centering Identifiability

To satisfy **A1** (centering), note that the likelihood is invariant to translations:

$$\begin{aligned}\delta^i_{k,t} + \delta^j_{k,t} + \mathbf{X}^{i\mathrm{T}}_t \Lambda_k \mathbf{X}^j_t &= \delta^i_{k,t} + \delta^j_{k,t} + (\mathbf{X}^i_t - \mathbf{c} + \mathbf{c})^{\mathrm{T}} \Lambda_k (\mathbf{X}^j_t - \mathbf{c} + \mathbf{c}),\\ &= \tilde{\delta}^i_{k,t} + \tilde{\delta}^j_{k,t} + \tilde{\mathbf{X}}^{i\mathrm{T}}_t \Lambda_k \tilde{\mathbf{X}}^j_t,\end{aligned}$$

where $\tilde{\mathbf{X}}^i_t = \mathbf{X}^i_t - \mathbf{c}$ and $\tilde{\delta}^i_{k,t} = \delta^i_{k,t} + \tilde{\mathbf{X}}^{i\mathrm{T}}_t \Lambda_k \mathbf{c} + \mathbf{c}^{\mathrm{T}} \Lambda_k \mathbf{c}/2$.

Therefore, given an estimate of the approximite posterior, we can estimate the centered solutions $\tilde{\mathbf{X}}^i_t$ and $\tilde{\delta}^i_t$ with $\mathbf{c} = (1/n) \sum_{i=1}^n \mathbf{X}^i_t$.

School Contact Networks

Two days of contact data collected at a primary school in France (Stehlé et al., 2011).

- $Y^k_{ijt} = 1$: individual i and individual j had a at least one interaction (≤ 5 ft) lasting more than 20 seconds during the tth 20 minute interval on $\{$Thursday, Friday$\}$.
- Thursday is taken as the reference layer.
- $n = 242$ individuals, $T = 24$ time steps, $K = 2$ days.

Estimating the Epidemiological Branching Factor

Often a disease's epidemic propensity is summarized by the *epidemic branching factor*

$$\kappa = \frac{\sum_{i=1}^{n} d_i^2/n}{\sum_{i=1}^{n} d_i/n}.$$

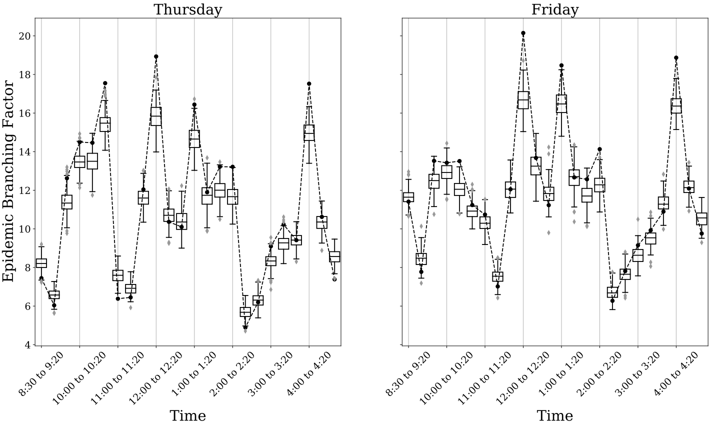
In network-based SIR and SEIR models, the *basic reproduction number*

$$R_0 = \frac{\tau}{\tau + \gamma}\ (\kappa - 1),$$

where τ and γ are infection and recovery rates, respectively.

Branching Factor Estimates

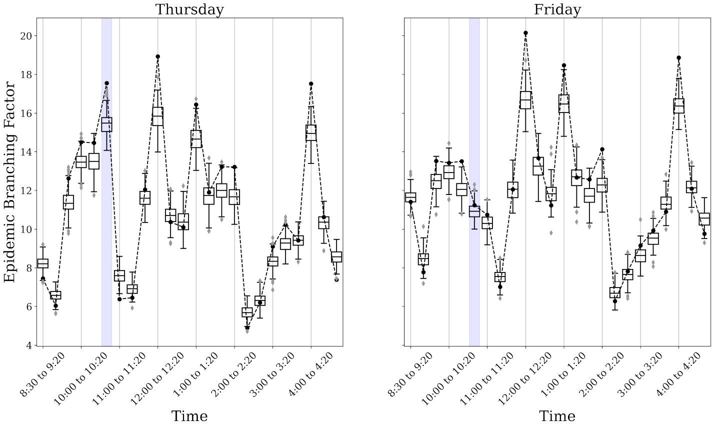
Estimated with 250 samples from the approximate posterior.



Captures spikes during the two breaks (around 10:30 am and 3:30 pm) and lunch (12:00 - 1:00 pm).

Branching Factor Estimates

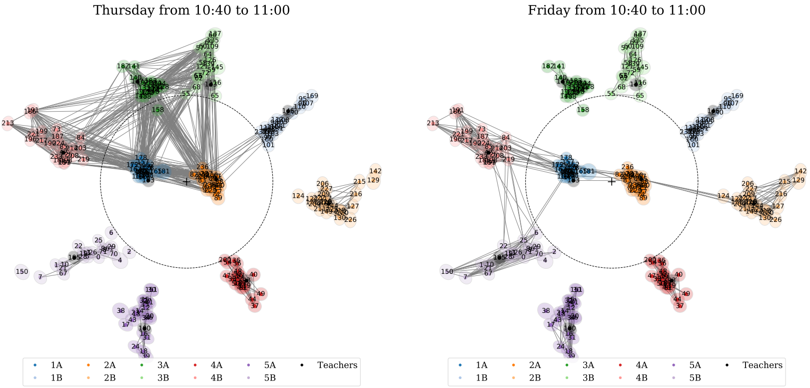
Estimated with 250 samples from the approximate posterior.



Captures spikes during the two breaks (around 10:30 am and 3:30 pm) and lunch (12:00 - 1:00 pm).

Heterogeneous Contact Patterns

The latent space reveals heterogeneous connectivity patterns between classrooms.



The Role of the Homophily Matrix (Λ_k)

Toy Model: For $i = 1, \ldots, n$, assign a binary latent feature $X_t^i \in \{-1, 1\}$.

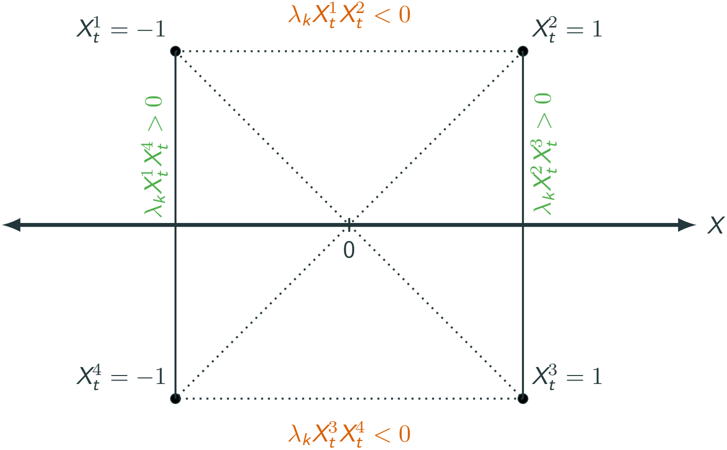
Deterministically form an edge as follows:

$$Y_{ijt}^k = \mathbb{1}\{\lambda_k X_t^i X_t^j > 0\}.$$

The Role of the Homophily Matrix (Λ_k)

Toy Model: $Y^k_{ijt} = \mathbb{1}\{\lambda_k X^i_t X^j_t > 0\}$.

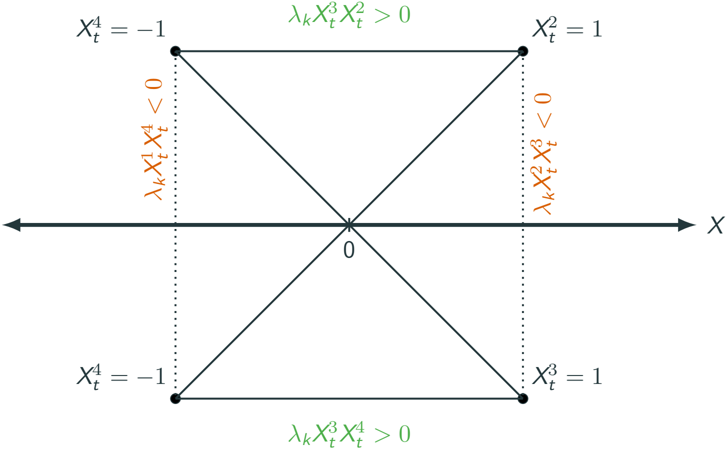
When $\lambda_k > 0$, the relationship is **homophilic** (same features connect):



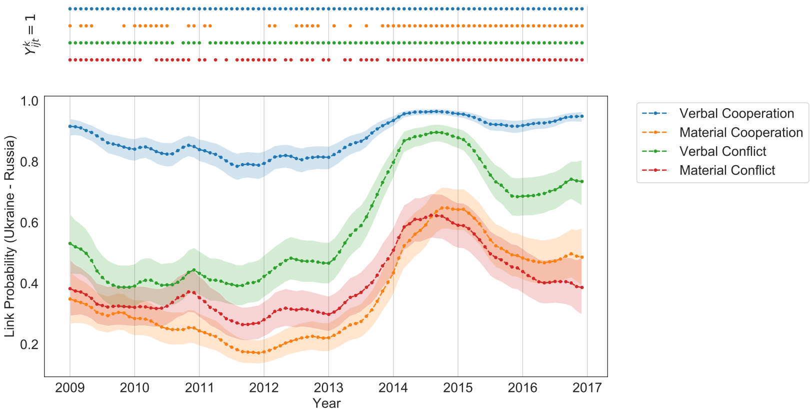
The Role of the Homophily Matrix (Λ_k)

Toy Model: $Y_{ijt}^k = \mathbb{1}\{\lambda_k X_t^i X_t^j > 0\}$.

When $\lambda_k < 0$, the relationship is **heterophilic** (opposites connect):



2014 Crimean Crisis



The estimated link probability between Russia and Ukraine increases dramatically around the Crimean Crisis.