Journal of Machine Learning Research 24 (2023) 1-69 Submitted 3/21; Published 4/23

An Eigenmodel for Dynamic Multilayer Networks

Joshua Daniel Loyal JLOYAL@FSU.EDU
Department of Statistics
Florida State University
Tallahassee, FL 32308, USA

Yuguo Chen YUGUO@ILLINOIS.EDU
Department of Statistics
University of Illinois at Urbana-Champaign
Champaign, IL 61820, USA

Editor: Ji Zhu

Abstract

Dynamic multilayer networks frequently represent the structure of multiple co-evolving relations; however, statistical models are not well-developed for this prevalent network type. Here, we propose a new latent space model for dynamic multilayer networks. The key feature of our model is its ability to identify common time-varying structures shared by all layers while also accounting for layer-wise variation and degree heterogeneity. We establish the identifiability of the model's parameters and develop a structured mean-field variational inference approach to estimate the model's posterior, which scales to networks previously intractable to dynamic latent space models. We demonstrate the estimation procedure's accuracy and scalability on simulated networks. We apply the model to two real-world problems: discerning regional conflicts in a data set of international relations and quantifying infectious disease spread throughout a school based on the student's daily contact patterns.

Keywords: dynamic multilayer network, epidemics on networks, latent space model, statistical network analysis, variational inference

1. Introduction

Dynamic multilayer networks are a prevalent form of relational data with applications in epidemiology, sociology, biology, and other fields (Boccaletti et al., 2014). Unlike static single-layer networks, which are limited to recording one dyadic relation among a set of actors at a single point in time, dynamic multilayer networks contain several types of dyadic relations, called layers, observed over a sequence of times. For instance, social networks contain several types of social relationships jointly evolving over time: friendship, vicinity, coworker-ship, partnership, and others. Also, international relations unfold through daily political events involving two countries, e.g., offering aid, verbally condemning, or participating in military conflict (Hoff, 2015). Lastly, the spread of information on social media occurs on a dynamic multilayer network, e.g., hourly interactions among Twitter users such as liking, replying to, and re-tweeting each other's content (Domenico et al., 2013). Proper

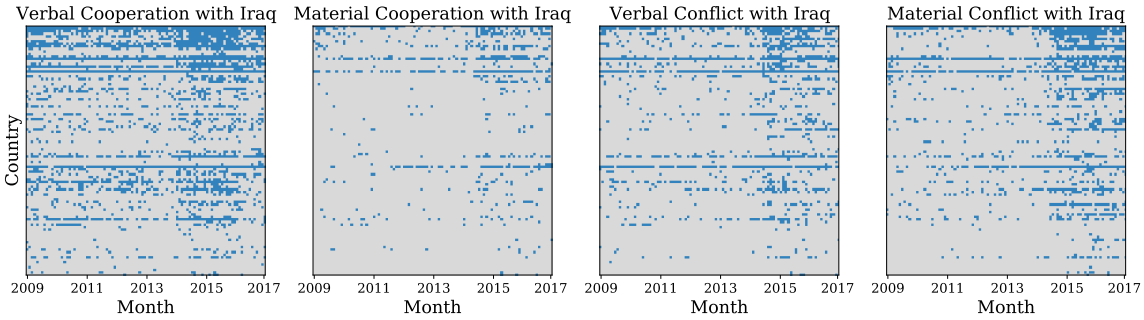


Figure 1: Monthly cooperation and conflict relations between Iraq and other countries from 2009 to 2017. A blue (gray) square indicates a relation occurred (did not occur) between Iraq and that nation during that month.

statistical modeling of dynamic multilayer networks is essential for an accurate understanding of these complex systems.

The statistical challenge in modeling multiple co-evolving networks is maintaining a concise representation while also adequately describing important network characteristics. These characteristics include the dyadic dependencies in each individual static relation, such as degree heterogeneity and transitivity, the autocorrelation of the individual dyadic time series, and the common structures shared among the various relations. We provide an example of such network characteristics in Figure 1, which displays the monthly time series of four dyadic relations between Iraq and other countries from 2009 to 2017. A complete description of the data can be found in Section 5. Within a layer (e.g., verbal cooperation), the individual time series (rows) are correlated with each other while also exhibiting strong autocorrelation. Furthermore, the four relations share a clear homogeneous structure, which is made especially evident after the abrupt change in all dyadic time series in late 2014 due to an American-led intervention in Iraq. We explore this event in more detail in Section 5. A statistical network model should decompose these dependencies in an interpretable way.

To date, the statistics literature contains an expansive collection of network models designed to capture specific network properties. See Goldenberg et al. (2010) and Loyal and Chen (2020) for a comprehensive review. An important class of network models is latent space models (LSMs) proposed in Hoff et al. (2002). The key idea behind LSMs is that each actor is assigned a vector in some low-dimensional latent space whose pairwise distances under a specified similarity measure determine the network's dyad-wise connection probabilities. The LSM interprets these latent features as an actor's unmeasured characteristics such that actors that are close in the latent space are more likely to form a connection. This interpretation naturally explains the high levels of homophily (assortativity) and transitivity in real-world networks. A series of works expanded the network characteristics captured by LSMs (Handcock et al., 2007; Hoff, 2008; Krivitsky et al., 2009; Hoff, 2005; Ma et al., 2020), such as community structure, degree heterogeneity, heterophily (disassortativity), etc. Furthermore, researchers have adopted the LSM formulation to model both dynamic networks (Sarkar and Moore, 2005; Durante and Dunson, 2014; Sewell and

Chen, 2015; He and Hoff, 2019) and static multilayer networks (Gollini and Murphy, 2016; Salter-Townshend and McCormick, 2017; D'Angelo et al., 2019; Wang et al., 2019; Zhang et al., 2020).

Currently, the statistical methodology for modeling dynamic multilayer networks is limited. Snijders et al. (2013) introduced a stochastic actor-oriented model which represents the networks as co-evolving continuous-time Markov processes. In addition, Hoff (2015) introduced a multilinear tensor regression framework where real-valued dynamic multilayer networks are modeled through tensor autoregression. To our knowledge, the only existing LSM for dynamic multilayer networks is the Bayesian nonparametric model proposed in Durante et al. (2017). Although highly flexible, this model lacks interpretability due to strong non-identifiable issues. Furthermore, this model's applications are limited to small networks with only a few dozen nodes and time points due to the model's high computational complexity. Currently, the LSM literature lacks models that decompose the complexity of dynamic multilayer networks into interpretable components and scale to the large networks commonly analyzed in practice.

To address these needs, we develop a new Bayesian dynamic bilinear latent space model that is flexible, interpretable, and computationally efficient. Our approach identifies a common time-varying structure shared by all layers while also accounting for layer-wise variation. Intuitively, our model posits that actors have intrinsic traits that influence how they connect in each layer. Specifically, we identify a common structure in which we represent each node by a single latent vector shared across layers. Also, we introduce node-specific additive random effects (or socialities) to adjust for heavy-tailed degree distributions (Rastelli et al., 2016). The model accounts for layer-wise heterogeneity in two ways. First, the layers assign different amounts of homophily (heterophily) to each latent trait. Second, to capture the dependence of an actor's degree on relation type, we allow the additive random effects to vary by layer. Lastly, we propagate the latent variables through time via a discrete Markov process (Sarkar and Moore, 2005; Sewell and Chen, 2015). These correlated changes capture the network's structural evolution and temporal autocorrelation.

To estimate our model, we derive a variational inference algorithm (Wainwright and Jordan, 2008; Blei et al., 2017) that scales to networks much larger than those analyzed by previous approaches. We base our inference on a structured mean-field approximation to the posterior. Our approximation improves upon previous variational approximations found in the dynamic latent space literature (Sewell and Chen, 2017; Liu and Chen, 2022) by retaining the latent variable's temporal dependencies. Furthermore, we derive a coordinate ascent variational inference algorithm that consists of closed-form updates. Our work leads to a novel approach to fitting dynamic latent space models using techniques from the linear Gaussian state space model (GSSM) literature.

The structure of our paper is as follows. In Section 2, we present our Bayesian parametric model for dynamic multilayer networks, discuss identifiability issues, and connect it to existing models for static multilayer and single-layer dynamic networks. Section 3 outlines our structured mean-field approximation and the coordinate ascent variational inference algorithm used for estimation. Section 4 demonstrates the accuracy of our inference algorithm and compares it to other estimation methods on simulated networks. In Section 5, we apply our model to two real-world networks taken from international relations and epidemiology. Finally, Section 6 concludes with a discussion of various model extensions and future

research directions. The Appendices contain proofs, in-depth derivations of the variational inference algorithm, implementation details, and additional results and figures.

Notation. We write $[N] = \{1, \ldots, N\}$. We use the notation $B_{1:L}$ to refer to the sequence $(B_1, B_2, \ldots, B_L)$ where B_ℓ is any indexed object. Also, for objects with a double index, we use the notation $C_{1:M,1:N}$ to refer to the collection $(C_{mn})_{(m,n)\in[M]\times[N]}$. We use $\mathbb{1}_{\{x=a\}}$ to denote the Boolean indicator function, which evaluates to 1 when $x = a$ and 0 otherwise. We denote an n-dimensional vector of ones by $\mathbf{1}_n$, an n-dimension vector of zeros by $\mathbf{0}_n$, and the $n \times n$ identity matrix by I_n. Furthermore, given a vector $\mathbf{v} \in \mathbb{R}^d$, we use $\text{diag}(\mathbf{v})$ to indicate a $d \times d$ diagonal matrix with the elements of $\mathbf{v}$ on the diagonal. We use $I_{p,q} = \text{diag}(1, \ldots, 1, -1, \ldots, -1)$ to denote a diagonal matrix with p ones followed by q negative ones on the diagonal. Lastly, we use $\text{BlockDiagonal}(A, B)$ to denote a block diagonal matrix with matrices A and B on the diagonal.

2. An Eigenmodel for Dynamic Multilayer Networks

In this section, we develop our Bayesian model for dynamic multilayer networks. To begin, we formally introduce dynamic multilayer network data. Dynamic multilayer networks consist of K relations measured over T time points between the same set of n nodes (or actors). We collect these relations in binary adjacency matrices $\mathbf{Y}_t^k \in \{0,1\}^{n\times n}$ for $1 \le k \le K$ and $1 \le t \le T$. The entries Y_{ijt}^k indicate the presence ($Y_{ijt}^k = 1$) or absence ($Y_{ijt}^k = 0$) of an edge between actors i and j in layer k at time t. This article only considers undirected networks without self-loops so that $\mathbf{Y}_t^k$ is a symmetric matrix. We discuss extensions of our model to weighted and directed networks in Section 6.

2.1 The Model

Here, we propose our new eigenmodel for dynamic multilayer networks with the goal of capturing the correlations between different dyads within a network, the dyads' autocorrelation over time, and the dependence between layers. Specifically, we assume that the dyads are independent Bernoulli random variables conditioned on the latent parameters:

$$\mathbb{P}\left(\mathbf{Y}_{1:T}^1, \ldots, \mathbf{Y}_{1:T}^K \mid \boldsymbol{\delta}_{1:K,1:T}, \Lambda_{1:K}, \mathcal{X}_{1:T}\right) = \prod_{k=1}^{K}\prod_{t=1}^{T}\prod_{j<i} \mathbb{P}\left(Y_{ijt}^k \mid \delta_{k,t}^i, \delta_{k,t}^j, \Lambda_k, \mathbf{X}_t^i, \mathbf{X}_t^j\right),$$

where

$$\text{logit}\left[\mathbb{P}\left(Y_{ijt}^k = 1 \mid \delta_{k,t}^i, \delta_{k,t}^j, \Lambda_k, \mathbf{X}_t^i, \mathbf{X}_t^j\right)\right] = \Theta_{ijt}^k = \delta_{k,t}^i + \delta_{k,t}^j + \mathbf{X}_t^{i\,\mathrm{T}} \Lambda_k \mathbf{X}_t^j. \tag{1}$$

In Equation (1), layer k's log-odds matrix at time t, $\boldsymbol{\Theta}_t^k \in \mathbb{R}^{n\times n}$ with elements Θ_{ijt}^k, contains two latent random effects that induce essential unconditional dependencies in the dynamic multilayer network's dyads. We defer specification of their distributions until the next section. First, the sociality effects $\boldsymbol{\delta}_{k,t} = (\delta_{k,t}^1, \ldots, \delta_{k,t}^n)^{\mathrm{T}} \in \mathbb{R}^n$ model degree heterogeneity and node-level autocorrelation. Second, the time-varying latent positions $\mathcal{X}_t = (\mathbf{X}_t^1, \ldots, \mathbf{X}_t^n)^{\mathrm{T}} \in \mathbb{R}^{n\times d}$, where d is the latent space's dimension, induce clusterability (Hoff, 2008) in the networks and dyadic autocorrelation. Finally, the homophily coefficients $\Lambda_k = \text{diag}(\boldsymbol{\lambda}_k) \in \mathbb{R}^{d\times d}$ are diagonal matrices that quantify each relation's level of

homophily along a latent dimension. For each λ_{kh} ($1 \leq h \leq d$), positive values ($\lambda_{kh} > 0$) indicate homophily along the hth latent dimension in layer k, negative values ($\lambda_{kh} < 0$) indicate heterophily along the hth latent dimension in layer k, and a zero value ($\lambda_{kh} = 0$) indicates the hth latent dimension does not contribute to the connection probability in layer k. Furthermore, the model captures common structures among the layers by sharing a common set of latent trajectories.

In matrix form, the log-odds matrices are

$$\boldsymbol{\Theta}_t^k = \boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\Lambda_k\mathcal{X}_t^{\mathrm{T}}. \tag{2}$$

To ensure identifiability of the model parameters, we require both a centered latent space, that is $J_n\mathcal{X}_t = \mathcal{X}_t$ where $J_n = I_n - (1/n)\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$, and $\Lambda_r = I_{p,q}$, where $p + q = d$, for some reference layer $r \in \{1, \ldots, K\}$. In an applied setting, one could take a particular interesting layer as the reference layer. Without loss of generality, we set $r = 1$ as the reference. As we elaborate in Section 2.5, we use these conditions to identify the socialities $\boldsymbol{\delta}_{k,t}$, and to identify $\mathcal{X}_t$ up to a signed-permutation of its columns, where the permutation is common to all time points but the sign-flips can vary over time. At the same time, we show that the bilinear term, $\mathcal{X}_t\Lambda_k\mathcal{X}_t^{\mathrm{T}}$, is directly identifiable.

Overall, our proposed eigenmodel for dynamic networks's parameterization reduces the dimensionality of dynamic multi-relational data. The model contains $nTK + nTd + Kd$ parameters, which, for typical values of d, is much less than the $KTn(n-1)/2$ dyads that originally summarized the dynamic multilayer network. Next, we further elaborate on the interpretation of the model's parameters and our inclusion of temporal correlation in the random effects.

2.2 Layer-Specific Social Trajectories

An actor's sociality, $\delta_{k,t}^i$, represents their global popularity in layer k at time t. In particular, holding all other parameters fixed, the larger an actor's sociality $\delta_{k,t}^i$, the more likely they are to connect with other nodes in the kth layer at time t regardless of their position in the latent space. Formally, $\delta_{k,t}^i$ is the conditional log-odds ratio of actor i forming a connection with another actor in layer k at time t compared to an actor with the same latent position as actor i but with $\delta_{k,t}^i = 0$. Hub nodes are an example of nodes with a high sociality, while isolated nodes have a low sociality. An actor's sociality can differ between layers. We find this flexibility necessary to model real-world multilayer relations. For example, in the international relations network presented in the introduction, a peaceful nation might participate in many cooperative relations while rarely engaging in conflict relations.

The ith actor's social trajectory in layer k, $\delta_{k,1:T}^i$, measures their time-varying sociality in the kth layer. For example, a nation's propensity to engage in militaristic relations might increase after a regime change. We assume that the social trajectories are independent across layers k and individuals i and propagate them through time via a shared Markov process:

$$\delta_{k,1}^i \overset{\text{iid}}{\sim} N(0, \tau_\delta^2), \qquad \delta_{k,t}^i \sim N(\delta_{k,t-1}^i, \sigma_\delta^2), \qquad t = 2, \ldots, T, \quad k = 1, \ldots, K,$$

where iid stands for independent and identically distributed. In the previous expression, τ_δ^2 measures the sociality effects' initial variation over all layers. Similarly, σ_δ^2 measures the

sociality effects' variation over time. In particular, a small value of σ_δ^2 indicates that most social trajectories are flat with little dynamic variability. We place the following conjugate priors on the variance parameters: $\tau_\delta^2 \sim \Gamma^{-1}(a_{\tau_\delta^2}/2, b_{\tau_\delta^2}/2)$ and $\sigma_\delta^2 \sim \Gamma^{-1}(c_{\sigma_\delta^2}/2, d_{\sigma_\delta^2}/2)$.

2.3 Dynamic Latent Features Shared Between Layers

Like other latent space models, we assume that the probability of two actors forming a connection depends on their latent representations in an unobserved latent space. Specifically, we assign every actor a latent feature $\mathbf{X}_t^i \in \mathbb{R}^d$ at each time point. Relations in dynamic networks typically have strong autocorrelations wherein the dyadic relations and latent features slowly vary over time. These autocorrelations are captured by a distribution that assumes the latent positions propagate through time via a shared Markov process:

$$\mathbf{X}_1^i \overset{\text{iid}}{\sim} N(\mathbf{0}_d, \Psi_0), \qquad \mathbf{X}_t^i \sim N(\mathbf{X}_{t-1}^i, \sigma^2 I_d), \qquad t = 2, \ldots, T.$$

Sewell and Chen (2015) first proposed this Gaussian random-walk process for the latent positions in the context of single-layer ($K = 1$) dynamic networks. Intuitively, these dynamics assume that changes in the network's connectivity patterns are partly due to changes in the actor's latent features. Unlike Sewell and Chen (2015), we do not restrict the initial covariance matrix $\Psi_0 \in \mathbb{R}^{d\times d}$ to be spherical, that is, a constant multiple of the identity matrix. There are two reasons for this choice. First, recall that the homophily coefficients take on values ± 1 for the reference layer. Therefore, the different variances (or scales) assigned to each latent dimension by the non-spherical covariance matrix reflect their influence on the reference layer's bilinear term. Second, Hoff et al. (2002) recommended a spherical covariance matrix because of the latent positions' rotational invariance. This argument is inappropriate in our context because the latent positions are identified up to a signed-permutation. Lastly, σ^2 measures the step size of each latent position's Gaussian random-walk. We place the following conjugate priors on the variance parameters: $\Psi_0 \sim \text{Wishart}^{-1}(\nu, V)$ and $\sigma^2 \sim \Gamma^{-1}(c_{\sigma^2}/2, d_{\sigma^2}/2)$.

2.4 Layer-Specific Homophily Levels

The proposed model posits that the multilayer networks are correlated because they share a single set of latent positions $\mathcal{X}_t$ among all layers. Mathematically, this restriction allows the model to capture common structures across layers. The homophily coefficients $\Lambda_k = \text{diag}(\boldsymbol{\lambda}_k)$ for $\boldsymbol{\lambda}_k \in \mathbb{R}^d$ allow for variability between the layers. Intuitively, the model assumes that two relations differ because they put distinct weights on the latent features. For example, homophilic features in a friendship relation may be heterophilic in a combative relation. For interpretability, we restrict Λ_k to a diagonal matrix. We place independent multivariate Gaussian priors on the diagonal elements

$$\boldsymbol{\lambda}_k \overset{\text{iid}}{\sim} N(\mathbf{0}_d, \sigma_\lambda^2 I_d), \qquad k = 2, \ldots, K.$$

For identifiability reasons, the homophily coefficients take values of ± 1 in the first layer. We enforce this reference layer constraint by re-parameterizing the reference layer's diagonal elements in terms of Bernoulli random variables

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

where ρ is the prior probability of an assortative relationship along a latent dimension. Under this constraint, we interpret the other layer's homophily coefficients in comparison to the reference. For example, if $\lambda_{11} = -1$ and $\lambda_{21} = 2$, then the second layer weights dimension one twice as heavily as the reference layer while exhibiting homophily instead of heterophily.

2.5 Identifiability

Here, we present sufficient conditions for identifiability and their implications on inference. For bilinear latent space models with sociality effects, it is natural to require the matrix of latent positions to be centered and full rank (Zhang et al., 2020; Macdonald et al., 2022). In addition to the previous conditions, Theorem 1 shows that restricting the reference layer's homophily coefficients to take values ± 1 and placing a mild distinctness condition on the remaining homophily coefficients is sufficient to identify our model up to a signed-permutation of the latent space, where the permutation is common to all time points and the sign-flips can change between time-points. We provide the proof in Appendix B.

Theorem 1 (Identifiability Conditions) *Suppose that two sets of parameters* $\{\boldsymbol{\delta}_{1:K,1:T}, \mathcal{X}_{1:T}, \Lambda_{1:K}\}$ *and* $\{\tilde{\boldsymbol{\delta}}_{1:K,1:T}, \tilde{\mathcal{X}}_{1:T}, \tilde{\Lambda}_{1:K}\}$ *satisfy the following conditions:*

A1. $J_n\mathcal{X}_t = \mathcal{X}_t$ *and* $J_n\tilde{\mathcal{X}}_t = \tilde{\mathcal{X}}_t$ *for* $t = 1, \ldots, T$ *where* $J_n = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$.

A2. $\mathrm{rank}(\mathcal{X}_t) = \mathrm{rank}(\tilde{\mathcal{X}}_t) = d$ *for* $t = 1, \ldots, T$.

A3. *For one* $r \in \{1, \ldots, K\}$, $\Lambda_r = I_{p,q}$ *and* $\tilde{\Lambda}_r = I_{p',q'}$.

A4. *For at least one layer* $\ell \neq r$, $\mathrm{rank}(\Lambda_\ell) = \mathrm{rank}(\tilde{\Lambda}_\ell) = d$ *and both* $\Lambda_\ell\Lambda_r$ *and* $\tilde{\Lambda}_\ell\tilde{\Lambda}_r$ *have distinct diagonal elements, i.e.,* $(\Lambda_\ell\, I_{p,q})_{gg} \neq (\Lambda_\ell\, I_{p,q})_{hh}$ *and* $(\tilde{\Lambda}_\ell I_{p',q'})_{gg} \neq (\tilde{\Lambda}_\ell\, I_{p',q'})_{hh}$ *for* $1 \leq g \neq h \leq d$.

Then the model is identifiable up to a signed-permutation of the latent space, that is, if for all $1 \leq k \leq K$ *and* $1 \leq t \leq T$ *we have that*

$$\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\Lambda_k\mathcal{X}_t^{\mathrm{T}} = \tilde{\boldsymbol{\delta}}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\tilde{\boldsymbol{\delta}}_{k,t}^{\mathrm{T}} + \tilde{\mathcal{X}}_t\tilde{\Lambda}_k\tilde{\mathcal{X}}_t^{\mathrm{T}},$$

then $I_{p,q} = I_{p',q'}$ *and for all* $1 \leq k \leq K$ *and* $1 \leq t \leq T$ *we have that*

$$\tilde{\boldsymbol{\delta}}_{k,t} = \boldsymbol{\delta}_{k,t},\ \tilde{\mathcal{X}}_t = \mathcal{X}_t P\,\mathrm{diag}(\mathbf{s}_t),\ \tilde{\Lambda}_k = P^{\mathrm{T}}\Lambda_k P,$$

where $\mathbf{s}_t \in \{\pm 1\}^d$, $P = \mathrm{BlockDiagonal}(P_1, P_2)$ *is a* $d \times d$ *block-diagonal permutation matrix, and* P_1 *and* P_2 *are permutation matrices on* p *and* q *elements, respectively.*

Assumption A1 alone, which centers the latent space, is sufficient to remove any confounding between the social trajectories and the latent positions. This issue arises because the likelihood is invariant to translations in the latent space. Indeed,

$$\begin{aligned}\delta_{k,t}^i + \delta_{k,t}^j + \mathbf{X}_t^{i\,\mathrm{T}}\Lambda_k\mathbf{X}_t^j &= \delta_{k,t}^i + \delta_{k,t}^j + (\mathbf{X}_t^i - \mathbf{c} + \mathbf{c})^{\mathrm{T}}\Lambda_k(\mathbf{X}_t^j - \mathbf{c} + \mathbf{c}),\\ &= \tilde{\delta}_{k,t}^i + \tilde{\delta}_{k,t}^j + \tilde{\mathbf{X}}_t^{i\,\mathrm{T}}\Lambda_k\tilde{\mathbf{X}}_t^j,\end{aligned}$$

where $\tilde{\mathbf{X}}_t^i = \mathbf{X}_t^i - \mathbf{c}$ and $\tilde{\delta}_{k,t}^i = \delta_{k,t}^i + \tilde{\mathbf{X}}_t^{i\,\mathrm{T}} \Lambda_k \mathbf{c} + \mathbf{c}^{\mathrm{T}} \Lambda_k \mathbf{c}/2$. Such confounding is present in previous bilinear latent space models that treat the latent positions as random effects (Hoff et al., 2002; Krivitsky et al., 2009). Our prior specification does not directly enforce the centering constraint. Instead, we let all the parameters float because including this redundancy can speed up the variational algorithm proposed in the next section (Liu and Wu, 1999; van Dyk and Meng, 2001; Qi and Jaakkola, 2006). However, when summarizing the model, we want to identify the social and latent trajectories so that the sociality effects are no longer confounded with the bilinear term. Therefore, after estimation, we perform posterior inference on $\tilde{\mathbf{X}}_t^i$ and $\tilde{\delta}_{k,t}^i$ with $\mathbf{c} = (1/n) \sum_{i=1}^n \mathbf{X}_t^i$. See Section 3.5 for details.

Assumptions A1—A4 are strong enough to identify each $\mathcal{X}_t$ up to sign-flips and permutations of its columns and the homophily coefficients $\Lambda_k = \mathrm{diag}(\boldsymbol{\lambda}_k)$ up to the same set of permutations applied to the rows of $\boldsymbol{\lambda}_k$. Furthermore, the same permutation P applies to all time points. Standard results for the generalized random dot product graph model (Rubin-Delanchy et al., 2022) suggest that Assumptions A1—A3 are sufficient to identify the latent space up to an indefinite orthogonal transformation, which differs across time points. Theorem 1 states that the mild distinctness condition in Assumption A4 is sufficient to fix the orientation of these indefinite orthogonal transformations up to a signed-permutation, where the permutation is common to all time points but the sign-flips still vary over time.

Intuitively, Assumption A4 asserts that at least one layer should measure a homophily pattern distinct from the reference layer. For example, Assumption A4 is not satisfied for layers whose homophily coefficients satisfy $\Lambda_k = \alpha I_{p,q}$ for any scalar α. While mildly restrictive, we expect Assumption A4 to hold when the layers measure different phenomena. For example, we expect the cooperation and conflict layers that make up the international relations network studied in the real data analysis of Section 5 to have distinct homophily patterns. Most importantly, Assumption A4 holds with probability one under our choice of priors. As such, we will assume the model satisfies Assumptions A1—A4 going forward. For an analysis of parameter identifiability without Assumption A4, see Appendix A.

Lastly, although Assumption A3 is always theoretically satisfied, one does not know a priori that the reference layer chosen in applications has full rank. Incorrect selection of a reference layer that is not full rank can result in under-performance due to model misspecification. As a heuristic, we check that the maximum rank of the reference layer's adjacency matrices over all time points, i.e., $\max_{1 \le t \le T} \mathrm{rank}(\mathbf{Y}_t^r)$, is the largest among all the layers. One may also check the sensitivity of the results to the choice of reference layer.

2.6 Connections to Existing Work in Special Cases ($T = 1$ or $K = 1$)

Dynamic multilayer networks encompass static multilayer networks ($T = 1$) and single-layer dynamic networks ($K = 1$). As such, our proposed model naturally generalizes or connects to existing models in these cases. When we observe a static multilayer network ($T = 1$), the decomposition in Equation (2) resembles the one proposed by Zhang et al. (2020). Unlike our decomposition, Zhang et al. (2020) allowed Λ_k to be an unconstrained $d \times d$ matrix but constrained $\mathcal{X}_1$ to be a scaled semi-orthogonal matrix, i.e., $n^{-1} \mathcal{X}_1^{\mathrm{T}} \mathcal{X}_1 = I_d$. A temporal extension of this orthogonality constraint is unwieldy for the development of a computationally efficient Bayesian model. For example, existing priors on dynamic processes for semi-orthogonal matrices (Chikuse, 2006) lead to impractical doubly-intractable posterior

distributions that often require computationally inefficient estimation methods (Murray et al., 2006; Møller et al., 2006). As such, we avoid such complexities by introducing an alternative decomposition in Equation (2) that remains meaningfully identifiable and allows for efficient inference through a novel structured mean-field variational inference algorithm.

For single-layer dynamic networks ($K = 1$), the model reduces to a latent space model for dynamic networks under a bilinear similarity measure (Durante and Dunson, 2014; Sewell and Chen, 2017); however, unlike these models, we include time-varying sociality effects. Although single-layer dynamic networks are not the focus of this work, a significant contribution we make to the field of Bayesian dynamic latent space models is the introduction of a structured mean-field variational inference algorithm. As we elaborate on in the next section, this approach provides a higher fidelity approximation to the parameter's posterior distribution compared to existing approaches while matching their time complexity. Regardless, our primary contribution is a Bayesian model for dynamic multilayer network data, a data type that none of these existing methods can handle, along with a computationally efficient estimation method, which is lacking in the existing literature.

3. Variational Inference

Now, we turn to the problem of parameter estimation and inference. We take a Bayesian approach to inference with the goal of providing both posterior mean and credible intervals for the model's parameters. However, the large amount of dyadic relations that comprise multilayer dynamic networks makes Markov chain Monte Carlo inference impractical for all but small networks. For this reason, we employ a variational approach (Wainwright and Jordan, 2008). For notational convenience, we collect the latent variables in $\boldsymbol{\theta} = \{\boldsymbol{\delta}_{1:K,1:T}, \Lambda_{1:K}, \mathcal{X}_{1:T}\}$ and the state space parameters in $\boldsymbol{\phi} = \{\Psi_0, \sigma^2, \tau_\delta^2, \sigma_\delta^2\}$.

We aim to approximate the intractable posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{Y}_{1:T}^1, \ldots, \mathbf{Y}_{1:T}^K)$ with a tractable variational distribution $q(\boldsymbol{\theta}, \boldsymbol{\phi})$ that minimizes the KL divergence between $q(\boldsymbol{\theta}, \boldsymbol{\phi})$ and $p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{Y}_{1:T}^1, \ldots, \mathbf{Y}_{1:T}^K)$. It can be shown that minimizing this divergence is equivalent to maximizing the evidence lower bound (ELBO), a lower bound on the data's marginal log-likelihood

$$\mathcal{L}(q) = \mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{\phi})}\left[\log p(\mathbf{Y}_{1:T}^1, \ldots, \mathbf{Y}_{1:T}^K, \boldsymbol{\theta}, \boldsymbol{\phi}) - \log q(\boldsymbol{\theta}, \boldsymbol{\phi})\right] \leq \log p(\mathbf{Y}_{1:T}^1, \ldots, \mathbf{Y}_{1:T}^K).$$

In general, the ELBO is not concave; however, optimization procedures often converge to a reasonable optimum. One still has the flexibility to specify the variational distribution's form, although the need to evaluate and sample from it often guides this choice. A convenient form is the structured mean-field approximation, which factors $q(\boldsymbol{\theta}, \boldsymbol{\phi})$ into a product over groups of dependent latent variables. Furthermore, when the model consists of conjugate exponential family distributions, this form lends itself to a simple coordinate ascent optimization algorithm with optimal closed-form coordinate updates. For an introduction to variational inference, see Blei et al. (2017).

In what follows, we present a structured mean-field variational inference algorithm that preserves the eignmodel's essential statistical dependencies and maintains closed-form coordinate updates. Normally, the absence of conditional conjugacy in latent space models poses a challenge for closed-form variational inference. Indeed, popular solutions require additional approximations of the expected log-likelihood (Salter-Townshend and Murphy,

2013; Gollini and Murphy, 2016; Liu and Chen, 2022), which may bias parameter estimates. Another challenge is that existing mean-field variational approximations are inadequate for our model due to the latent variable's temporal dependencies. Our solution employs Pólya-gamma augmentation (Polson et al., 2013) and variational Kalman smoothing (Beal, 2003) to produce a new and widely applicable variational inference algorithm for bilinear latent space models for dynamic networks. As we elaborate in Section 3.3.5, an additional benefit of using Kalman smoothing is that the proposed structured mean-field algorithm has the same time complexity as existing variational algorithms for dynamic latent space models.

3.1 Pólya-gamma Augmentation

As previously mentioned, we use Pólya-gamma augmentation to render the model conditionally conjugate. For each dyad in the dynamic multilayer network, we introduce auxiliary Pólya-gamma latent variables $\omega_{ijt}^k \overset{\text{iid}}{\sim} \text{PG}(1,0)$, where $\text{PG}(b,c)$ denotes a Pólya-gamma distribution with parameters $b > 0$ and $c \in \mathbb{R}$. For convenience, we use $\boldsymbol{\omega}$ to denote the collection of all Pólya-gamma auxiliary variables. As shown in Polson et al. (2013), the joint distribution is now proportional to

$$p(\mathbf{Y}_{1:T}^1, \ldots, \mathbf{Y}_{1:T}^k, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega}) \propto p(\boldsymbol{\theta} \mid \boldsymbol{\phi}) p(\boldsymbol{\phi}) p(\boldsymbol{\omega}) \prod_{k=1}^{K} \prod_{t=1}^{T} \prod_{j<i} \exp\left\{ z_{ijt}^k \psi_{ijt}^k - \omega_{ijt}^k (\psi_{ijt}^k)^2 / 2 \right\},$$

where $z_{ijt}^k = Y_{ijt}^k - 1/2$ and $\psi_{ijt}^k = \delta_{k,t}^i + \delta_{k,t}^j + \mathbf{X}_t^{i\,\mathrm{T}} \Lambda_k \mathbf{X}_t^j$. This joint distribution results in each latent variable's full conditional distribution lying within the exponential family, a property sufficient for closed-form variational inference.

3.2 The Structured Mean-Field Approximation

We use the following structured mean-field (SMF) approximation to the augmented model's posterior

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega}) = \left[\prod_{h=1}^{d} q(\lambda_{1h}) \right] \left[\prod_{k=2}^{K} q(\boldsymbol{\lambda}_k) \right] \left[\prod_{k=1}^{K} \prod_{i=1}^{n} q(\delta_{k,1:T}^i) \right] \left[\prod_{i=1}^{n} q(\mathbf{X}_{1:T}^i) \right] \left[\prod_{k=1}^{K} \prod_{t=1}^{T} \prod_{j<i} q(\omega_{ijt}^k) \right] \times q(\Psi_0) q(\sigma^2) q(\tau_\delta^2) q(\sigma_\delta^2). \tag{3}$$

This factorization is attractive because it maintains the essential temporal dependencies in the posterior distribution. Since we use optimal variational factors, preserving these dependencies increases the approximate posterior distribution's accuracy.

In contrast, existing variational approximations (Sewell and Chen, 2017; Liu and Chen, 2022) for dynamic latent space models use the following mean-field (MF) variational family for the latent positions:

$$\prod_{i=1}^{n} q(\mathbf{X}_{1:T}^i) = \prod_{i=1}^{n} \prod_{t=1}^{T} q(\mathbf{X}_t^i).$$

Unlike the proposed SMF approximation, the MF approximation assumes that all nodes' latent positions are independent across all time points. Despite its adoption in the literature, the MF approximation has known statistical and computational deficiencies. Most

importantly, Wang and Titterington (2004) showed that enforcing temporal independence in the variational approximation can lead to inconsistent estimation in dynamic state-space models when there is dependence between the latent states over time. Furthermore, removing strong posterior dependencies in the variational family decreases the fidelity of the approximate posterior, which can result in less accurate (often too narrow) approximate credible intervals. Lastly, mean-field approximations have more local optima than their structured mean-field counterparts (Wainwright and Jordan, 2008; Hoffman et al., 2013), which makes optimization more challenging under the MF approximation.

3.3 Coordinate Ascent Variational Inference Algorithm

To maximize the ELBO, we employ coordinate ascent variational inference (CAVI). CAVI performs coordinate ascent on one variational factor at a time, holding the rest fixed. The optimal coordinate updates take a simple form: set each variational factor to the corresponding latent variable's expected full conditional probability under the remaining factors. For example, the update for $q(\mathbf{X}^i_{1:T})$ is given by

$$\log q(\mathbf{X}^i_{1:T}) = \mathbb{E}_{-q(\mathbf{X}^i_{1:T})}\left[\log p(\mathbf{X}^i_{1:T} \mid \cdot)\right] + c,$$

where $\mathbb{E}_{-q(\mathbf{X}^i_{1:T})}[\cdot]$ indicates an expectation taken with respect to all variational factors except $q(\mathbf{X}^i_{1:T})$, $p(\mathbf{X}^i_{1:T} \mid \cdot)$ is the full conditional distribution of $\mathbf{X}^i_{1:T}$, and c is a normalizing constant. When the full conditionals are members of the exponential family, a coordinate update involves calculating the natural parameter's expectations under the remaining variational factors.

The CAVI algorithm alternates between optimizing $q(\boldsymbol{\omega})$, $q(\boldsymbol{\delta}_{1:K,1:T})$, $q(\mathcal{X}_{1:T})$, $q(\Lambda_{1:K})$, and $q(\boldsymbol{\phi})$. Algorithm 1 outlines the full CAVI algorithm and defines some notation used throughout the rest of the article. We summarize each variational factor's coordinate update in the following sections. Appendix C and Appendix D provide the full details and derivations of the coordinate updates and the variational Kalman smoothers, respectively.

3.3.1 Updating $q(\omega^k_{ijt})$

By the exponential tilting property of the Pólya-gamma distribution, we have

$$\log q(\omega^k_{ijt}) = \mathbb{E}_{-q(\omega^k_{ijt})}\left[p_{\mathrm{PG}}(\omega^k_{ijt} \mid 1, \psi^k_{ijt})\right] + c,$$

where $p_{PG}(\omega \mid b, c)$ is the density of $\mathrm{PG}(b, c)$ random variable. This density is a member of the exponential family with natural parameter $-(\psi^k_{ijt})^2/2$. We provide the full coordinate update, which involves taking the expectation of $(\psi^k_{ijt})^2$, in Algorithm 2 of Appendix C.

3.3.2 Updating $q(\delta^i_{k,1:T})$, $q(\tau^2_\delta)$, $q(\sigma^2_\delta)$

Under the Pólya-gamma augmentation scheme, the conditional distributions of the social trajectories take the form of linear Gaussian state space models. In particular,

$$\log q(\delta^i_{k,1:T}) = \log h(\delta^i_{k,1}) + \sum_{t=2}^{T} \log h(\delta^i_{k,t} \mid \delta^i_{k,t-1}) + \sum_{t=1}^{T} \log h(\mathbf{z}^i_{k,t} \mid \delta^i_{k,t}) + c, \tag{4}$$

Define the following expectations taken with respect to the full variational posterior:

$$\mathbb{E}\left[\mathbf{X}_t^i\right] = \boldsymbol{\mu}_t^i, \quad \mathrm{Var}(\mathbf{X}_t^i) = \Sigma_t^i, \quad \mathrm{Cov}(\mathbf{X}_t^i, \mathbf{X}_{t+1}^i) = \Sigma_{t,t+1}^i,$$
$$\mathbb{E}\left[\delta_{k,t}^i\right] = \mu_{\delta_{k,t}^i}, \quad \mathrm{Var}(\delta_{k,t}^i) = \sigma^2_{\delta_{k,t}^i}, \quad \mathrm{Cov}(\delta_{k,t}^i, \delta_{k,t+1}^i) = \sigma^2_{\delta_{k,t,t+1}^i},$$
$$\mathbb{E}\left[\boldsymbol{\lambda}_k\right] = \boldsymbol{\mu}_{\boldsymbol{\lambda}_k}, \quad \mathrm{Var}(\boldsymbol{\lambda}_k) = \Sigma_{\boldsymbol{\lambda}_k}, \quad \mathbb{E}\left[\omega_{ijt}^k\right] = \mu_{\omega_{ijt}^k}.$$

Iterate the following steps until convergence:

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

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

   using a variational Kalman smoother as in Algorithm 3.
3. Update

   $$q(\mathbf{X}_{1:T}^i) : \text{ a Gaussian state space model for } i \in \{1,\dots,n\},$$
   $$q(\Psi_0) = \text{Wishart}^{-1}(\bar{\nu}, \bar{V}),$$
   $$q(\sigma^2) = \Gamma^{-1}(\bar{c}_{\sigma^2}/2, \bar{d}_{\sigma^2}/2),$$

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

Algorithm 1: Coordinate ascent variational inference for the eigenmodel for dynamic multilayer networks. Appendix C contains the details of Algorithms 2—5. Iterations are performed until successive differences of the expected log-likelihood, Equation (6), drop below a tolerance threshold.

where

$$\begin{aligned}
\log h(\delta^i_{k,1}) &= \mathbb{E}_{q(\tau^2_\delta)}\left[\log N(\delta^i_{k,1} \mid 0, \tau^2_\delta)\right], \\
\log h(\delta^i_{k,t} \mid \delta^i_{k,t-1}) &= \mathbb{E}_{q(\sigma^2_\delta)}\left[\log N(\delta^i_{k,t} \mid \delta^i_{k,t-1}, \sigma^2_\delta)\right], \\
\log h(\mathbf{z}^i_{k,t} \mid \delta^i_{k,t}) &= \mathbb{E}_{-q(\delta^i_{k,1:T})}\left[\sum_{j\neq i} \log N(z^k_{ijt} \mid \omega^k_{ijt}\,\delta^i_{k,t} + \omega^k_{ijt}(\delta^j_{k,t} + \mathbf{X}^{i\,\mathrm{T}}_t \Lambda_k \mathbf{X}^j_t),\ \omega^k_{ijt})\right].
\end{aligned}$$

In the previous expressions, $\mathbf{z}^i_{k,t} \in \mathbb{R}^{n-1}$ is a vector that consists of stacking z^k_{ijt} for $j \neq i$ and $N(\mathbf{x} \mid \boldsymbol{\mu}, \Sigma)$ is the density of a $N(\boldsymbol{\mu}, \Sigma)$ random variable. Because all densities involved are Gaussian, the expectations yield Gaussian densities with natural parameters that depend on the remaining variational factors. Thus, we recognize the optimal variational distribution as a GSSM. The expected sufficient statistics needed to update the remaining variational factors can be computed with either the variational Kalman smoother (Beal, 2003) or a standard Kalman smoother under an augmented state space model (Barber and Chiappa, 2007). We use the variational Kalman smoother. Furthermore, the inverse-gamma priors on the state space parameters result in fully conjugate coordinate updates for τ^2_δ and σ^2_δ. The update for the social trajectories is presented in Algorithm 3 of Appendix C.

3.3.3 Updating $q(\mathbf{X}^i_{1:T})$, $q(\Psi_0)$, $q(\sigma^2)$

Similar to the social trajectories, the conditional distributions of the latent trajectories are also GSSMs. Specifically,

$$\log q(\mathbf{X}^i_{1:T}) = \log h(\mathbf{X}^i_1) + \sum_{t=2}^{T} \log h(\mathbf{X}^i_t \mid \mathbf{X}^i_{t-1}) + \sum_{t=1}^{T} \log h(\mathbf{z}^i_t \mid \mathbf{X}^i_t) + c, \tag{5}$$

where

$$\begin{aligned}
\log h(\mathbf{X}^i_1) &= \mathbb{E}_{q(\Psi_0)}\left[\log N(\mathbf{X}^i_1 \mid \mathbf{0}_d, \Psi_0)\right], \\
\log h(\mathbf{X}^i_t \mid \mathbf{X}^i_{t-1}) &= \mathbb{E}_{q(\sigma^2)}\left[\log N(\mathbf{X}^i_t \mid \mathbf{X}^i_{t-1}, \sigma^2 I_d)\right], \\
\log h(\mathbf{z}^i_t \mid \mathbf{X}^i_t) &= \mathbb{E}_{-q(\mathbf{X}^i_{1:T})}\left[\sum_{k=1}^{K}\sum_{j\neq i} \log N(z^k_{ijt} \mid \omega^k_{ijt}(\delta^i_{k,t} + \delta^j_{k,t}) + \omega^k_{ijt}\mathbf{X}^{j\,\mathrm{T}}_t \Lambda_k \mathbf{X}^i_t,\ \omega^k_{ijt})\right].
\end{aligned}$$

In the previous expressions, $\mathbf{z}^i_t \in \mathbb{R}^{K(n-1)}$ is a vector formed by stacking z^k_{ijt} for $j \neq i$ and $k = 1, \ldots, K$. Again, we recognize that $q(\mathbf{X}^i_{1:T})$ is a GSSM; therefore, we can calculate the expected sufficient statistics with the variational Kalman smoother. Also, the inverse-Wishart and inverse-gamma priors on Ψ_0 and σ^2 result in closed form coordinate updates. The updates for the latent trajectories are presented in Algorithm 4 of Appendix C.

3.3.4 Updating $q(\Lambda_k)$

Given the augmented model's conjugacy, the homophily coefficients will be Bernoulli for the reference layer and Gaussian for the other layers. The corresponding coordinate updates, which involve calculating the Bernoulli probabilities and performing standard Bayesian linear regression, are presented in Algorithm 5 of Appendix C.

3.3.5 Time Complexity

The closed-form Kalman smoothing updates lead to an efficient SMF algorithm with the same time complexity as the MF algorithm. Note that the MF and SMF algorithms only significantly differ in the updates for $q(\mathcal{X}_{1:T})$ and $q(\delta_{1:K,1:T})$. For each social trajectory, the time complexity for Kalman smoothing is $O(T)$, and the time complexity to calculate the sufficient statistics is $O(nT)$. As such, the time complexity to update all nK social trajectories under the SMF algorithm is $O(nKT + n^2KT) = O(n^2KT)$. Similarly, for each latent trajectory, the time complexity for Kalman smoothing is $O(T)$, and the time complexity to calculate the sufficient statistics is $O(nKT)$. As such, the time complexity to update all n latent trajectories under the SMF algorithm is $O(nT + n^2KT) = O(n^2KT)$. We present the CAVI updating scheme for the MF variational family in Appendix F, which also has a time complexity of $O(n^2KT)$. As such, the two algorithms only differ by a constant factor, which we empirically quantify in Section 4. These results show that the SMF algorithm is 3.5 to 2.5 times slower than the MF algorithm for most network sizes.

3.4 Convergence Criteria

Although it is possible to calculate the ELBO to determine convergence, evaluating the state space terms is computationally expensive. Instead, we monitor the expected log-likelihood

$$\mathcal{F}(q) = \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{j<i}(Y_{ijt}^k - 1/2)\mathbb{E}_{q(\boldsymbol{\theta})}\left[\psi_{ijt}^k\right] - \frac{1}{2}\mathbb{E}_{q(\omega_{ijt}^k)}\left[\omega_{ijt}^k\right]\mathbb{E}_{q(\boldsymbol{\theta})}\left[(\psi_{ijt}^k)^2\right], \tag{6}$$

which upper bounds the ELBO. We say the algorithm converged when the difference in the expected log-likelihood is less than 10^{-2} between iterations or the number of iterations exceeded 1,000. Due to the ELBO's non-convexity, we run the algorithm with four different initializations (one non-random and three random) and choose the model with the highest AUC (area under the receiver operating characteristic curve) for predicting in-sample edges. For details on our initialization procedure and hyper-parameter settings, see Appendix E.

3.5 Inference of Identifiable Parameters

Recall that a centered latent space is a sufficient condition for parameter identifiability. As such, we make inference on the following parameters based on the approximate posterior:

$$\tilde{\mathbf{X}}_t^i = \mathbf{X}_t^i - \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_t^j,$$
$$\tilde{\delta}_{k,t}^i = \delta_{k,t}^i + \tilde{\mathbf{X}}_t^{i\,\mathrm{T}}\Lambda_k\mathbf{c} + \frac{1}{2}\mathbf{c}^{\mathrm{T}}\Lambda_k\mathbf{c},$$

where $\mathbf{c} = (1/n)\sum_{j=1}^{n} \mathbf{X}_t^j$. Under our approximation, the marginal posterior distributions of the $\tilde{\mathbf{X}}_t^i$'s are Gaussian with moments

$$\begin{aligned} \mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\omega})}\left[\tilde{\mathbf{X}}_t^i\right] &= \tilde{\boldsymbol{\mu}}_t^i = \boldsymbol{\mu}_t^i - \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{\mu}_t^j, \\ \mathrm{Var}(\tilde{\mathbf{X}}_t^i) &= \tilde{\Sigma}_t^i = \left(1-\frac{1}{n}\right)^2 \Sigma_t^i + \left(\frac{1}{n}\right)^2 \sum_{j\neq i} \Sigma_t^j, \end{aligned} \tag{7}$$

where the variance is with respect to $q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega})$ as well. We calculate each $\tilde{\delta}_{k,t}^i$'s posterior mean and 95% credible interval using 2,500 samples from the approximate posterior distribution because their approximate posterior distributions lack an analytic form.

3.6 Choice of Latent Space Dimension

Similar to other latent space models, selection of the latent space dimension d depends on the purpose of the analysis. If the goal is exploratory or descriptive, then setting $d = 2$ allows for easy visualization of the latent space. However, when the goal is predictive, the choice of dimension can significantly influence the model's performance. In this case, we recommend using information criteria to estimate d. We found that the Akaike information criterion (AIC) outperformed competing criteria on simulated data, see Appendix H for details. We used this method to select the dimension d for the real data examples in Section 5.

4. Simulation Studies

We perform multiple simulation studies to evaluate the performance of the proposed structured variational inference algorithm for inferring the eigenmodel for dynamic multilayer network's parameters. In each simulation study, we generated parameter settings as follows:

1. Generate the reference homophily coefficients: $\lambda_{1h} = 2u_h - 1$ for $1 \leq h \leq d$, where $u_h \overset{\text{iid}}{\sim} \text{Bernoulli}(0.5)$.

2. Generate the remaining homophily coefficients from a Gaussian mixture: $\lambda_{kh} \overset{\text{iid}}{\sim} 0.5\, N(-1, 0.25) + 0.5\, N(1, 0.25)$ for $2 \leq k \leq K$ and $1 \leq h \leq d$.

3. Generate initial sociality effects: $\delta_{k,1}^i \overset{\text{iid}}{\sim} U[-4, 1]$ for $1 \leq i \leq n$ and $1 \leq k \leq K$.

4. Generate the social trajectories: For $t = 2, \ldots, T$, $1 \leq i \leq n$, and $1 \leq k \leq K$, set $\delta_{k,t}^i = \delta_{k,t-1}^i + \epsilon_{k,t}^i$. We independently drew $(\epsilon_{k,2}^i, \ldots, \epsilon_{k,T}^i)^{\mathrm{T}} \sim N(\mathbf{0}_{T-1}, \sigma^2 R)$, where R is a correlation matrix with elements $R_{tt'} = \rho^{|t-t'|}$ for $1 \leq t, t' \leq T-1$.

5. Generate initial latent positions from a Gaussian mixture: For $1 \leq i \leq n$, sample $\mathbf{X}_1^i \overset{\text{iid}}{\sim} 0.5\, N(\boldsymbol{\mu}, \Sigma) + 0.5\, N(-\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (0.75, 0.75)^{\mathrm{T}}$ and $\Sigma = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$.

6. Generate the latent trajectories: For $t = 2, \ldots, T$ and $1 \leq i \leq n$, set $\mathbf{X}_t^i = \mathbf{X}_{t-1}^i + \boldsymbol{\epsilon}_t^i$. We independently drew $(\epsilon_{2g}^i, \ldots \epsilon_{Tg}^i)^{\mathrm{T}} \sim N(\mathbf{0}_{T-1}, \sigma^2 R)$ for $1 \leq g \leq d$, where R is the same correlation matrix defined in step 4.

7. Center the latent space: For $t = 1, \ldots, T$, set $\tilde{\mathbf{X}}_t^i = \mathbf{X}_t^i - (1/n)\sum_{j=1}^n \mathbf{X}_t^j$ for $1 \leq i \leq n$.

The dimension of the latent space $d = 2$ in the simulated data. The parameters σ and ρ control the simulation's temporal smoothness. Specifically, σ controls the magnitude of both the latent positions' and the sociality parameters' transitions, and ρ controls the transitions' temporal correlation. For a given set of model parameters, we sampled a single undirected adjacency matrix using the dyad-wise probabilities in Equation (1). In all simulations in the main text, we assumed the dimension d is known and did not perform dimension selection.

4.1 Parameter Recovery for Varying Network Sizes (n, K, T)

First, we present a simulation study that assessed how the proposed algorithm's estimation error scaled with network size. We considered three scenarios: *Scenario 1.* an increase in the number of nodes with $(n, K, T) \in \{50, 100, 200, 500, 1000\} \times \{5\} \times \{10\}$, *Scenario 2.* an increase in the number of layers with $(n, K, T) \in \{100\} \times \{5, 10, 20\} \times \{10\}$, and *Scenario 3.* an increase in the number of time points with $(n, K, T) \in \{100\} \times \{5\} \times \{10, 50, 100\}$. For each scenario, we fixed $\sigma = 0.05$ and $\rho = 0.4$ and sampled 50 independent networks. For an additional scenario that decreased σ as the number of time steps increased, see Appendix I.

To evaluate the estimates' accuracy, we computed relative errors between the parameters' posterior means and their true values according to the Frobenius norm. Because the true homophily coefficients are distinct, Theorem 1 states that the latent positions are identifiable up to shared column permutations and time-varying sign-flips. To account for this invariance, we calculated the latent position's time-averaged relative error as

$$\min_{P \in \Pi_d} \frac{1}{T} \sum_{t=1}^{T} \min_{\mathbf{s}_t \in \{-1,1\}^d} \frac{\|\tilde{\mathcal{X}}_t - \hat{\tilde{\mathcal{X}}}_t P \operatorname{diag}(\mathbf{s}_t)\|_F^2}{\|\tilde{\mathcal{X}}_t\|_F^2},$$

where Π_d is the set of permutation matrices on d elements, $\|\cdot\|_F$ is the Frobenius norm, $\tilde{\mathcal{X}}_t = (\tilde{\mathbf{X}}_t^1, \ldots, \tilde{\mathbf{X}}_t^i)^{\mathrm{T}}$, and $\hat{\tilde{\mathcal{X}}}_t = (\tilde{\boldsymbol{\mu}}_t^1, \ldots, \tilde{\boldsymbol{\mu}}_t^n)^{\mathrm{T}}$ where $\tilde{\boldsymbol{\mu}}_t^i$ is defined in Equation (7). Similarly, we computed the relative error of the homophily coefficients accounting for invariance under simultaneous permutations of their rows and columns:

$$\min_{P \in \Pi_d} \frac{\sum_{k=1}^K \|\Lambda_k - P^{\mathrm{T}} \operatorname{diag}(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k}) P\|_F^2}{\sum_{k=1}^K \|\Lambda_k\|_F^2}.$$

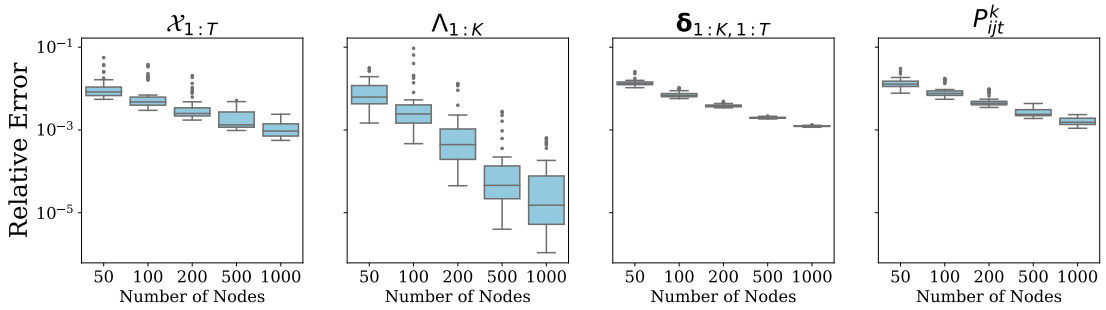
Lastly, we calculated the relative errors for the centered social trajectories and the dyad-wise probabilities, both of which do not have identifiability issues. For computational expediency, we calculated the dyad-wise probabilities by plugging-in the posterior means into Equation (1), e.g.,

$$\widehat{\mathbb{P}}(Y_{ijt}^k = 1 \mid \mu_{\delta_{k,t}^i}, \mu_{\delta_{k,t}^j}, \boldsymbol{\mu}_{\boldsymbol{\lambda}_k}, \boldsymbol{\mu}_t^i, \boldsymbol{\mu}_t^j) = \operatorname{logit}^{-1}\left[\mu_{\delta_{k,t}^i} + \mu_{\delta_{k,t}^j} + \boldsymbol{\mu}_t^{i\mathrm{T}} \operatorname{diag}(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k}) \boldsymbol{\mu}_t^j\right],$$

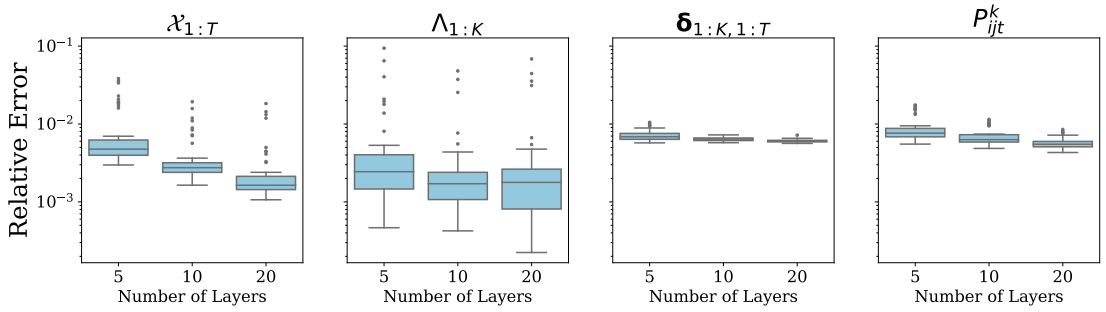
which is an upper-bound on the approximate posterior mean of the dyad-wise probability. We can use Monte Carlo to estimate the dyad-wise probabilities' approximate posterior mean by sampling from the approximate posterior if desired.

The estimation errors for varying n, K, T are displayed in the boxplots in Figure 2a, Figure 2b, and Figure 2c, respectively. Overall, the CAVI algorithm accurately recovers the

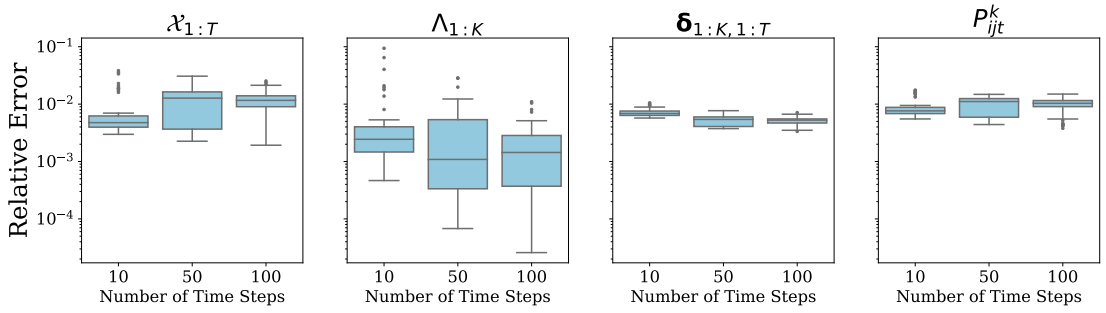
model's parameters. The starkest improvement in estimation accuracy occurs as the number of nodes increases. This improvement is partially due to the more accurate estimation of the homophily coefficients. Due to the model's ability to pool information across layers, the latent positions' relative error decreases as K increases. Such an improvement is not observed for the social trajectories because the number of social trajectories grows with the number of layers. Surprisingly, the homophily coefficients' estimation error does not improve as T increases, although the estimation error is already low at roughly 10^{-2} to 10^{-3}. Furthermore, the latent positions' estimation error slightly degrades as the number of time steps increases. Such deterioration is typical in smoothing problems.



(a) $n \in \{50, 100, 200, 500, 1000\}$, $K = 5$, $T = 10$.



(b) $n = 100, K \in \{5, 10, 20\}, T = 10$



(c) $n = 100, K = 5, T \in \{10, 50, 100\}$

Figure 2: Relative estimation errors of the model's parameters as (a) the number of nodes n increases, (b) the number of layers K increases, and (c) the number time steps T increases. The boxplots show the distribution over 50 simulations.

4.2 The Effect of Temporal Smoothness (σ^2, ρ) on Predictive Accuracy

Next, we evaluated how the parameters' smoothness over time affects the estimates' predictive accuracy. In particular, we expected the proposed methodology's predictive accuracy to improve as the parameters' temporal smoothness increases compared with methods that do not take advantage of such temporal smoothness.

4.2.1 Comparison of Joint and Separate Estimation

Here, we compare the eigenmodel for dynamic multilayer networks with a static multilayer version of the model that does not pool information over time. In particular, the alternative approach fits the following model for static multilayer networks separately at each time step $t = 1, \ldots, T$:

$$
\begin{aligned}
&Y_{ijt}^k \overset{\text{ind.}}{\sim} \text{Bernoulli}\left[\text{logit}^{-1}\left(\delta_{k,t}^i + \delta_{k,t}^j + \mathbf{X}_t^{i\,\mathrm{T}} \Lambda_{kt} \mathbf{X}_t^j\right)\right], && 1 \le j < i \le n, \quad 1 \le k \le K,\\
&\lambda_{1t,h} = 2u_h - 1, \qquad u_h \overset{\text{iid}}{\sim} \text{Bernoulli}(\rho), && 1 \le h \le d,\\
&\boldsymbol{\lambda}_{kt} \overset{\text{iid}}{\sim} N(\mathbf{0}_d, \sigma_\lambda^2 I_d), && 2 \le k \le K,\\
&\delta_{k,t}^i \overset{\text{iid}}{\sim} N(0, \tau_{\delta,t}^2), && 1 \le i \le n, \quad 1 \le k \le K,\\
&\mathbf{X}_t^i \overset{\text{iid}}{\sim} N(\mathbf{0}_d, \Psi_t), && 1 \le i \le n,\\
&\Psi_t \sim \text{Wishart}^{-1}(\nu, V), \qquad \tau_{\delta,t}^2 \sim \Gamma^{-1}(c_\delta, d_\delta),
\end{aligned}
$$

where ind. stands for independently distributed. This model is similar to the latent space model for static multilayer networks proposed by Zhang et al. (2020). The difference is that they allow Λ_{kt} to be any $d \times d$ real matrix, treat the parameters as fixed effects, and estimate them by maximizing the likelihood. Despite these similarities, we chose the above model as a comparison because it isolates the impact that the latent positions' temporal smoothness and the pooling of the shared homophily parameters across time have on estimation accuracy by only differing in this aspect. We fit the above model using a modified mean-field variational inference algorithm similar to the one proposed in Section 3, which we outline in Appendix G.

We generated 50 networks from the data-generating process outlined in the beginning of Section 4 with $n = 100$, $K = 5$, and $T = 10$ while varying both σ and ρ. Figure 3 contains the results for $\sigma \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$, and $\rho \in \{0.0, 0.4, 0.8\}$. We refer to the proposed eigenmodel for dynamic multilayer networks as the "joint" model and the static multilayer version as the "separate" model. To evaluate each model's predictive performance, we calculated the AUC for edge predictions and the sample Pearson correlation coefficient (PCC) between the true and estimated dyad-wise probabilities. Each metric was calculated on held-out data. Specifically, we randomly labeled 20% of the dyads as held-out data and removed them during estimation. The same set of dyads were removed from the network at each layer and time step. Note that AUC ranges from 0.5 to 1, and PCC ranges from -1 to 1, with larger values being better.

According to Figure 3, the proposed joint estimation procedure outperforms the separate estimation procedure for all values of σ and ρ. Specifically, the median metrics for the joint estimator are higher than the separate estimator. We see that the difference in performance

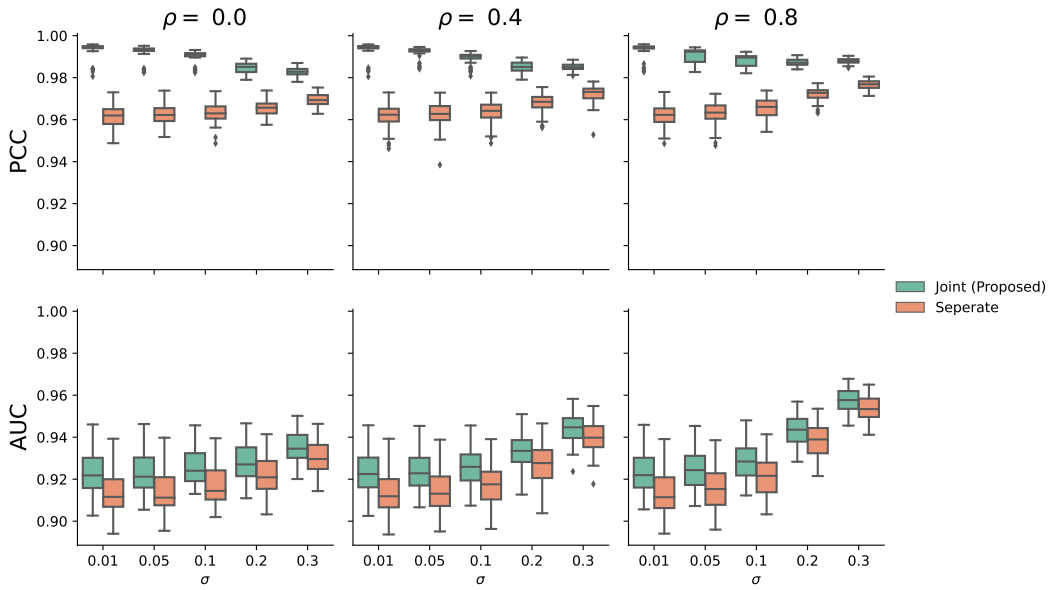


Figure 3: Comparison of joint and separate estimation for different values of σ and ρ. The first row contains boxplots of the PCC between the true and estimated link probabilities on held-out data. The second row contains boxplots of the AUC for predicting held-out edges.

decreases as the transition variance σ increases because there is less temporal similarity between the latent variables. Also, the performance gap between the two estimators appears unaffected by ρ. Lastly, we note that for networks measured over many time points (large T), the separate estimation scheme is practically more challenging because one has to solve multiple non-convex optimization problems accurately. Based on these results and practical considerations, we recommend the joint estimation scheme.

4.2.2 Comparison of the SMF and MF Variational Inference Algorithms

Here, we compare the performance of the SMF algorithm to the MF algorithm described in Section 3.2. We generated 50 networks from the process outlined at the beginning of Section 4 with $n = 100$, $K = 5$, and $T = 30$ while varying both σ and ρ. We assigned 20% of the dyads to a held-out set according to the same procedure described in the previous section. Unlike in the previous sections, we initialized both algorithms once with the non-random singular value thresholding scheme (Appendix E), so that both algorithms started from the same initial values. Each algorithm was run for a maximum of 100 iterations, and the PCC between the true and estimated link probabilities was calculated after each iteration. Figure 4 contains the results for $\sigma \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$, and $\rho \in \{0.0, 0.4, 0.8\}$.

From Figure 4, it is clear that the SMF algorithm finds a better or equivalent solution to the solution found by the MF algorithm. For small values of $\sigma \in \{0.01, 0.05, 0.1\}$, the SMF algorithm often finds a better solution in fewer iterations than the corresponding MF

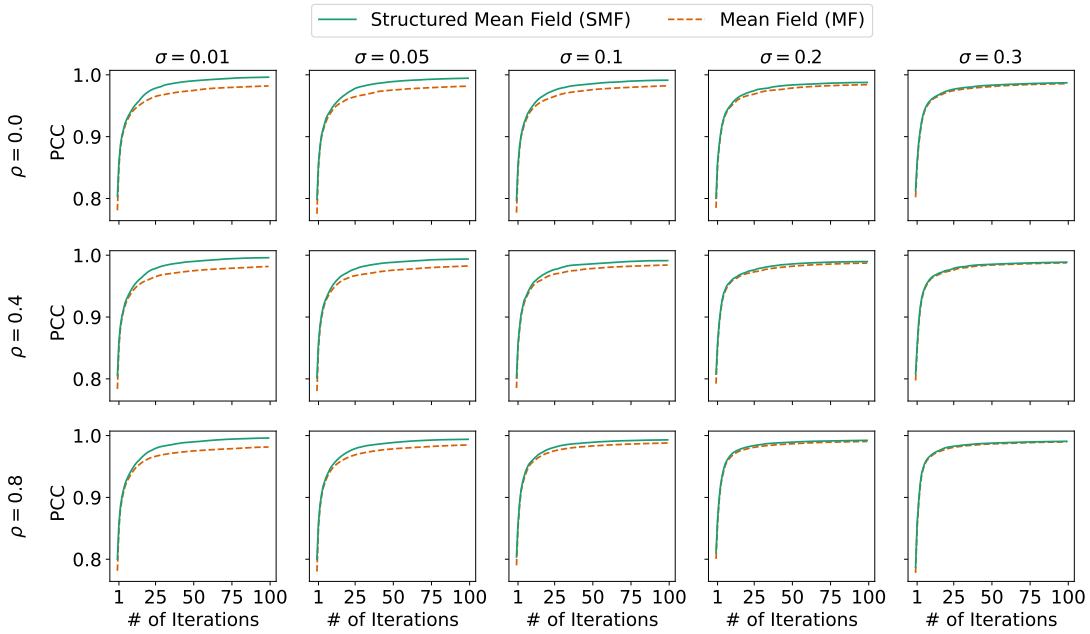


Figure 4: Performance comparison between the SMF and MF algorithms. The lines indicate the median PCC at a given iteration for the SMF (solid green) and MF (dashed red) algorithms.

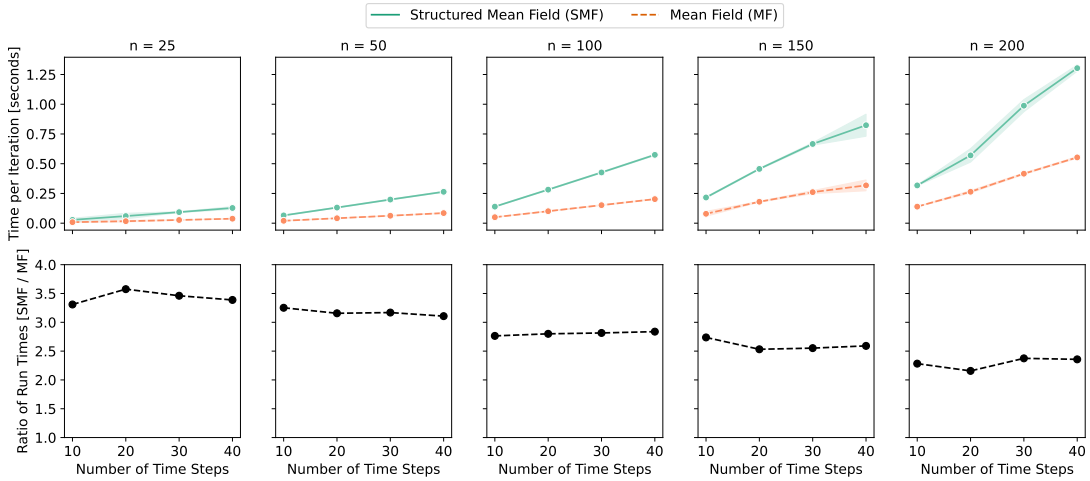


Figure 5: (Top) Median run time in seconds for one iteration of the SMF algorithm (solid green lines) and the MF algorithm (dashed red line) for $K = 5$ and various values of T and n. Lines indicate the median and the bands represent one standard deviation over 10 repetitions. (Bottom) Ratio of the median per iteration run time of the SMF algorithm compared to the MF algorithm for $K = 5$ and various values of T and n.

algorithm. The overall gap in performance decreases as σ increases, and the curves mostly overlap for larger values of $\sigma \in \{0.2, 0.3\}$. These conclusions are the same for all values of ρ. Overall, the SMF algorithm outperforms the MF algorithm when the temporal dependence is strong (i.e., σ is small), which is often the case in real-world dynamic networks.

The previous results are only meaningful if the time per iteration between the SMF and MF algorithms are comparable. As outlined in Section 3, both algorithms' per iteration time complexity is $O(n^2KT)$, so their run times only differ by a constant factor regardless of network size. Figure 5 compares the per iteration run times of the two algorithms for $K = 5$, and various values of T and n. We recorded the median per iteration run time over 100 iterations for each network and repeated this 10 times to produce the plots in Figure 5. The machine used for these benchmarks was a 2021 MacBook Pro with an Apple M1 Pro processor and a memory of 32 GB. We observe that both algorithms are linear in T. In addition, the ratio of the per iteration run time of the SMF algorithm is roughly 3.5 to 2.5 times slower than the MF algorithm, with the gap decreasing as n increases. Once we factor in the per iteration run times, we conclude that the SMF algorithm tends to find more accurate solutions faster than the MF algorithm for temporally smooth (σ is small) latent positions. However, the two solutions are similar when the latent positions are highly variable (σ is large), with the MF algorithm finding the solution faster. Overall, the SMF algorithm is computationally efficient, with the time to perform one iteration taking less than a second for most of the network sizes in this simulation study.

5. Real Data Applications

In this section, we demonstrate how to use the proposed model to analyze real-world data sets. We consider networks from political science and epidemiology. The first example studies a time series of different international relations between 100 countries over eight years. The second example applies the model to a contact network of 242 individuals at a primary school measured over two days to quantify heterogeneities in infectious disease spread throughout the school day.

5.1 International Relations

This application explores the temporal evolution of different relations between socio-political actors. The raw data consists of (source actor, target actor, event type, time-stamp) tuples collected by the Integrated Crisis Early Warning System (ICEWS) project (Boschee et al., 2015), which automatically identifies and extracts international events from news articles. The event types are labeled according to the CAMEO taxonomy (Gerner et al., 2008). The CAMEO scheme includes twenty labels ranging from the most neutral "1 — make public statement" to the most negative "20 — engage in unconventional mass violence."

Our sample consists of monthly event data between countries during the eight years of the Obama administration (2009 - 2017). We grouped the event types into four categories known as "QuadClass" (Duval and Thompson, 1980). These classes split events along four dimensions: (1) *verbal cooperation* (labels 2 to 5), (2) *material cooperation* (labels 6 to 7), (3) *verbal conflict* (labels 8 to 16), and (4) *material conflict* (labels 17 to 20). At a high-level, the first two classes represent friendly relations such as "5 — engage in diplomatic

cooperation" and "7 — provide aid", while the last two classes reflect hostile relations such as "13 — threaten" and "19 — assault".

5.1.1 Statistical Network Analysis of the ICEWS Data

We structured the ICEWS data as a dynamic multilayer network recording which four event types occurred between nations each month from 2009 until the end of 2016. We chose a monthly time-window to match previous statistical analyses of these networks (Minhas et al., 2017; He and Hoff, 2019). Each event type is a layer in the multilayer networks. We chose verbal cooperation as the reference layer because its associated adjacency matrices have the largest maximum rank across all time points. We limited the actors to the 100 most active countries during this period. This preprocessing resulted in a dynamic multilayer network with $K = 4$ layers, $T = 96$ time steps, and $n = 100$ actors. An edge ($Y^k_{ijt} = 1$) means that country i and country j had at least one event of type k during the tth month, where $t = 1$ corresponds to January 2009.

We fit six models with dimensions ranging from 1 to 6 using the SMF algorithm and determined that a latent space dimension of $d = 3$ fit the data best based on the AIC, see Figure 15 in Appendix J. The chosen model's in-sample AUC was 0.91, which indicates a good fit to the data. Since the model's latent positions are only identifiable up to a signed-permutation, the latent positions at time t (for $t \geq 2$) are matched to the positions closest to their previous positions at time $t - 1$ through a signed-permutation.

5.1.2 Detection of Historical Events During the Obama Administration

We validate the model by demonstrating that the inferred social trajectories and latent space dynamics reflect major international events. We focus on three events: the Arab Spring, the American-led intervention in Iraq, and the Crimea Crisis. Specifically, we concentrate on interpreting the latent parameters for Libya, Syria, Iraq, the United States, Russia, and Ukraine since they played a large role in these events.

Because these events involve conflict, we start by analyzing each country's material conflict social trajectory, i.e., $\delta^i_{4,1:T}$ for $1 \leq i \leq n$. Figure 6 plots these social trajectories' posterior means with a few select countries highlighted. Appendix J contains the same plot for the remaining three layers. Most social trajectories are relatively flat. Indeed, the 95% credible interval for the step size standard deviation σ_δ is $(0.0437, 0.0443)$, which is much smaller than that of the initial standard deviation τ_δ, which equals $(3.20, 3.68)$. However, the social trajectories of Iraq, Syria, and Libya demonstrate dramatic changes. Specifically, Libya and Syria both increase their material conflict sociality at the start of the Arab Spring in 2011. In particular, Libya's sociality spikes during the Libyan Civil War in 2011 that saw Muammar Gaddafi's regime overthrown. Iraq's sociality increases leading up to and throughout the United States' escalated military presence in 2014. Note that the Crimea Crisis, which began with Russia annexing the Crimea Peninsula in February 2014, is not reflected in Russia's social trajectory and only moderately reflected in Ukraine's social trajectory. This conflict is hard to detect based on the socialites because an actor's social trajectory reflects their global standing in the network while the Crimea Crisis is primarily a regional conflict. In contrast, the latent space, which captures local transitive effects, should better reflect this more localized conflict.

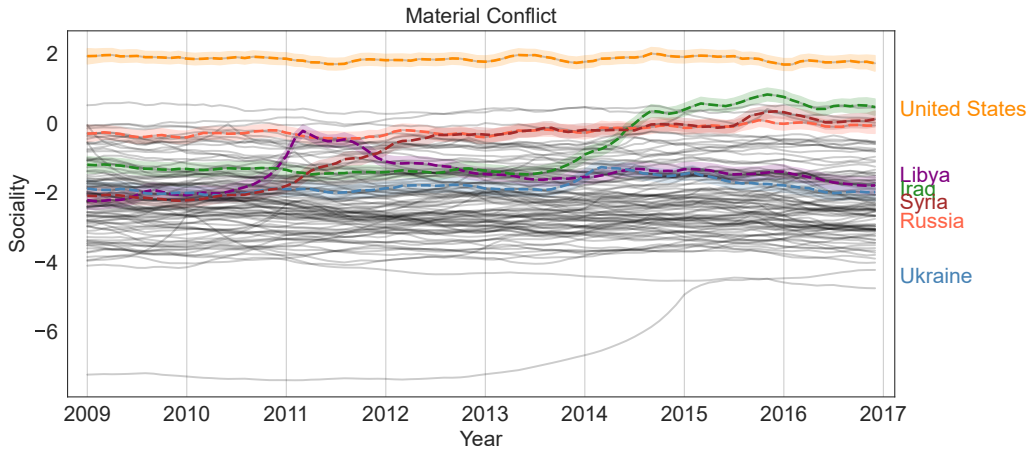


Figure 6: Posterior means of the material conflict social trajectories. Select countries are highlighted in color with bands that represent 95% credible intervals. The remaining countries' social trajectories are displayed with gray curves.

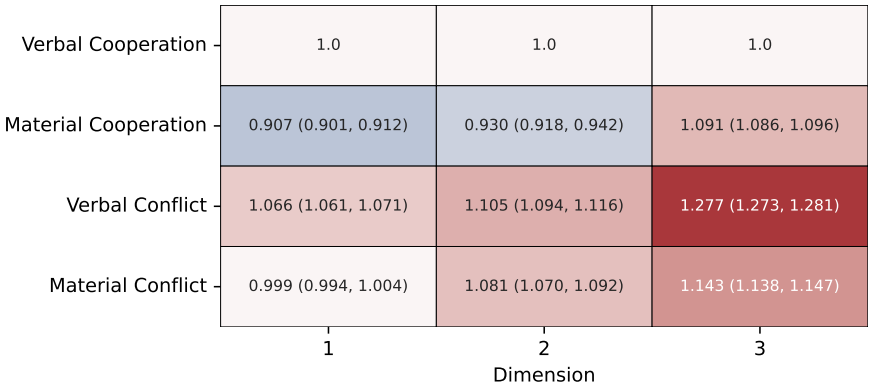


Figure 7: Heatmap of the homophily coefficients' posterior means for the ICEWS network. Each cell contains the coefficient's posterior mean and 95% credible interval. Red/blue cells indicate values greater/less than the reference layer (verbal cooperation).

We begin analyzing the latent space by interpreting the estimated homophily coefficients, Λ_k (Figure 7). All layers exhibit assortativity along all latent dimensions. Interestingly, we notice similarities in how the cooperation and the conflict layers use the second and third dimensions of the latent space. For these dimensions, the conflict layers have larger homophily coefficients than the cooperation layers, which can be seen by comparing the first and second rows to the third and fourth rows of the heatmap. To interpret this result, we visualize the latent space's layout.

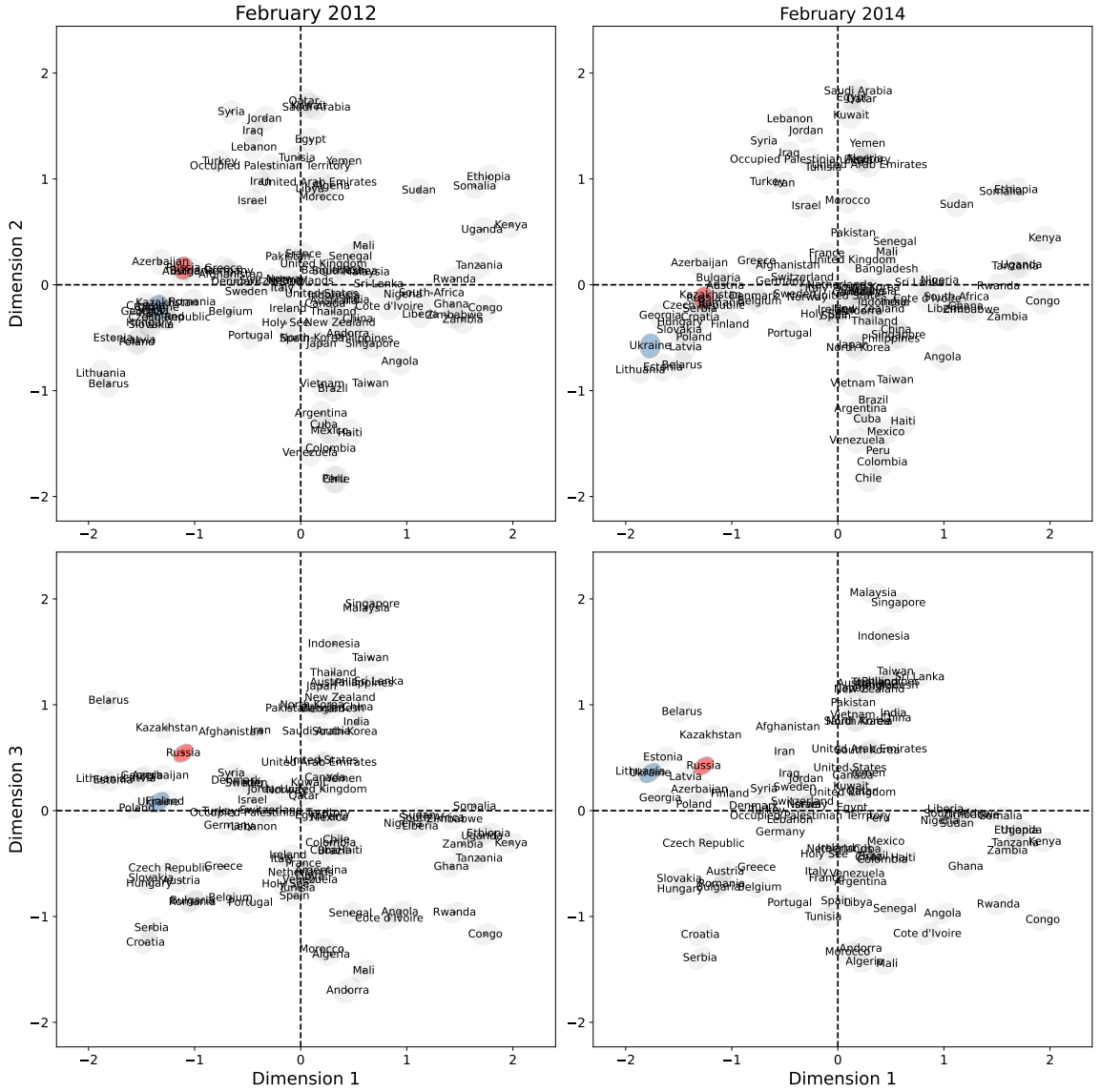


Figure 8: Estimated latent space for the ICEWS networks on February 2012 (left) and February 2014 (right). The first row plots the second dimension (vertical) against the first dimension (horizontal), and the second row plots the third dimension (vertical) against the first dimension (horizontal). The x and y axes are denoted by the dashed horizontal and vertical lines, respectively. The names of each nation are annotated. The ellipses are two standard deviation ($\sim 95\%$) credible ellipses for each actor's latent position. Ukraine and Russia are highlighted in blue and red, respectively.

Figure 8 displays the estimated latent space during February 2012 and February 2014. The latent space encodes the geographic locations of the countries. Due to the positive homophily of the relations, actors are more likely to connect when their latent positions share a common angle. Figure 8's first row plots the second latent dimension (vertical axis) against the first latent dimension (horizontal axis). In these plots, Middle Eastern nations are in the top left quadrant, Eastern African nations are in the top right quadrant, Latin American nations are in the bottom right quadrant, and Eastern European countries are in the bottom left quadrant. Furthermore, the second latent dimension mainly separates Middle Eastern and African nations from nations in Eastern Europe and Latin America. Figure 8's second row plots the third latent dimension against the first latent dimension. In these plots, the third dimension (the vertical axis) is used to separate Eastern European nations into northern countries near Russia's border and Southeast European countries, and to separate Asian nations from African nations. Furthermore, highly sociable nations, such as the United States, are near the center of the latent space in both plots because their high sociality explains most of their interactions. Overall, we conclude that the higher values of the conflict homophily coefficients indicate that the regional (geographic) effects encoded by the second and third latent dimensions play a more prominent role in predicting conflict than cooperation.

Finally, we demonstrate how the latent space reflects the regional Crimea Crisis between Russia and Ukraine in early 2014. Figure 9 displays the latent trajectories for the two nations by plotting the magnitudes of their latent positions as well as the angle between them. Recall that the relation's positive homophily means that actors are more likely to connect when their latent positions have a large magnitude and a small angle between them. Unlike the actor's social trajectories, their latent trajectories are highly variable during the

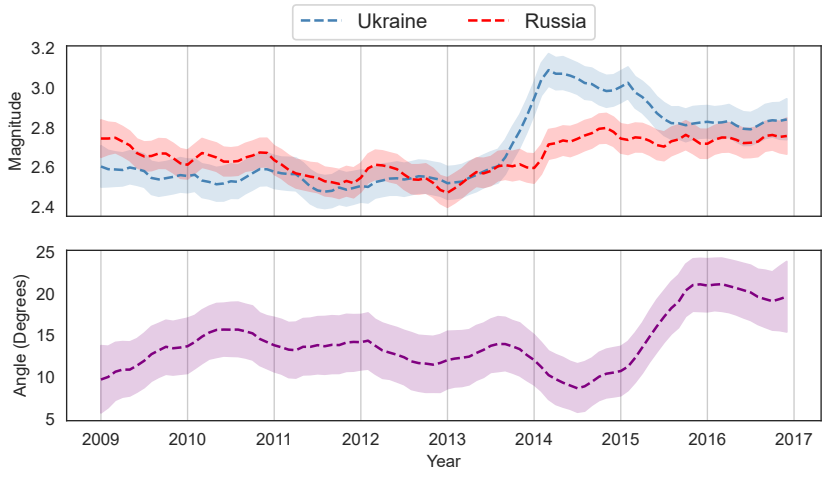


Figure 9: Posterior means and 95% credible intervals of Ukraine and Russia's latent trajectories. The top and bottom plots give estimates for each position's magnitude and the angle between Russia and Ukraine's positions, respectively. Estimates are calculated using 5,000 samples from the approximate posterior.

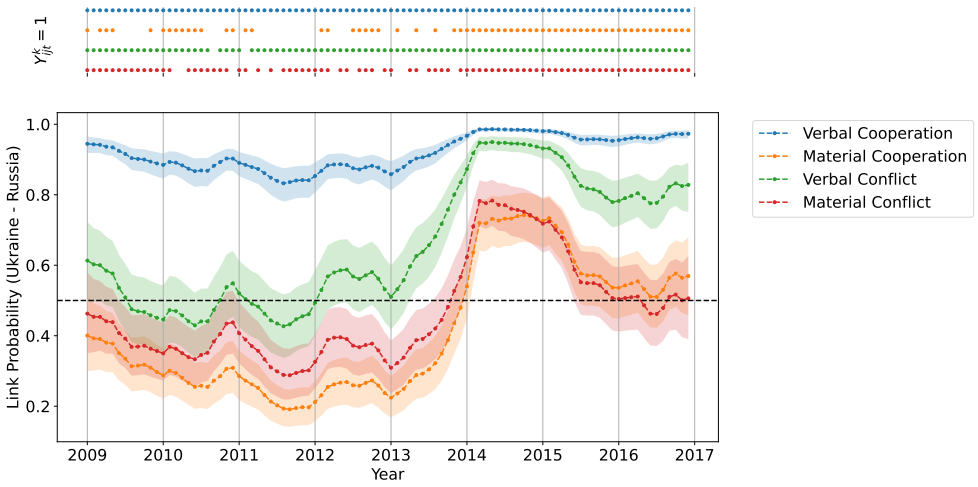


Figure 10: Observed values of the adjacency matrices where a dot indicates that Ukraine and Russia had a particular relation during that month (top). The posterior means and 95% credible intervals for the monthly link probability between the two nations across the four international relations (bottom). Estimates are calculated using 5,000 samples from the approximate posterior. The horizontal dashed black line indicates a probability of 0.5.

Crimea Crisis. Around the second half of 2013, the magnitude of Ukraine's latent position increases significantly, reaching a maximum in early 2014. During this time, the magnitude of Russia's latent position also increases slightly. In addition, the angle between the two countries decreases during that time. Comparing Ukraine and Russia's latent positions in February 2012 to those in February 2014 in Figure 8, we see that they align themselves while moving toward the periphery of the latent space. Furthermore, these dynamics mostly result from changes in the second and third dimensions of their latent positions. These dynamics result in an increased connection probability between the two nations in all layers during the crisis, see Figure 10. Overall, we conclude that the latent trajectories reflect regional events in the ICEWS data.

5.2 Epidemiological Face-to-Face Contact Networks

This case study uses our proposed model to analyze longitudinal face-to-face contact networks drawn from an epidemiological survey of students at a primary school (grades 1 to 5) in Lyon, France. Such contact networks influence mathematical models of infectious disease spread in varying populations (Wallinga et al., 2006; Zagheni et al., 2008). Also, the analysis of these contact patterns allows school administrators to mitigate infectious disease spread in classrooms by determining the times during the day when spread is most prevalent. In the exploratory phase, these analyses often have difficulty visualizing the complicated dynamic networks. Furthermore, they often do not formally quantify the un-

certainty in network statistics. In this section, we demonstrate how our model provides a meaningful network visualization and quantification of uncertainty.

The contact networks were collected by the SocioPatterns collaboration (`http://www.sociopatterns.org`) and initially analyzed in Stehlé et al. (2011). Contact data is available for 242 individuals (232 children and 10 teachers) belonging to grades 1 through 5. Each grade is split into two sections (A and B) so that there are ten classes overall. Each class has its own classroom and teacher. The school day runs from 8:30 am to 4:30 pm, with a lunch break from 12:00 pm to 2:00 pm and two breaks of 20 to 25 minutes around 10:30 am and 3:30 pm.

The face-to-face contacts occurred over two days: Thursday, October 1st, 2009, and Friday, October 2nd, 2009. Data was collected from 8:45 am to 5:20 pm on the first day and from 8:30 am to 5:05 pm on the second day. Radio-frequency identification (RFID) devices measured the contacts between individuals. The RFID sensor registered a contact when two individuals were within 1 to 1.5 meters during a 20-second interval. This distance range was chosen to correspond to the range over which a communicable infectious disease could spread. For more details on the data collection technology, see Cattuto et al. (2010).

5.2.1 Statistical Network Analysis of the School Contact Network

We structured the face-to-face contact data as a dynamic multilayer network recording face-to-face interactions each day. We treated each day as a layer so that the layers correspond to Thursday and Friday. We set Friday as the reference layer because the associated adjacency matrices have the largest maximum rank across all time points. We divided the daily contact networks into 20-minute time intervals between 9:00 am and 5:00 pm and extended the first and last time intervals to accommodate the different starting and ending times of the experiment on the two days. We chose this bin size to match Stehlé et al. (2011), who found that a 20-minute time window appropriately filtered out the noisy fluctuations of the dynamic contact networks and adequately retained the slowly-varying information about the contact network's evolution. This preprocessing resulted in a dynamic multilayer network with $K = 2$ layers, $T = 24$ time steps, and $n = 242$ actors. Specifically, an edge ($Y^k_{ijt} = 1$) means that actor i and actor j had at least one registered interaction during the tth 20-minute interval on day k.

We fit six models with dimensions ranging from 1 to 6 using the SMF algorithm and determined that a latent space dimension of $d = 2$ fit the data best based on the AIC, see Figure 19 in Appendix J. The chosen model's in-sample AUC was 0.97, which indicates a good fit to the data. Since the model's latent positions are only identifiable up to a signed-permutation, the latent positions at time t (for $t \geq 2$) are matched to the positions closest to their previous positions at time $t - 1$ through a signed-permutation.

5.2.2 Dynamics of the Epidemic Branching Factor

Here, we demonstrate how to use our model to (1) determine periods in the school day most susceptible to the spread of infectious disease and (2) identify differences in the contact patterns between the two days. To quantify a network's contribution to the spread of

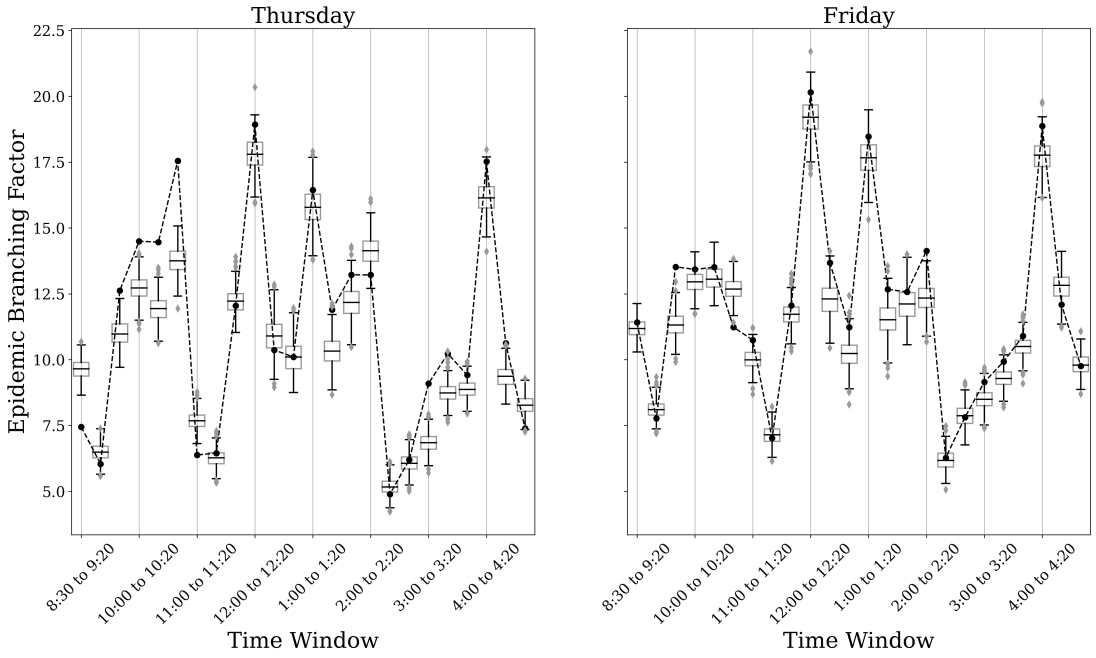


Figure 11: Epidemic branching factors for the face-to-face contact networks on Thursday (left) and Friday (right) at different times throughout the school day. The dashed black curves depict the observed network's branching factor. Boxplots show the range of the branch factor's posterior distribution.

infectious disease, we use the epidemic branching factor (Andersson, 1998), defined as

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

where d_i is the ith node's degree. The epidemic branching factor is related to the basic reproduction number, R_0, which is (loosely) equal to the number of secondary infections caused by a typical infectious individual during an epidemic's early stages (Anderson and May, 1991). In network-based susceptible-infected-recovered (SIR) models, R_0 equals $\tau(\kappa - 1)/(\tau + \gamma)$, where τ and γ are infection and recovery rates, respectively (Andersson, 1997). This relation implies that larger branching factors lead to more massive epidemics.

Figure 11 depicts the posterior distribution of the epidemic branching factor. The boxplots contain 500 networks, each sampled from a different set of latent variables drawn from the model's approximate posterior. The model matches the observed branching factor for most time steps, which include some of the most dramatic changes at 12:00 pm to 12:20 pm, 1:00 pm to 1:20 pm, and 4:00 pm to 4:20 pm. However, the model underestimates the change at 10:40 am to 11:00 am on Thursday. Regardless, the model still captures these four spikes in branching factor. Intuitively, the timings of these spikes occur during lunchtime (12:00 pm to 2:00 pm) and the two short breaks (around 10:30 am and 3:30 pm). More surprisingly, the branching factor's dynamics differ between the two days. The most

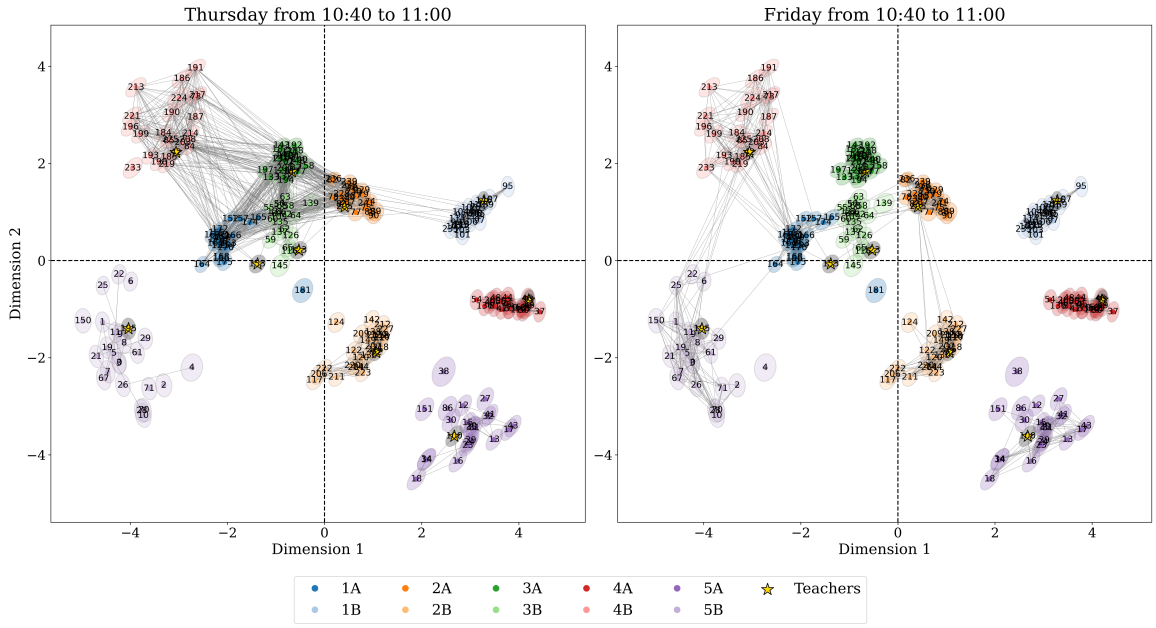


Figure 12: Estimated latent space for the primary school face-to-face contact networks from 10:40 am to 11:00 am. The gray lines indicate observed edges on Thursday (left) and Friday (right). The x and y axes are denoted by the dashed horizontal and vertical lines, respectively. The students are colored by their classroom and section, while teachers are displayed with a yellow star. The ellipses are two standard deviation ($\sim 95\%$) credible ellipses for each actor's latent position.

apparent difference is the spike from 10:40 am to 11:00 am on Thursday that is not present on Friday. To understand what caused this difference, we analyzed the shared latent space. We defer a discussion of the actor's social trajectories to Appendix J.

Figure 12 depicts the latent positions' posterior means and the observed edges on Thursday and Friday during the first short break from 10:40 am to 11:00 am. The inferred homophily coefficients are all positive and significantly different between layers (see Figure 22 in Appendix J). The latent space accurately clusters the students into their ten classrooms. The two layers share the same classroom structure, which affirms our choice of a shared latent space. The difference in branching factors is due to the varying mixing patterns between the classrooms on the two days. Specifically, the classrooms that interact on the two days are different. On Thursday, there are many contacts between students in classes 1A, 1B, 2A, 3A, 3B, and 4B. In contrast, on Friday, classes 1A, 2A, 2B, 4B, and 5B interact. Furthermore, the number of edges between classrooms is much lower on Friday than on Thursday. This observation implies a simple intervention to mitigate disease spread: stagger each classroom's break time in order to limit contacts between students of different classes, which will lower the epidemic branching factor.

6. Discussion

This article proposed a flexible, interpretable, and computationally efficient latent space model for dynamic multilayer networks. Our eigenmodel for dynamic multilayer networks decomposes the dyadic data into a common time-varying latent space used differently by the layers through layer-specific homophily levels and additive node-specific social trajectories that account for further degree heterogeneity. Also, we determined and corrected for various identifiability issues. This accomplishment allows for an intuitive interpretation of the latent space, unlike previous nonparametric models (Durante et al., 2017). Next, we developed an efficient variational inference algorithm for parameter estimation. Unlike previous variational approaches, we maintain the essential temporal dependencies in the posterior approximation. Furthermore, our variational algorithm is widely applicable to general dynamic bilinear latent space models. A simulation study established the effectiveness of our estimation procedure to scale to various network sizes. Finally, we demonstrated how to use our model to analyze international relations from 2009 to 2016 and understand the spread of an infectious disease in a primary school contact network.

Many real-world networks contain non-binary relations. One can adopt the proposed model to networks with non-binary edges with minor changes. For example, replacing the Bernoulli likelihood in Equation (1) with a Gaussian likelihood can model real-valued networks with minimal changes to the variational algorithm. However, extending the variational algorithm to general exponential family likelihoods, such as Poisson or negative binomial, is a direction for future research.

Relations are often directed in nature; therefore, it is natural to generalize the model to directed networks. Such a model needs to allow for varying levels of reciprocity in the directed relations. A simple extension of our model to directed networks is

$$Y_{ijt}^k \overset{\text{ind.}}{\sim} \text{Bernoulli}\left(\text{logit}^{-1}\left[\Theta_{ijt}^k\right]\right), \quad \text{with} \quad \Theta_{ijt}^k = \delta_{k,t}^i + \gamma_{k,t}^j + \mathbf{X}_t^{i\,\mathrm{T}} \Lambda_k \mathbf{Z}_t^j,$$

where $Y_{ijt}^k = 1$ ($Y_{ijt}^k = 0$) denotes the presence (absence) of a directed edge from i to j in layer k at time t. The latent variables' distributions are

$$\gamma_{k,1}^i \overset{\text{iid}}{\sim} N(0, \tau_\gamma^2), \quad \gamma_{k,t}^i \sim N(\gamma_{k,t-1}^i, \sigma_\gamma^2), \quad \mathbf{Z}_1^i \overset{\text{iid}}{\sim} N(\mathbf{0}_d, \Psi_{0,z}), \quad \mathbf{Z}_t^i \sim N(\mathbf{Z}_{t-1}^i, \sigma_z^2 I_d),$$

and the priors on the remaining parameters are left unchanged from the undirected case. In this case, $\delta_{k,t}^i, \gamma_{k,t}^i \in \mathbb{R}$ model degree heterogeneity in outgoing and incoming edges, respectively. The asymmetric latent positions $\mathbf{X}_t^i, \mathbf{Z}_t^i \in \mathbb{R}^d$ allow an actor's features to differ depending on whether they are receiving or initiating the relation. The variational inference algorithm for this model remains mostly unchanged. However, a model that does not drastically increase the number of parameters compared to the undirected case, such as the one in Sewell and Chen (2015), is an area of research interest.

Further research directions include increasing the algorithm's scalability through stochastic variational inference (Hoffman et al., 2013; Aliverti and Russo, 2022) and exploring the variational estimates' statistical properties. Overall, our proposed eigenmodel for dynamic multilayer networks is an interpretable statistical network model with applications to various real-world scientific problems. A repository for the replication code is available on Github (Loyal, 2023).

Acknowledgments

This work was supported in part by National Science Foundation grant DMS-2015561 and a grant from Sandia National Laboratories. We thank the action editor and referees for their helpful comments and suggestions, which significantly improved this paper.

Appendix A. Parameter Identifiability Without Assumption A4

In this section, we analyze the identifiability of the proposed eigenmodel for dynamic multilayer networks when Assumption A4 does not hold. First, Proposition 2 shows that Assumptions A1—A3 are sufficient to identify the model up to a restricted linear transformation of the latent space that varies by time.

Proposition 2 *Suppose that two sets of parameters* $\{\boldsymbol{\delta}_{1:K,1:T}, \mathcal{X}_{1:T}, \Lambda_{1:K}\}$ *and* $\{\tilde{\boldsymbol{\delta}}_{1:K,1:T}, \tilde{\mathcal{X}}_{1:T}, \tilde{\Lambda}_{1:K}\}$ *satisfy Assumptions A1—A3 from Theorem 1, then the model is identifiable up to a linear transformation of the latent space at each time point. That is, if for all* $1 \leq k \leq K$ *and* $1 \leq t \leq T$ *we have that*

$$\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\Lambda_k\mathcal{X}_t^{\mathrm{T}} = \tilde{\boldsymbol{\delta}}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\tilde{\boldsymbol{\delta}}_{k,t}^{\mathrm{T}} + \tilde{\mathcal{X}}_t\tilde{\Lambda}_k\tilde{\mathcal{X}}_t^{\mathrm{T}},$$

then for all $1 \leq k \leq K$ *and* $1 \leq t \leq T$ *we have that*

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

where each $M_t \in \mathbb{R}^{d\times d}$ *satisfies* $M_t I_{p',q'} M_t^{\mathrm{T}} = I_{p,q}$ *for* $1 \leq t \leq T$.

Under Proposition 2, the latent space is identifiable up to a restricted set of linear transformations, M_t, that are difficult to interpret. For this reason, we provide the following proposition, which reduces the set of linear transformations to a well-studied group of transformations when the latent space dimension $d \leq 3$, which is often the case in practice.

Proposition 3 *Consider the same setup as in Proposition 2. If* $1 \leq d \leq 3$, *then* $I_{p,q} = I_{p',q'}$ *so that each* M_t *is in the indefinite orthogonal group, i.e.,* $M_t I_{p,q} M_t^{\mathrm{T}} = I_{p,q}$ *for* $1 \leq t \leq T$.

Invariance under the indefinite orthogonal group is common in LSMs that allow a disassortative latent space (Rubin-Delanchy et al., 2022). Furthermore, this group reduces to the orthogonal group, another source of non-identifiability in many LSMs, when p or q equals d. The most notable property of the indefinite orthogonal group is that it does not preserve Euclidean distances. This implies that any inference based on Euclidean distances in the latent space is not well-defined. For a detailed discussion, we refer the reader to Rubin-Delanchy et al. (2022), who studied inference under such a non-identifiability in the context of a generalized random dot product graph (RDPG) model. In particular, they showed that any post hoc clustering of the latent positions should use a Gaussian mixture model with elliptical covariance matrices since the clustering results are invariant to indefinite orthogonal transformations. This observation is essential if one intends to use the latent positions for community detection without Assumption A4.

Appendix B. Proofs of Proposition 2, Proposition 3, and Theorem 1

This section demonstrates the identifiability of our model under the assumptions proposed in Proposition 2, Proposition 3, and Theorem 1. Before stating the proofs, we need the following lemma.

Lemma 4 *For any* $\mathbf{v} = (v_1, \ldots, v_n)^{\mathrm{T}} \in \mathbb{R}^n$, *if* $\mathbf{v}\mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n + \mathbf{1}_n\mathbf{v}^{\mathrm{T}}\mathbf{1}_n = 0$, *then* $\mathbf{v} = 0$.

Proof The condition can be written as

$$n\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^n v_i \\ \vdots \\ \sum_{i=1}^n v_i \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

which implies $v_1 = \cdots = v_n = -(1/n)\sum_{i=1}^n v_i$. Thus, we have $\mathbf{v} = 0$. ■

Proof [Proof of Proposition 2]. We begin by showing that under Assumption A1, the social trajectories $\boldsymbol{\delta}_{k,t}$ are identifiable. Under Assumption A1, $J_n\mathcal{X}_t = \mathcal{X}_t$ and $J_n\tilde{\mathcal{X}}_t = \tilde{\mathcal{X}}_t$, which implies that $\mathcal{X}_t\Lambda_k\mathcal{X}_t^{\mathrm{T}}\mathbf{1}_n = \tilde{\mathcal{X}}_t\Lambda_k\tilde{\mathcal{X}}_t^{\mathrm{T}}\mathbf{1}_n = 0$. Now assume two sets of parameters satisfy

$$\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}} + \mathcal{X}_t\Lambda_k\mathcal{X}_t^{\mathrm{T}} = \tilde{\boldsymbol{\delta}}_{k,t}\mathbf{1}_n^{\mathrm{T}} + \mathbf{1}_n\tilde{\boldsymbol{\delta}}_{k,t}^{\mathrm{T}} + \tilde{\mathcal{X}}_t\tilde{\Lambda}_k\tilde{\mathcal{X}}_t^{\mathrm{T}} \tag{8}$$

for $k = 1, \ldots, K$ and $t = 1, \ldots, T$. Right multiplying $\mathbf{1}_n$ on both sides of the above equation gives

$$\boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n + \mathbf{1}_n\boldsymbol{\delta}_{k,t}^{\mathrm{T}}\mathbf{1}_n = \tilde{\boldsymbol{\delta}}_{k,t}\mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n + \mathbf{1}_n\tilde{\boldsymbol{\delta}}_{k,t}^{\mathrm{T}}\mathbf{1}_n,$$

or

$$(\boldsymbol{\delta}_{k,t} - \tilde{\boldsymbol{\delta}}_{k,t})\mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n + \mathbf{1}_n(\boldsymbol{\delta}_{k,t} - \tilde{\boldsymbol{\delta}}_{k,t})^{\mathrm{T}}\mathbf{1}_n = 0.$$

Applying Lemma 4, we conclude that

$$\boldsymbol{\delta}_{k,t} = \tilde{\boldsymbol{\delta}}_{k,t} \tag{9}$$

for all k and t.

Now, we focus on the identifiability of the latent space and the homophily coefficients. By Assumption A3, for the reference layer r, Equation (8) and Equation (9) imply

$$\mathcal{X}_t I_{p,q}\mathcal{X}_t^{\mathrm{T}} = \tilde{\mathcal{X}}_t I_{p',q'}\tilde{\mathcal{X}}_t^{\mathrm{T}}. \tag{10}$$

By Assumption A2, $\mathcal{X}_t$ and $\tilde{\mathcal{X}}_t$ are full rank so have left inverses B and $\tilde{B}$, respectively. In other words, $B\mathcal{X}_t = \tilde{B}\tilde{\mathcal{X}}_t = I_d$. Multiplying Equation (10) on the right by $\tilde{B}^{\mathrm{T}}I_{p',q'}$, we have that

$$\tilde{\mathcal{X}}_t = \mathcal{X}_t I_{p,q}\mathcal{X}_t^{\mathrm{T}}\tilde{B}^{\mathrm{T}}I_{p',q'} = \mathcal{X}_t M_t, \tag{11}$$

where $M_t = I_{p,q}\mathcal{X}_t^{\mathrm{T}}\tilde{B}^T I_{p',q'} \in \mathbb{R}^{d\times d}$. More generally, for all layers $k \in \{1, \ldots K\}$, we have

$$\mathcal{X}_t\Lambda_k\mathcal{X}_t^{\mathrm{T}} = \tilde{\mathcal{X}}_t\tilde{\Lambda}_k\tilde{\mathcal{X}}_t^{\mathrm{T}} = \mathcal{X}_t M_t\tilde{\Lambda}_k M_t^{\mathrm{T}}\mathcal{X}_t^{\mathrm{T}},$$

where the last equality used the identity in Equation (11). Multiplying each side of the previous identity on the left by B and on the right by B^{T}, we conclude that

$$\Lambda_k = M_t \tilde{\Lambda}_k M_t^{\mathrm{T}}. \tag{12}$$

However, we want a transformation that takes Λ_k to $\tilde{\Lambda}_k$. To proceed, we note that M_t is invertable. Indeed, since Equation (12) holds for the reference layer, we conclude that $M_t I_{p',q'} M_t^{\mathrm{T}} = I_{p,q}$. It is then easy to check that $M_t^{-1} = I_{p',q'} M_t^{\mathrm{T}} I_{p,q}$. Therefore, $(M_t^{\mathrm{T}})^{-1} = (M_t^{-1})^{\mathrm{T}} = I_{p,q} M_t I_{p',q'}$. Multiplying Equation (12) on the left by M_t^{-1} and on the right by $(M_t^{\mathrm{T}})^{-1}$, we find that

$$\tilde{\Lambda}_k = \left[I_{p',q'} M_t^{\mathrm{T}} I_{p,q}\right] \Lambda_k \left[I_{p,q} M_t I_{p',q'}\right].$$

Lastly, multiplying on the left and the right by $I_{p',q'}$ and noting that $I_{p',q'} \tilde{\Lambda}_k I_{p',q'} = \tilde{\Lambda}_k$ and $I_{p,q} \Lambda_k I_{p,q} = \Lambda_k$, we find that

$$\tilde{\Lambda}_k = M_t^{\mathrm{T}} \Lambda_k M_t,$$

which completes the proof. ■

Proof [Proof of Proposition 3]. From Proposition 2, we have that each matrix M_t satisfies $M_t I_{p',q'} M_t^{\mathrm{T}} = I_{p,q}$ for $1 \leq t \leq T$. Consider a single matrix $M \in \{M_t\}_{t=1}^T$. Taking the determinant of both sides of $M I_{p',q'} M^{\mathrm{T}} = I_{p,q}$, we conclude that $\det(M)^2 (-1)^{q'} = (-1)^q$, so that $q - q'$ is an even number. Without loss of generality, assume that $q \geq q'$. We proceed case by case:

(i) $d = 1$. Since $q - q'$ can only equal zero, the result is immediate.

(ii) $d = 2$. In this case, the only non-trivial case is $q - q' = 2$, which corresponds to $q = 2, q' = 0$. Now, we show that this combination leads to a contradiction. In this case, M satisfies $M M^{\mathrm{T}} = -I_2$, which is a contradiction because $M M^{\mathrm{T}}$ is a positive-definite matrix while $-I_2$ is not. Thus, $q = q'$.

(iii) $d = 3$. Once again, the only non-trivial case is $q - q' = 2$, where we are have the following two cases: $q = 3, q' = 1$ and $q = 2, q' = 0$. We proceed by showing that both scenarios lead to a contradiction.

For the case $q = 3, q' = 1$, we have $M I_{2,1} M^{\mathrm{T}} = -I_3$ which implies $-M^{\mathrm{T}} M = I_{2,1}$, where we used the fact that $M^{\mathrm{T}} I_{p,q} M = I_{p',q'}$ shown during the proof of Proposition 2. Letting $\mathbf{m}_j$ be the jth column of M, we have that

$$-\|\mathbf{m}_1\|_2^2 = 1,$$

which cannot be satisfied by a real vector.

Similarly for $q = 2, q' = 0$, we have $M M^{\mathrm{T}} = I_{1,2}$. Letting $\tilde{\mathbf{m}}_j$ be the jth row vector of M, we have that

$$\|\tilde{\mathbf{m}}_2\|_2^2 = -1.$$

which is impossible for a real vector.

Therefore, $I_{p,q} = I_{p',q'}$ when $1 \leq d \leq 3$, which completes the proof. ■

Proof [Proof of Theorem 1]. Without loss of generality, consider a single matrix $M \in \{M_t\}_{t=1}^T$. Further let $\Lambda_\ell = \operatorname{diag}(\boldsymbol{\lambda}_\ell)$ and $\tilde{\Lambda}_\ell = \operatorname{diag}(\tilde{\boldsymbol{\lambda}}_\ell)$, so that by Proposition 2 we have that

$$\operatorname{diag}(\tilde{\boldsymbol{\lambda}}_\ell) = M^{\mathrm{T}} \operatorname{diag}(\boldsymbol{\lambda}_\ell) M. \tag{13}$$

Now, left multiplying $M I_{p',q'}$ on both sides of Equation (13) and applying the identity $M I_{p',q'} M^{\mathrm{T}} = I_{p,q}$, we have that

$$M \operatorname{diag}(\tilde{\boldsymbol{\lambda}}_\ell) I_{p',q'} = \operatorname{diag}(\boldsymbol{\lambda}_\ell) I_{p,q} M, \tag{14}$$

Denoting the jth columns of M by $\mathbf{m}_j \in \mathbb{R}^d$, we can re-express the linear system in Equation (14) as

$$\left(\tilde{\lambda}_{\ell,j} (I_{p',q'})_{jj}\, I_d - \operatorname{diag}(\boldsymbol{\lambda}_\ell) I_{p,q}\right) \mathbf{m}_j = \mathbf{0}_d \quad \text{for } j = 1, \ldots, d, \tag{15}$$

where $\mathbf{0}_d$ is a d-dimensional vector of zeros.

Now, we determine what relationship Equation (15) imposes on $\operatorname{diag}(\boldsymbol{\lambda}_\ell)$ and $\operatorname{diag}(\tilde{\boldsymbol{\lambda}}_\ell)$. Let $A_j = \tilde{\lambda}_{j,\ell} (I_{p',q'})_{jj}\, I_d - \operatorname{diag}(\boldsymbol{\lambda}_\ell) I_{p,q}$ for $j = 1, \ldots, d$. Since M is full rank, A_j must be a singular matrix for $j = 1, \ldots, d$. Combining the facts that $\operatorname{diag}(\tilde{\boldsymbol{\lambda}}_\ell) I_{p',q'}$ and $\operatorname{diag}(\boldsymbol{\lambda}_\ell) I_{p,q}$ are both full rank with d distinct elements and that there are d singular diagonal matrices A_j, it is easy to see that $\operatorname{diag}(\boldsymbol{\lambda}_\ell) I_{p,q} = P \operatorname{diag}(\tilde{\boldsymbol{\lambda}}_\ell) I_{p',q'} P^{\mathrm{T}}$ for some permutation matrix P. In other words, $\operatorname{diag}(\boldsymbol{\lambda}_\ell) I_{p,q}$ equals $\operatorname{diag}(\tilde{\boldsymbol{\lambda}}_\ell) I_{p',q'}$ with permuted diagonal entries.

Now we focus on the consequences for M. As a result of the argument in the previous paragraph, each A_j is a rank $d-1$ diagonal matrix. This means that each A_j is a diagonal matrix with $d-1$ non-zero entries and a single zero entry on the diagonal. Therefore, Equation (15) holds if and only if $\{\mathbf{m}_j\}_{j=1}^d$ are d-dimensional vectors with a single non-zero entry where A_j is zero on the diagonal. Also, since M is full rank, $\{\mathbf{m}_1, \ldots, \mathbf{m}_d\}$ are linearly independent. This implies that M is a generalized permutation matrix: $M = P \operatorname{diag}(\mathbf{s})$ where $\operatorname{diag}(\mathbf{s})$ is a full-rank diagonal matrix and P is a permutation matrix.

To complete the proof, we focus on the diagonal entries of $M I_{p',q'} M^{\mathrm{T}} = I_{p,q}$. From the previous paragraph, we have that $M I_{p',q'} M^{\mathrm{T}} = P \operatorname{diag}(\mathbf{s}) I_{p',q'} \operatorname{diag}(\mathbf{s}) P^{\mathrm{T}}$. Let $\sigma : \{1, \ldots, d\} \to \{1, \ldots, d\}$, denote the permutation encoded by P, so that

$$P = \begin{pmatrix} \mathbf{e}_{\sigma(1)}^{\mathrm{T}} \\ \vdots \\ \mathbf{e}_{\sigma(d)}^{\mathrm{T}} \end{pmatrix},$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_d$ are the standard basis vectors. Thus, the diagonal entries must satisfy

$$(I_{p',q'})_{\sigma(j)\sigma(j)}\, s_{\sigma(j)}^2 = (I_{p,q})_{jj} \text{ for } j = 1, \ldots, d,$$

which holds if and only if $\mathbf{s} = \{\pm 1\}^d$, $I_{p,q} = I_{p',q'}$, and σ only permutes the first p and last q diagonal elements of $I_{p,q}$. In other words, for $1 \leq t \leq T$, $M_t = \text{BlockDiagonal}(P_{1,t}, P_{2,t}) \operatorname{diag}(\mathbf{s}_t)$, where $P_{1,t}$ and $P_{2,t}$ are permutation matrices on p and q elements, respectively.

It remains to show that $P_t = P_s$, where $P_t = \text{BlockDiagonal}(P_{1,t}, P_{2,t})$, for all $1 \leq s \neq t \leq T$. For any s, t, we have from Proposition 2 that

$$M_s^{\mathrm{T}} \Lambda_\ell M_s = M_t^{\mathrm{T}} \Lambda_\ell M_t,$$

which simplifies to

$$P_t P_s^{\mathrm{T}} \Lambda_\ell P_s P_t^{\mathrm{T}} = \Lambda_\ell,$$

using the fact that $P^{-1} = P^{\mathrm{T}}$ for any permutation matrix P. Let $\Pi_{s,t} = P_s P_t^{\mathrm{T}} = \text{BlockDiagonal}(P_{1,s} P_{1,t}^{\mathrm{T}}, P_{2,s} P_{2,t}^{\mathrm{T}})$, which is also a permutation matrix. Since $\Lambda_\ell I_{p,q}$ has distinct diagonal elements, we also have that the first p elements of Λ_ℓ must be distinct from one another and the last q elements must be distinct from one another. Using the same argument as in the previous paragraph, it is follows that $\Pi_{s,t}$ is the identity permutation. Therefore, $P_s P_t^{\mathrm{T}} = I_d$ or $P_s = P_t$ for all $1 \leq s \neq t \leq T$, which completes the proof.

■

Appendix C. Derivation of Variational Updates

This section contains detailed derivations of the variational updates presented in Section 3 of the main text. For notational simplicity, we use the shorthand $\mathbb{E}_{q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega})}[\cdot] = \langle \cdot \rangle$, where $q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega})$ is defined in Equation (3) of the main text, to denote expectations with respect to the full variational posterior throughout this section. For a definition of the notation used in this section, see Algorithm 1.

Throughout this section, we encounter the following Gaussian state space model

$$\mathbf{x}_1 \sim N(\mathbf{0}_d, \Psi_0), \tag{16}$$

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}_d, \sigma^2 I_d), \tag{17}$$

$$\mathbf{y}_t = A_t \mathbf{x}_t + \mathbf{b}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}_n, C_t), \tag{18}$$

where $\mathbf{x}_t \in \mathbb{R}^d$, $\mathbf{y}_t \in \mathbb{R}^n$, $A_t \in \mathbb{R}^{n \times d}$, $\mathbf{b}_t \in \mathbb{R}^n$, $C_t \in \mathbb{R}^{n \times n}$. In this context, d is not necessarily the dimension of the latent space and n is not necessarily the number of nodes in the network. Specifically, the full conditional distributions of the social and latent trajectories are of this form. Before proceeding, we state a lemma used throughout Appendix C and Appendix D.

Lemma 5 *For the Gaussian state space model specified by Equations (16)—(18), the conditional distribution $p(\mathbf{x}_{1:T} \mid \mathbf{y}_{1:T})$ is in the exponential family with natural parameters*

$$\psi = (-\Psi_0^{-1}/2, -1/2\sigma^2, \Gamma_{1:T}^1, -\Gamma_{1:T}^2/2), \tag{19}$$

where $\Gamma_t^1 = A_t^{\mathrm{T}} C_t^{-1} \mathbf{y}_t - A_t C_t^{-1} \mathbf{b}_t$ and $\Gamma_t^2 = A_t^{\mathrm{T}} C_t^{-1} A_t$ for $1 \leq t \leq T$.

Proof We have

$$\begin{aligned}\log p(\mathbf{x}_{1:T} \mid \mathbf{y}_{1:T}) &\propto -\frac{1}{2}\mathbf{x}_1^{\mathrm{T}}\Psi_0^{-1}\mathbf{x}_1 - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 - \\ &\qquad \frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t - A_t\mathbf{x}_t - \mathbf{b}_t)^{\mathrm{T}}C_t^{-1}(\mathbf{y}_t - A_t\mathbf{x}_t - \mathbf{b}_t), \\ &\propto -\frac{1}{2}\operatorname{tr}(\Psi_0^{-1}\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 + \\ &\qquad (A_t^{\mathrm{T}}C_t^{-1}\mathbf{y}_t - A_tC_t^{-1}\mathbf{b}_t)^{\mathrm{T}}\mathbf{x}_t - \frac{1}{2}\sum_{t=1}^{T}\operatorname{tr}(A_t^{\mathrm{T}}C_t^{-1}A_t\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}),\end{aligned}$$

which is in exponential family form with natural parameters given in Equation (19). ■

The variational distributions of the social and latent trajectories—Equation (4) and Equation (5) in the main text—are GSSMs that are in the form assumed by Lemma 5. This observation implies that the expected natural parameters, $\mathbb{E}_{-q(\mathbf{x}_{1:T})}[\psi]$, are sufficient for calculating the variational distribution's moments, i.e., $\mathbb{E}_{q(\mathbf{x}_{1:T})}[\mathbf{x}_t]$, $\mathbb{E}_{q(\mathbf{x}_{1:T})}\left[\mathbf{x}_t\mathbf{x}_t^{\mathrm{T}}\right]$, and $\mathbb{E}_{q(\mathbf{x}_{1:T})}\left[\mathbf{x}_t\mathbf{x}_{t+1}^{\mathrm{T}}\right]$. In Appendix D, we derive a variational Kalman smoother that calculates these moments recursively.

C.1 Derivation of Algorithm 2

The coordinate updates for $q(\omega_{ijt}^k)$ are given in Algorithm 2, which we formally derive in the remainder of this section.

Proposition 6 *Under the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text,* $q(\omega_{ijt}^k) = \mathrm{PG}(1, c_{ijt}^k)$ *where* c_{ijt}^k *is given in Equation (20). Furthermore, the mean of this distribution is given by Equation (21).*

Proof From the exponential tilting property of the Pólya-gamma distribution, we have that

$$\omega_{ijt}^k \mid \cdot \sim \mathrm{PG}(1, \psi_{ijt}^k), \tag{22}$$

where $\psi_{ijt}^k = \delta_{k,t}^i + \delta_{k,t}^j + \mathbf{X}_t^{i\,\mathrm{T}}\Lambda_k\mathbf{X}_t^j$. This distribution is in the exponential family with natural parameter $-(\psi_{ijt}^k)^2/2$. The variational distribution is then a $\mathrm{PG}(1, c_{ijt}^k)$ where $(c_{ijt}^k)^2 = \mathbb{E}_{-q(\omega_{ijt}^k)}\left[(\psi_{ijt}^k)^2\right]$.

Update $q(\omega_{ijt}^k) = \text{PG}(1, c_{ijt}^k)$:

For each $k \in \{1, \ldots, K\}$, $t \in \{1, \ldots T\}$, and $(i, j) \in \{(i, j) : 1 \leq i \leq n, j < i\}$:

$$c_{ijt}^k = (\sigma^2_{\delta^i_{k,t}} + \mu^2_{\delta^i_{k,t}} + \sigma^2_{\delta^j_{k,t}} + \mu^2_{\delta^j_{k,t}} + 2\mu_{\delta^i_{k,t}}\mu_{\delta^j_{k,t}} + 2(\mu_{\delta^i_{k,t}} + \mu_{\delta^j_{k,t}})\boldsymbol{\mu}_t^{i\,\mathrm{T}} \operatorname{diag}(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k})\boldsymbol{\mu}_t^j + \|(\Sigma_{\boldsymbol{\lambda}_k} + \boldsymbol{\mu}_{\boldsymbol{\lambda}_k}\boldsymbol{\mu}_{\boldsymbol{\lambda}_k}^{\mathrm{T}}) \odot (\Sigma_t^i + \boldsymbol{\mu}_t^i\boldsymbol{\mu}_t^{i\,\mathrm{T}}) \odot (\Sigma_t^j + \boldsymbol{\mu}_t^j\boldsymbol{\mu}_t^{j\,\mathrm{T}})\|)^{1/2}, \tag{20}$$

$$\mu_{\omega_{ijt}^k} = \frac{1}{2c_{ijt}^k}\left(\frac{e^{c_{ijt}^k} - 1}{1 + e^{c_{ijt}^k}}\right). \tag{21}$$

Algorithm 2: Coordinate ascent updates for the auxiliary Pólya-gamma variables. Here, $\odot$ is the Hadamard product between two matrices, i.e., $(A \odot B)_{ij} = A_{ij}B_{ij}$, and $\|A\| = \sum_i \sum_j A_{ij}$.

It remains to calculate the natural parameter c_{ijt}^k. We have that

$$\begin{aligned}
(c_{ijt}^k)^2 = \mathbb{E}_{-q(\omega_{ijt}^k)}\left[(\psi_{ijt}^k)^2\right] &= \langle(\delta_{k,t}^i + \delta_{k,t}^j + \mathbf{X}_t^{i\,\mathrm{T}}\Lambda_k\mathbf{X}_t^j)^2\rangle, \\
&= \langle(\delta_{k,t}^i + \delta_{k,t}^j)^2\rangle + 2\langle\delta_{k,t}^i + \delta_{k,t}^j\rangle\langle\mathbf{X}_t^i\rangle^{\mathrm{T}}\langle\Lambda_k\rangle\langle\mathbf{X}_t^j\rangle + \langle(\mathbf{X}_t^{i\,\mathrm{T}}\Lambda_k\mathbf{X}_t^j)^2\rangle, \\
&= \sigma^2_{\delta^i_{k,t}} + \mu^2_{\delta^i_{k,t}} + \sigma^2_{\delta^j_{k,t}} + \mu^2_{\delta^j_{k,t}} + 2\mu_{\delta^i_{k,t}}\mu_{\delta^j_{k,t}} + \\
&\qquad 2(\mu_{\delta^i_{k,t}} + \mu_{\delta^j_{k,t}})\boldsymbol{\mu}_t^{i\,\mathrm{T}}\operatorname{diag}(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k})\boldsymbol{\mu}_t^j + \mathbb{E}_{-q(\omega_{ijt}^k)}\left[(\mathbf{X}_t^{i\,\mathrm{T}}\Lambda_k\mathbf{X}_t^j)^2\right].
\end{aligned}$$

Note that the last term is equal to

$$\begin{aligned}
\mathbb{E}_{-q(\omega_{ijt}^k)}\left[(\mathbf{X}_t^{i\,\mathrm{T}}\Lambda_k\mathbf{X}_t^j)^2\right] &= \mathbb{E}_{-q(\omega_{ijt}^k)}\left[\sum_{g=1}^d\sum_{h=1}^d \lambda_g^k\lambda_h^k X_{tg}^i X_{th}^i X_{tg}^j X_{th}^j\right], \\
&= \sum_{g=1}^d\sum_{h=1}^d \mathbb{E}_{q(\boldsymbol{\lambda}_k)}\left[\lambda_g^k\lambda_h^k\right]\mathbb{E}_{q(\mathbf{X}_t^i)}\left[X_{tg}^i X_{th}^i\right]\mathbb{E}_{q(\mathbf{X}_t^j)}\left[X_{tg}^j X_{th}^j\right], \\
&= \|\mathbb{E}_{q(\boldsymbol{\lambda}_k)}\left[\boldsymbol{\lambda}_k\boldsymbol{\lambda}_k^{\mathrm{T}}\right] \odot \mathbb{E}_{q(\mathbf{X}_t^i)}\left[\mathbf{X}_t^i\mathbf{X}_t^{i\,\mathrm{T}}\right] \odot \mathbb{E}_{q(\mathbf{X}_t^j)}\left[\mathbf{X}_t^j\mathbf{X}_t^{j\,\mathrm{T}}\right]\|, \\
&= \|(\Sigma_{\boldsymbol{\lambda}_k} + \boldsymbol{\mu}_{\boldsymbol{\lambda}_k}\boldsymbol{\mu}_{\boldsymbol{\lambda}_k}^{\mathrm{T}}) \odot (\Sigma_t^i + \boldsymbol{\mu}_t^i\boldsymbol{\mu}_t^{i\,\mathrm{T}}) \odot (\Sigma_t^j + \boldsymbol{\mu}_t^j\boldsymbol{\mu}_t^{j\,\mathrm{T}})\|,
\end{aligned}$$

where $\odot$ is the Hadamard product, i.e, $(A \odot B)_{ij} = A_{ij}B_{ij}$, and $\|A\| = \sum_i \sum_j A_{ij}$.

Lastly, the moments of the Pólya-gamma distribution are available in closed form. In particular, we have that

$$\mu_{\omega_{ijt}^k} = \mathbb{E}_{q(\omega_{ijt}^k)}\left[\omega_{ijt}^k\right] = \frac{1}{2c_{ijt}^k}\left(\frac{e^{c_{ijt}^k} - 1}{1 + e^{c_{ijt}^k}}\right).$$

■

C.2 Derivation of Algorithm 3

The coordinate updates for $q(\delta^i_{1:T}), q(\tau^2_\delta)$, and $q(\sigma^2_\delta)$ are given in Algorithm 3, which we formally derive in the remainder of this section.

Proposition 7 *Under the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text,* $q(\tau^2_\delta) = \Gamma^{-1}(\bar{a}_{\tau^2_\delta}/2, \bar{b}_{\tau^2_\delta}/2)$ *where* $\bar{a}_{\tau^2_\delta}$ *and* $\bar{b}_{\tau^2_\delta}$ *are defined in Equation (27) and Equation (28), respectively.*

Proof Standard calculations show that

$$p(\tau^2_\delta \mid \cdot) \propto \left(\frac{1}{\tau^2_\delta}\right)^{(a_{\tau^2_\delta}+nK)/2} \exp\left(-\frac{1}{2\tau^2_\delta}\sum_{k=1}^{K}\sum_{i=1}^{n}(\delta^i_{k,1})^2 - \frac{b_{\tau^2_\delta}}{2\tau^2_\delta}\right),$$
$$\propto \Gamma^{-1}\left(\frac{a_{\tau^2_\delta}+nK}{2}, \frac{1}{2}\left\{\sum_{k=1}^{K}\sum_{i=1}^{n}(\delta^i_{k,1})^2 + b_{\tau^2_\delta}\right\}\right).$$

Note that $\Gamma^{-1}(a/2, b/2)$ is in the exponential family with natural parameters a and b. Thus, the variational distribution is also an inverse-gamma distribution with natural parameters

$$\bar{a}_{\tau^2_\delta} = a_{\tau^2_\delta} + nK,$$
$$\bar{b}_{\tau^2_\delta} = b_{\tau^2_\delta} + \sum_{k=1}^{K}\sum_{i=1}^{n}\mathbb{E}_{q(\delta^i_{k,1:T})}\left[(\delta^i_{k,1})^2\right],$$
$$= b_{\tau^2_\delta} + \sum_{k=1}^{K}\sum_{i=1}^{n}\left(\sigma^2_{\delta^i_{k,1}} + \mu^2_{\delta^i_{k,1}}\right).$$

■

Proposition 8 *Under the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text,* $q(\sigma^2_\delta) = \Gamma^{-1}(\bar{c}_{\tau^2_\delta}/2, \bar{d}_{\tau^2_\delta}/2)$ *where* $\bar{c}_{\tau^2_\delta}$ *and* $\bar{d}_{\tau^2_\delta}$ *are defined in Equation (29) and Equation (30), respectively.*

Proof Standard calculations show that

$$p(\sigma^2_\delta \mid \cdot) \propto \left(\frac{1}{\sigma^2_\delta}\right)^{(c_{\sigma^2_\delta}+nK(T-1))/2} \exp\left(-\frac{1}{2\sigma^2_\delta}\sum_{k=1}^{K}\sum_{t=2}^{T}\sum_{i=1}^{n}(\delta^i_{k,t} - \delta^i_{k,t-1})^2 - \frac{d_{\sigma^2_\delta}}{2\sigma^2_\delta}\right),$$
$$\propto \Gamma^{-1}\left(\frac{c_{\sigma^2_\delta}+nK(T-1)}{2}, \frac{1}{2}\left\{\sum_{k=1}^{K}\sum_{t=2}^{T}\sum_{i=1}^{n}(\delta^i_{k,t} - \delta^i_{k,t-1})^2 + d_{\sigma^2_\delta}\right\}\right).$$

1. Update $q(\delta^i_{k,1:T})$, a linear Gaussian state space model (GSSM):
   For each $k \in \{1, \dots K\}$ and $i \in \{1, \dots, n\}$:

   (a) For $t \in \{1, \dots T\}$, update the natural parameters of the GSSM:

$$\Gamma^1_t = \sum_{j \neq i} [Y^k_{ijt} - 1/2 - \mu_{\omega^k_{ijt}}(\mu_{\delta^j_{k,t}} + \boldsymbol{\mu}^{i\,\mathrm{T}}_t \operatorname{diag}(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k})\boldsymbol{\mu}^j_t)], \tag{23}$$

$$\Gamma^2_t = \sum_{j \neq i} \mu_{\omega^k_{ijt}}, \tag{24}$$

$$\left\langle 1/\tau^2_\delta \right\rangle = \bar{a}_{\tau^2_\delta}/\bar{b}_{\tau^2_\delta}, \tag{25}$$

$$\left\langle 1/\sigma^2_\delta \right\rangle = \bar{c}_{\sigma^2_\delta}/\bar{d}_{\sigma^2_\delta}. \tag{26}$$

   (b) Update marginal distributions and cross-covariances as in Algorithm 7:

$$\mu_{\delta^i_{k,1:T}}, \sigma^2_{\delta^i_{k,1:T}}, \{\sigma^2_{\delta^i_{k,t,t+1}}\}_{t=1}^{T-1} = \texttt{KalmanSmoother}(\Gamma^1_{1:T}, \Gamma^2_{1:T}, \bar{a}_{\tau^2_\delta}/\bar{b}_{\tau^2_\delta}, \bar{c}_{\sigma^2_\delta}/\bar{d}_{\sigma^2_\delta}).$$

2. Update $q(\tau^2_\delta) = \Gamma^{-1}(\bar{a}_{\tau^2_\delta}/2, \bar{b}_{\tau^2_\delta}/2)$:

$$\bar{a}_{\tau^2_\delta} = a_{\tau^2_\delta} + nK, \tag{27}$$

$$\bar{b}_{\tau^2_\delta} = b_{\tau^2_\delta} + \sum_{k=1}^{K}\sum_{i=1}^{n}\left(\sigma^2_{\delta^i_{k,1}} + \mu^2_{\delta^i_{k,1}}\right). \tag{28}$$

3. Update $q(\sigma^2_\delta) = \Gamma^{-1}(\bar{c}_{\sigma^2_\delta}/2, \bar{d}_{\sigma^2_\delta}/2)$:

$$\bar{c}_{\sigma^2_\delta} = c_{\sigma^2\delta} + nK(T-1), \tag{29}$$

$$\bar{d}_{\sigma^2_\delta} = d_{\sigma^2_\delta} + \sum_{k=1}^{K}\sum_{t=2}^{T}\sum_{i=1}^{n}\left\{\sigma^2_{\delta^i_{k,t}} + \mu^2_{\delta^i_{k,t}} + \sigma^2_{\delta^i_{k,t-1}} + \mu^2_{\delta^i_{k,t-1}} - 2(\sigma^2_{\delta^i_{k,t-1,t}} + \mu_{\delta^i_{k,t-1}}\mu_{\delta^i_{k,t}})\right\}. \tag{30}$$

Algorithm 3: Coordinate ascent updates for the social trajectories. `KalmanSmoother` is the variational Kalman smoother defined in Algorithm 7 of Appendix D.

Note that $\Gamma^{-1}(a/2, b/2)$ is in the exponential family with natural parameters a and b. Thus, the variational distribution is also an inverse-gamma distribution with natural parameters

$$\begin{aligned}
\bar{c}_{\sigma_\delta^2} &= c_{\sigma_\delta^2} + nK(T-1), \\
\bar{d}_{\sigma_\delta^2} &= d_{\sigma_\delta^2} + \sum_{k=1}^{K}\sum_{t=2}^{T}\sum_{i=1}^{n} \mathbb{E}_{q(\delta^i_{k,1:T})}\left[(\delta^i_{k,t} - \delta^i_{k,t-1})^2\right], \\
&= d_{\sigma_\delta^2} + \sum_{k=1}^{K}\sum_{t=2}^{T}\sum_{i=1}^{n} \left\{\sigma^2_{\delta^i_{k,t}} + \mu^2_{\delta^i_{k,t}} + \sigma^2_{\delta^i_{k,t-1}} + \mu^2_{\delta^i_{k,t-1}} - 2(\sigma^2_{\delta_{k,t-1,t}} + \mu_{\delta^i_{k,t}}\mu_{\delta^i_{k,t-1}})\right\}.
\end{aligned}$$

■

Proposition 9 *Under the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text, $q(\delta^i_{k,1:T})$ is a Gaussian state space model with natural parameters given by Equations (23)—(26).*

Proof From Equation (4), we associate $p(\delta^i_{k,1:T} \mid \cdot)$ with a GSSM parameterized by

$$\begin{aligned}
\mathbf{x}_t &= \delta^i_{k,t} \in \mathbb{R}, \\
\mathbf{y}_t &= \mathbf{z}^i_{k,t} \in \mathbb{R}^{n-1}, \\
A_t &= (\omega^k_{i1t}, \ldots, \omega^k_{i(i-1)t}, \omega^k_{i(i+1)t}, \ldots, \omega^k_{int}) \in \mathbb{R}^{n-1}, \\
(\mathbf{b}_t)_j &= \omega^k_{ijt}(\delta^j_{k,t} + \mathbf{X}_t^{j\,\mathrm{T}}\Lambda_k\mathbf{X}^i_t), \quad \text{for } j \neq i, \\
C_t &= \mathrm{diag}(\omega^k_{i1t}, \ldots, \omega^k_{i(i-1)t}, \omega^k_{i(i+1)t}, \ldots, \omega^k_{int}).
\end{aligned} \tag{31}$$

We then apply Lemma 5 in Appendix C to identify the natural parameters. Note that $A_t^{\mathrm{T}} C_t^{-1} = \mathbf{1}_{n-1}$, so that expected natural parameters are

$$\Gamma^1_t = \langle \mathbf{1}^{\mathrm{T}}_{n-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle = \sum_{j\neq i}(Y^k_{ijt} - 1/2 - \mu_{\omega^k_{ijt}}(\mu_{\delta^j_{k,t}} + \boldsymbol{\mu}_t^{j\,\mathrm{T}}\,\mathrm{diag}(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k})\boldsymbol{\mu}^i_t)), \tag{32}$$

$$\Gamma^2_t = \langle \mathbf{1}^{\mathrm{T}}_{n-1} A_t\rangle = \sum_{j\neq i}\mu_{\omega^k_{ijt}}. \tag{33}$$

Furthermore, from Proposition 7 and Proposition 8, we have that $1/\tau_\delta^2$ and $1/\sigma_\delta^2$ are gamma distributed with means $\bar{a}_{\tau_\delta^2}/\bar{b}_{\tau_\delta^2}$ and $\bar{c}_{\sigma_\delta^2}/\bar{d}_{\sigma_\delta^2}$, respectively. Finally, we can apply the variational Kalman smoothing equations derived in Appendix D to calculate the moments of $q(\delta^i_{k,1:T})$. ■

C.3 Derivation of Algorithm 4

The coordinate updates for $q(\mathbf{X}^i_{1:T}), q(\Psi_0)$, and $q(\sigma^2)$ are given in Algorithm 4, which we formally derive in the remainder of this section.

1. Update $q(\mathbf{X}^i_{1:T})$, a linear Gaussian state space model (GSSM):
   For $i \in \{1, \ldots, n\}$:

   (a) For $t \in \{1, \ldots, T\}$, update the natural parameters of the GSSM:

$$\Gamma^1_t = \begin{pmatrix} \sum_{k=1}^K \sum_{j \neq i} \mu_{\lambda_{k1}} \mu^j_{t1} [Y^k_{ijt} - 1/2 - \mu_{\omega^k_{ijt}} (\mu_{\delta^i_{k,t}} + \mu_{\delta^j_{k,t}})] \\ \vdots \\ \sum_{k=1}^K \sum_{j \neq i} \mu_{\lambda_{kd}} \mu^j_{td} [Y^k_{ijt} - 1/2 - \mu_{\omega^k_{ijt}} (\mu_{\delta^i_{k,t}} + \mu_{\delta^j_{k,t}})] \end{pmatrix}, \tag{34}$$

$$\Gamma^2_t = \sum_{k=1}^K \sum_{j \neq i} \mu_{\omega^k_{ijt}} \left(\Sigma_{\boldsymbol{\lambda}_k} + \boldsymbol{\mu}_{\boldsymbol{\lambda}_k} \boldsymbol{\mu}^{\mathrm{T}}_{\boldsymbol{\lambda}_k} \right) \odot \left(\Sigma^j_t + \boldsymbol{\mu}^j_t \boldsymbol{\mu}^{j\,\mathrm{T}}_t \right), \tag{35}$$

$$\left\langle \Psi_0^{-1} \right\rangle = \bar{\nu}\, \bar{V}^{-1}, \tag{36}$$

$$\left\langle 1/\sigma^2 \right\rangle = \bar{c}_{\tau^2} / \bar{d}_{\tau^2}. \tag{37}$$

   (b) Update marginal distributions and cross-covariances as in Algorithm 7:

$$\boldsymbol{\mu}^i_{1:T}, \Sigma^i_{1:T}, \{\Sigma^i_{t,t+1}\}_{t=1}^{T-1} = \texttt{KalmanSmoother}(\Gamma^1_{1:T}, \Gamma^2_{1:T}, \bar{\nu}\, \bar{V}^{-1}, \bar{c}_{\sigma^2}/\bar{d}_{\sigma^2}).$$

2. Update $q(\Psi_0) = \text{Wishart}^{-1}(\bar{\nu}, \bar{V})$:

$$\bar{\nu} = \nu + n, \tag{38}$$

$$\bar{V} = V + \sum_{i=1}^n \left(\Sigma^i_1 + \boldsymbol{\mu}^i_1 \boldsymbol{\mu}^{i\,\mathrm{T}}_1 \right). \tag{39}$$

3. Update $q(\sigma^2) = \Gamma^{-1}(\bar{c}_{\sigma^2}/2, \bar{d}_{\sigma^2}/2)$:

$$\bar{c}_{\sigma^2} = c_{\sigma^2} + nd(T-1), \tag{40}$$

$$\bar{d}_{\sigma^2} = d_{\sigma^2} + \sum_{t=2}^T \sum_{i=1}^n \Big\{ \operatorname{tr}(\Sigma^i_t) + \boldsymbol{\mu}^{i\,\mathrm{T}}_t \boldsymbol{\mu}^i_t + \operatorname{tr}(\Sigma^i_{t-1}) + \boldsymbol{\mu}^{i\,\mathrm{T}}_{t-1} \boldsymbol{\mu}^i_{t-1} - 2(\operatorname{tr}(\Sigma^i_{t-1,t}) + \boldsymbol{\mu}^{i\,\mathrm{T}}_{t-1} \boldsymbol{\mu}^i_t) \Big\}. \tag{41}$$

Algorithm 4: Coordinate ascent updates for the latent trajectories. `KalmanSmoother` is the variational Kalman smoother defined in Algorithm 7 of Appendix D.

Proposition 10 *Under the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text,* $q(\Psi_0) = \text{Wishart}^{-1}(\bar{\nu}, \bar{V})$ *where* $\bar{\nu}$ *and* $\bar{V}$ *are defined in Equation (38) and Equation (39), respectively.*

Proof Standard calculations show that

$$\begin{aligned} p(\Psi_0 \mid \cdot) &\propto |\Psi_0|^{-(\nu+n+d+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}(V\Psi_0^{-1}) - \frac{1}{2}\sum_{i=1}^{n}\operatorname{tr}(\mathbf{X}_1^i \mathbf{X}_1^{i\,\mathrm{T}}\Psi_0^{-1})\right) \\ &\propto \text{Wishart}^{-1}\left(\nu + n, V + \sum_{i=1}^{n}\mathbf{X}_1^i\mathbf{X}_1^{i\,\mathrm{T}}\right). \end{aligned}$$

Note that $\text{Wishart}^{-1}(\nu, V)$ is in the exponential family with natural parameters ν and V. Thus, the variational distribution is also an inverse-Wishart distribution with natural parameters

$$\begin{aligned} \bar{\nu} &= \nu + n, \\ \bar{V} &= V + \sum_{i=1}^{n}\mathbb{E}_{q(\mathbf{X}_{1:T}^i)}\left[\mathbf{X}_1^i\mathbf{X}_1^{i\,\mathrm{T}}\right] \\ &= V + \sum_{i=1}^{n}\left(\Sigma_1^i + \boldsymbol{\mu}_1^i\boldsymbol{\mu}_1^{i\,\mathrm{T}}\right). \end{aligned}$$

■

Proposition 11 *Under the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text,* $q(\sigma^2) = \Gamma^{-1}(\bar{c}_{\sigma^2}/2, \bar{d}_{\sigma^2}/2)$ *where* $\bar{c}_{\sigma^2}$ *and* $\bar{d}_{\sigma^2}$ *are defined in Equation (40) and Equation (41), respectively.*

Proof Standard calculations show that

$$\begin{aligned} p(\sigma^2 \mid \cdot) &\propto \left(\frac{1}{\sigma^2}\right)^{(c_{\sigma^2}+nd(T-1))/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=2}^{T}\sum_{i=1}^{n}\|\mathbf{X}_t^i - \mathbf{X}_{t-1}^i\|_2^2 - \frac{d_{\sigma^2}}{2\sigma^2}\right), \\ &\propto \Gamma^{-1}\left(\frac{c_{\sigma^2}+nd(T-1)}{2}, \frac{1}{2}\left\{\sum_{t=2}^{T}\sum_{i=1}^{n}\|\mathbf{X}_t^i - \mathbf{X}_{t-1}^i\|_2^2 + d_{\sigma^2}\right\}\right). \end{aligned}$$

Note that $\Gamma^{-1}(a/2, b/2)$ is in the exponential family with natural parameters a and b. Thus, the variational distribution is also an inverse-gamma distribution with natural parameters

$$\begin{aligned} \bar{c}_{\sigma^2} &= c_{\sigma^2} + nd(T-1), \\ \bar{d}_{\sigma^2} &= d_{\sigma^2} + \sum_{t=2}^{T}\sum_{i=1}^{n}\mathbb{E}_{q(\mathbf{X}_{1:T}^i)}\left[\|\mathbf{X}_t^i - \mathbf{X}_{t-1}^i\|_2^2\right], \\ &= d_{\sigma^2} + \sum_{t=2}^{T}\sum_{i=1}^{n}\Big\{\operatorname{tr}(\Sigma_t^i) + \boldsymbol{\mu}_t^{i\,\mathrm{T}}\boldsymbol{\mu}_t^i + \operatorname{tr}(\Sigma_{t-1}^i) + \boldsymbol{\mu}_{t-1}^{i\,\mathrm{T}}\boldsymbol{\mu}_{t-1}^i \\ &\qquad - 2(\operatorname{tr}(\Sigma_{t-1,t}^i) + \boldsymbol{\mu}_{t-1}^{i\,\mathrm{T}}\boldsymbol{\mu}_t^i)\Big\}. \end{aligned}$$

■

Proposition 12 *Under the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text, $q(\mathbf{X}^i_{1:T})$ is a Gaussian state space model with natural parameters given in Equations (34)—(37).*

Proof First we lay out some notation. Let

$$\begin{aligned}
\boldsymbol{\theta}^i_{k,t} &= \delta^i_{k,t}\mathbf{1}_{n-1} + (\delta^1_{k,t}, \ldots, \delta^{i-1}_{k,t}, \delta^{i+1}_{k,t}, \ldots, \delta^n_{k,t})^{\mathrm{T}} \in \mathbb{R}^{n-1}, \\
\boldsymbol{\omega}^k_{it} &= (\omega^k_{i1t}, \ldots, \omega^k_{i(i-1)t}, \omega^k_{i(i+1)t}, \ldots, \omega^k_{int})^{\mathrm{T}} \in \mathbb{R}^{n-1}, \\
X^i_{k,t} &= (\mathbf{X}^1_t\Lambda_k, \ldots, \mathbf{X}^{i-1}_t\Lambda_k, \mathbf{X}^{i+1}_t\Lambda_k, \ldots, \mathbf{X}^n_t\Lambda_k)^{\mathrm{T}} \in \mathbb{R}^{(n-1)\times d}.
\end{aligned}$$

We then define the concatenated version of these quantities:

$$\begin{aligned}
\Omega^i_t &= \mathrm{diag}(\boldsymbol{\omega}^{1\,\mathrm{T}}_{it}, \ldots, \boldsymbol{\omega}^{K\,\mathrm{T}}_{it}) \in \mathbb{R}^{K(n-1)\times K(n-1)}, \\
\boldsymbol{\theta}^i_t &= (\boldsymbol{\delta}^{i\,\mathrm{T}}_{1,t}, \ldots, \boldsymbol{\delta}^{i\,\mathrm{T}}_{K,t})^{\mathrm{T}} \in \mathbb{R}^{K(n-1)},
\end{aligned}$$

and $X^i_t \in \mathbb{R}^{K(n-1)\times d}$ formed by stacking the matrices $X^i_{k,t}$ row-wise for $k = 1, \ldots, K$.

From Equation (5), we associate $p(\mathbf{X}^i_{1:T} \mid \cdot)$ with a GSSM parameterized by

$$\begin{aligned}
\mathbf{x}_t &= \mathbf{X}^i_t, \\
\mathbf{y}_t &= \mathbf{z}^i_t, \\
A_t &= \Omega^i_t\, X^i_t, \\
\mathbf{b}_t &= \Omega^i_t\, \boldsymbol{\theta}^i_t, \\
C_t &= \Omega^i_t.
\end{aligned} \tag{42}$$

We then apply Lemma 5 in Appendix C to identify the natural parameters. Note that $A_t^{\mathrm{T}} C_t^{-1} = X^i_t$. Taking into account the independence assumptions contained in the approximate posterior, we have

$$\begin{aligned}
\Gamma^1_t &= \langle X^i_t \rangle^{\mathrm{T}} \left(\mathbf{z}^i_t - \langle \Omega^i_t \rangle \langle \boldsymbol{\theta}^i_t \rangle \right), \\
&= \begin{pmatrix} \sum_{k=1}^K \sum_{j\neq i} \mu_{\lambda_{k1}} \mu^j_{t1} [Y^k_{ijt} - 1/2 - \mu_{\omega^k_{ijt}} (\mu_{\delta^i_{k,t}} + \mu_{\delta^j_{k,t}})] \\ \vdots \\ \sum_{k=1}^K \sum_{j\neq i} \mu_{\lambda_{kd}} \mu^j_{td} [Y^k_{ijt} - 1/2 - \mu_{\omega^k_{ijt}} (\mu_{\delta^i_{k,t}} + \mu_{\delta^j_{k,t}})] \end{pmatrix}.
\end{aligned} \tag{43}$$

Next, the individual elements of $\Gamma^2_t \in \mathbb{R}^{d\times d}$ are

$$(\Gamma^2_t)_{gh} = \left\langle X^{i\,\mathrm{T}}_t \Omega^i_t X^i_t \right\rangle_{gh} = \sum_{k=1}^K \sum_{j\neq i} \left\langle \omega^k_{ijt} \right\rangle \left\langle \lambda^k_g \lambda^k_h \right\rangle \left\langle X^j_{tg} X^j_{th} \right\rangle,$$

or

$$\Gamma_t^2 = \sum_{k=1}^{} \sum_{j \neq i} \boldsymbol{\mu}_{\omega_{ijt}^k} \mathbb{E}_{q(\boldsymbol{\lambda}_k)} \left[\boldsymbol{\lambda}_k \boldsymbol{\lambda}_k^{\mathrm{T}} \right] \odot \mathbb{E}_{q(\mathbf{X}_{1:T}^j)} \left[\mathbf{X}_t^j \mathbf{X}_t^{j\,\mathrm{T}} \right],$$

$$= \sum_{k=1}^{K} \sum_{j \neq i} \mu_{\omega_{ijt}^k} (\Sigma_{\boldsymbol{\lambda}_k} + \boldsymbol{\mu}_{\boldsymbol{\lambda}_k} \boldsymbol{\mu}_{\boldsymbol{\lambda}_k}^{\mathrm{T}}) \odot (\Sigma_t^j + \boldsymbol{\mu}_t^j \boldsymbol{\mu}_t^{j\,\mathrm{T}}). \tag{44}$$

Furthermore, from Proposition 10 and Proposition 11 we have that Ψ_0^{-1} is Wishart distributed with mean $\bar{\nu}\,\bar{V}^{-1}$ and $1/\sigma^2$ is gamma distributed with mean $\bar{c}_{\sigma^2}/\bar{d}_{\sigma^2}$, respectively. Finally, we can apply the variational Kalman smoothing equations derived in Appendix D to calculate the moments of $q(\mathbf{X}_{1:T}^i)$. ■

C.4 Derivation of Algorithm 5

The coordinate updates for $q(\boldsymbol{\lambda}_k)$ are given in Algorithm 5, which we formally derive in the remainder of this section.

Proposition 13 *Consider the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text. For $k \in \{2, \ldots, K\}$ (non-reference layers), $q(\boldsymbol{\lambda}_k) = N(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k}, \Sigma_{\boldsymbol{\lambda}_k})$ with parameters given in Equation (47) and Equation (48).*

Proof First we define some notation. Let $\mathcal{D} = \{(i,j) : j < i,\ 1 \leq i \leq n\}$ denote the set of dyads. Define $\mathbf{X}_t^{i \odot j} = \mathbf{X}_t^i \odot \mathbf{X}_t^j$ and $\theta_{k,t}^{ij} = \delta_{k,t}^i + \delta_{k,t}^j$. Let $\boldsymbol{\omega}_t^k = (\omega_{ijt}^k)_{((i,j) \in \mathcal{D})} \in \mathbb{R}^{|\mathcal{D}|}$ be a vector formed by stacking the ω_{ijt}^k by dyads. Also, let $\Omega_k = \mathrm{diag}(\boldsymbol{\omega}_1^k, \ldots, \boldsymbol{\omega}_T^k) \in \mathbb{R}^{|\mathcal{D}|T \times |\mathcal{D}|T}$. Finally, let $\mathbf{z}_k \in \mathbb{R}^{|\mathcal{D}|T}$, $X \in \mathbb{R}^{|\mathcal{D}|T \times d}$ and $\boldsymbol{\theta}_k \in \mathbb{R}^{|\mathcal{D}|T}$ be formed by stacking $z_{ijt}^k = Y_{ijt}^k - 1/2$, $\mathbf{X}_t^{i \odot j}$ and $\theta_{k,t}^{ij}$ first by dyads and then the result by time, respectively. Standard manipulations show that

$$p(\boldsymbol{\lambda}_k \mid \cdot) \propto p(\boldsymbol{\lambda}_k) N(\mathbf{z}_k \mid \Omega_k X \boldsymbol{\lambda}_k + \Omega_k \boldsymbol{\theta}_k, \Omega_k).$$

Since $p(\boldsymbol{\lambda}_k) = N(0, \sigma_\lambda^2 I_d)$, the full conditional distribution is Gaussian with the following natural parameters:

$$\Lambda = \left[X^{\mathrm{T}} \Omega_k X + \frac{1}{\sigma_\lambda^2} I_d \right],$$
$$\boldsymbol{\eta} = X^{\mathrm{T}} (\mathbf{z}_k - \Omega_k \boldsymbol{\theta}_k).$$

Taking expectations with respect to the approximate posterior and converting back to the mean and covariance parameters, we have

$$\Sigma_{\boldsymbol{\lambda}} = \left[\langle X^{\mathrm{T}} \Omega_k X \rangle + \frac{1}{\sigma_\lambda^2} I_d \right]^{-1},$$
$$\boldsymbol{\mu}_{\boldsymbol{\lambda}_k} = \Sigma_{\boldsymbol{\lambda}_k} \langle X \rangle^{\mathrm{T}} (\mathbf{z} - \langle \Omega_k \rangle \langle \boldsymbol{\theta}_k \rangle),$$

which is equivalent to the parameters in Equation (47) and Equation (48). ■

1. Update $q(\lambda_{1h}) = p_{\lambda_{1h}}^{\mathbb{1}\{\lambda_{1h}=1\}}(1-p_{\lambda_{1h}})^{\mathbb{1}\{\lambda_{1h}=-1\}}$:
   For $h \in \{1,\ldots,d\}$:

$$\eta_{\lambda_{1h}} = \log\left[\frac{\rho}{1-\rho}\right] + 2\sum_{t=1}^{T}\sum_{j<i}\Bigg\{(Y_{ijt}^1 - 1/2 - \mu_{\omega_{ijt}^i}(\mu_{\delta_{1,t}^i} + \mu_{\delta_{1,t}^j})\mu_{th}^i\mu_{th}^j - \mu_{\omega_{ijt}^1}\sum_{g\neq h}\mu_{\lambda_{1g}}((\Sigma_t^i)_{gh} + \mu_{tg}^i\mu_{th}^i)((\Sigma_t^i)_{gh} + \mu_{tg}^j\mu_{th}^j)\Bigg\}, \tag{45}$$

$$p_{\lambda_{1h}} = e^{\eta_{\lambda_{1h}}}/(1+e^{\eta_{\lambda_{1h}}}), \quad \mu_{\lambda_{1h}} = 2p_{\lambda_{1h}} - 1, \quad \sigma_{\lambda_{1h}}^2 = 1-(2p_{\lambda_{1h}}-1)^2. \tag{46}$$

2. Update $q(\boldsymbol{\lambda}_k) = N(\boldsymbol{\mu}_{\boldsymbol{\lambda}_k}, \Sigma_{\boldsymbol{\lambda}_k})$:
   For $k \in \{2,\ldots,K\}$:

$$\Sigma_{\boldsymbol{\lambda}_k} = \left[\sum_{t=1}^{T}\sum_{j<i}\mu_{\omega_{ijt}^k}(\Sigma_t^i + \boldsymbol{\mu}_t^i\boldsymbol{\mu}_t^{i\,\mathrm{T}}) \odot (\Sigma_t^j + \boldsymbol{\mu}_t^j\boldsymbol{\mu}_t^{j\,\mathrm{T}}) + \frac{1}{\sigma_\lambda^2}I_p\right]^{-1}, \tag{47}$$

$$\boldsymbol{\mu}_{\boldsymbol{\lambda}_k} = \Sigma_{\boldsymbol{\lambda}_k}\begin{pmatrix}\sum_{t=1}^{T}\sum_{j<i}[Y_{ijt}^k - 1/2 - \mu_{\omega_{ijt}^k}(\mu_{\delta_{k,t}^i} + \mu_{\delta_{k,t}^j})]\mu_{t1}^i\mu_{t1}^j \\ \vdots \\ \sum_{t=1}^{T}\sum_{j<i}[Y_{ijt}^k - 1/2 - \mu_{\omega_{ijt}^k}(\mu_{\delta_{k,t}^i} + \mu_{\delta_{k,t}^j})]\mu_{td}^i\mu_{td}^j\end{pmatrix}. \tag{48}$$

Algorithm 5: Coordinate ascent updates for the homophily coefficients.

Proposition 14 *Consider the eigenmodel for dynamic multilayer networks with the same prior distributions and variational factorization defined in the main text. For $h \in \{1, \ldots, d\}$, $q(\lambda_{1h}) = p_{\lambda_{1h}}^{\mathbb{1}\{\lambda_{1h}=1\}} (1 - p_{\lambda_{1h}})^{\mathbb{1}\{\lambda_{1h}=-1\}}$ where $p_{\lambda_{1h}}$ is given in Equation (46).*

Proof The full conditional distributions are

$$p(\lambda_{1h} \mid \cdot) \propto p(\lambda_{1h}) \exp\left\{ \sum_{t=1}^{T} \sum_{j<i} \left[(Y_{ijt}^1 - 1/2 - \omega_{ijt}^k(\delta_{1,t}^i + \delta_{1,t}^j))\lambda_{1h} X_{th}^i X_{th}^j - \right.\right.$$
$$\left.\left. \frac{1}{2}\omega_{ijt}^1 \lambda_{1q}^2 (X_{th}^i)^2 (X_{th}^j)^2 - \omega_{ijt}^1 \sum_{g \neq h} \lambda_{1g}\lambda_{1h} X_{tg}^i X_{th}^i X_{th}^j X_{tg}^j \right] \right\}. \tag{49}$$

The natural parameter is then

$$\begin{aligned}
\eta_{\lambda_{1h}} &= \mathbb{E}_{-q(\lambda_{1h})}\left[\log p(\lambda_{1h} = 1 \mid \cdot)\right] - \mathbb{E}_{-q(\lambda_{1h})}\left[\log p(\lambda_{1h} = -1 \mid \cdot)\right] \\
&= \log\left[\frac{\rho}{1-\rho}\right] + \\
&\quad 2\sum_{t=1}^{T}\sum_{j<i}\left\{ (Y_{ijt}^1 - 1/2\mu_{\omega_{ijt}^1}(\mu_{\delta_{1,t}^i} + \mu_{\delta_{1,t}^j}))\mu_{th}^i \mu_{th}^j - \right. \\
&\quad \left. \mu_{\omega_{ijt}^1} \sum_{g \neq h} \mu_{\lambda_{1g}}((\Sigma_t^i)_{gh} + \mu_{tg}^i \mu_{th}^i)((\Sigma_t^j)_{gh} + \mu_{tg}^j \mu_{th}^i) \right\}.
\end{aligned}$$

Converting back to the standard parameterization, we have

$$\begin{aligned}
p_{\lambda_{1h}} &= e^{\eta_{\lambda_{1h}}}/(1 + e^{\eta_{\lambda_{1h}}}), \\
\mu_{\lambda_{1h}} &= 2p_{\lambda_{1h}} - 1, \\
\sigma_{\lambda_{1h}}^2 &= 1 - (2p_{\lambda_{1h}} - 1)^2.
\end{aligned}$$

■

Appendix D. Derivation of the Variational Kalman Smoother

In this section, we derive the variational Kalman smoother used for inference in our model. Many of our results are based on the work in Beal (2003). The major difference between the two formulations is that we incorporate time-varying state space parameters and non-identity covariance matrices.

Consider the Gaussian state space model specified by Equations (16)—(18). Our goal is to perform variational inference based on the factorization $q(\mathbf{x}_{1:T})q(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ contains the parameters of the GSSM. Crucially, the variational distributions of the hidden states are also GSSMs. Indeed,

$$\begin{aligned}
q(\mathbf{x}_{1:T}) = c\exp(\langle \log p(\mathbf{x}_{1:T}, \mathbf{y}_{1:T})\rangle) &\triangleq h(\mathbf{x}_{1:T}, \mathbf{y}_{1:T}) \\
&= h(\mathbf{x}_1)h(\mathbf{y}_1 \mid \mathbf{x}_1)\prod_{t=2}^{T} h(\mathbf{x}_t \mid \mathbf{x}_{t-1})h(\mathbf{y}_t \mid \mathbf{x}_t),
\end{aligned}$$

Given the natural parameters $\langle \Gamma^1_{1:T} \rangle = \langle A^{\mathrm{T}}_{1:T} C^{-1}_{1:T} \rangle \mathbf{y}_{1:T} - \langle A^{\mathrm{T}}_{1:T} C^{-1}_{1:T} \mathbf{b}_{1:T} \rangle$, $\langle \Gamma^2_{1:T} \rangle = \langle A^{\mathrm{T}}_{1:T} C^{-1}_{1:T} A_{1:T} \rangle$, $\langle \Psi_0^{-1} \rangle$, and $\langle 1/\sigma^2 \rangle$, calculate the filtering distribution's moments as follows:

1. For $t = 1$:

$$\Sigma_1 = \left[\langle \Psi_0^{-1} \rangle + \langle A_1^{\mathrm{T}} C_1^{-1} A_1 \rangle \right]^{-1},$$
$$\boldsymbol{\mu}_1 = \Sigma_1 \left[\langle A_1^{\mathrm{T}} C_1^{-1} \rangle \mathbf{y}_1 - \langle A_1^{\mathrm{T}} C_1^{-1} \mathbf{b}_1 \rangle \right].$$

2. For $t = 2, \ldots, T$:

$$\Sigma^*_{t-1} = \left[\left\langle \frac{1}{\sigma^2} \right\rangle I_d + \Sigma^{-1}_{t-1} \right]^{-1},$$
$$\Sigma_t = \left[\left\langle \frac{1}{\sigma^2} \right\rangle I_d + \langle A_t^{\mathrm{T}} C_t^{-1} A_t \rangle - \left\langle \frac{1}{\sigma^2} \right\rangle^2 \Sigma^*_{t-1} \right]^{-1},$$
$$\boldsymbol{\mu}_t = \Sigma_t \left[\langle A_t^{\mathrm{T}} C_t^{-1} \rangle \mathbf{y}_t - \langle A_t^{\mathrm{T}} C_t^{-1} \mathbf{b}_t \rangle + \left\langle \frac{1}{\sigma^2} \right\rangle \Sigma^*_{t-1} \Sigma^{-1}_{t-1} \boldsymbol{\mu}_{t-1} \right].$$

Output $\{\boldsymbol{\mu}_{1:T}, \Sigma_{1:T}, \Sigma^*_{1:(T-1)}\}$.

Algorithm 6: Variational Kalman filter.

where

$$h(\mathbf{x}_1) = c_1 \exp(\langle \log N(\mathbf{x}_1 \mid 0, \Psi_0) \rangle),$$
$$h(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = c_2 \exp(\langle \log N(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \sigma^2 I_d) \rangle),$$
$$h(\mathbf{y}_t \mid \mathbf{x}_t) = c_3 \exp(\langle \log N(\mathbf{y}_t \mid A_t \mathbf{x}_t + \mathbf{b}_t, C_t) \rangle),$$

are Gaussian distributions and $\langle \cdot \rangle$ denotes an expectation with respect to $q(\boldsymbol{\theta})$. When calculating the distributions in $q(\mathbf{x}_{1:T})$, we use the Gaussian distribution's natural parameter form so that the expectations have an analytical form. The trade-off in using this parameterization is that when calculating the moments of these variational GSSMs, the standard Kalman smoothing equations are no longer applicable because we only have access to the natural parameters. Therefore, we derive Kalman smoothing equations that are expressed in terms of the variational distribution's natural parameters.

D.1 Variational Kalman Filter

Property 15 *For the Gaussian state space model specified in Equations (16)—(18), the variational filtering distributions $h(\mathbf{x}_t \mid \mathbf{y}_{1:t}) = N(\mathbf{x}_t \mid \boldsymbol{\mu}_t, \Sigma_t)$ with parameters that can be calculated recursively via Algorithm 6.*

Proof Define the forwards message variables as $\alpha_t(\mathbf{x}_t) = h(\mathbf{x}_t \mid \mathbf{y}_{1:t})$. The filter proceeds recursively starting at $t = 1$:

$$\begin{aligned}
\alpha_1(\mathbf{x}_1) &\propto h(\mathbf{x}_1)h(\mathbf{y}_1 \mid \mathbf{x}_1), \\
&\propto \exp\left(-\frac{1}{2}\langle \mathbf{x}_1^{\mathrm{T}}\Psi_0^{-1}\mathbf{x}_1 + (\mathbf{y}_1 - A_1\mathbf{x}_1 - \mathbf{b}_1)^{\mathrm{T}}C_1^{-1}(\mathbf{y}_1 - A_1\mathbf{x}_1 - \mathbf{b}_1)\rangle\right), \\
&\propto \exp\Bigg(-\frac{1}{2}\mathbf{x}_1^{\mathrm{T}}(\langle\Psi_0^{-1}\rangle + \langle A_1^{\mathrm{T}}C_1^{-1}A_1\rangle)\mathbf{x}_1 + \\
&\qquad\qquad (\langle A_1^{\mathrm{T}}C_1^{-1}\rangle\mathbf{y}_1 - \langle A_1^{\mathrm{T}}C_1^{-1}\mathbf{b}_1\rangle)^{\mathrm{T}}\mathbf{x}_1\Bigg), \\
&\propto N(\mathbf{x}_1 \mid \boldsymbol{\mu}_1, \Sigma_1),
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_1 &= \left[\langle\Psi_0^{-1}\rangle + \langle A_1^{\mathrm{T}}C_1^{-1}A_1\rangle\right]^{-1}, \\
\boldsymbol{\mu}_1 &= \Sigma_1\left[\langle A_1^{\mathrm{T}}C_1^{-1}\rangle\mathbf{y}_1 - \langle A_1^{\mathrm{T}}C_1^{-1}\mathbf{b}_1\rangle\right].
\end{aligned}$$

The last line follows from matching the natural parameters of a Gaussian density. This shows that $\alpha_1(\mathbf{x}_1)$ is Gaussian; therefore, we can proceed by induction. For $t > 1$, we have

$$\begin{aligned}
\alpha_t(\mathbf{x}_t) &\propto \int d\mathbf{x}_{t-1}\, h(\mathbf{x}_t, \mathbf{y}_{1:t}, \mathbf{x}_{t-1}) \\
&= \int d\mathbf{x}_{t-1}\, h(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1})h(\mathbf{x}_t \mid \mathbf{x}_{t-1})h(\mathbf{y}_t \mid \mathbf{x}_t) \\
&\propto \int d\mathbf{x}_{t-1}\, \alpha_{t-1}(\mathbf{x}_{t-1})\times \\
&\qquad \exp\Bigg(-\frac{1}{2}\Big[\mathbf{x}_{t-1}^{\mathrm{T}}\left\langle\frac{1}{\sigma^2}\right\rangle I_d\mathbf{x}_{t-1} - 2\mathbf{x}_{t-1}^{\mathrm{T}}\left\langle\frac{1}{\sigma^2}\right\rangle I_d\mathbf{x}_t + \\
&\qquad \mathbf{x}_t^{\mathrm{T}}\left(\left\langle\frac{1}{\sigma^2}\right\rangle I_d + \langle A_t^{\mathrm{T}}C_t^{-1}A_t\rangle\right)\mathbf{x}_t - 2\big(\langle A_t^{\mathrm{T}}C_t^{-1}\rangle\mathbf{y}_t - \langle A_t^{\mathrm{T}}C_t^{-1}\mathbf{b}_t\rangle\big)^{\mathrm{T}}\mathbf{x}_t\Big]\Bigg).
\end{aligned}$$

Inside the integral is a $N(\mathbf{x}_{t-1} \mid \boldsymbol{\mu}^*_{t-1}, \Sigma^*_{t-1})$ with

$$\begin{aligned}
\Sigma^*_{t-1} &= \left[\left\langle\frac{1}{\sigma^2}\right\rangle I_d + \Sigma^{-1}_{t-1}\right]^{-1}, \\
\boldsymbol{\mu}^*_{t-1} &= \Sigma^*_{t-1}\left[\left\langle\frac{1}{\sigma^2}\right\rangle\mathbf{x}_t + \Sigma^{-1}_{t-1}\boldsymbol{\mu}_{t-1}\right].
\end{aligned}$$

Marginalizing over this distribution leaves the following terms

$$\begin{aligned}
\alpha_t(\mathbf{x}_t) &\propto \exp\Bigg(-\frac{1}{2}\mathbf{x}_t^{\mathrm{T}}\Big(\Big\langle\frac{1}{\sigma^2}\Big\rangle I_d + \langle A_t^{\mathrm{T}} C_t^{-1} A_t\rangle - \Big\langle\frac{1}{\sigma^2}\Big\rangle\Sigma_{t-1}^*\Big\langle\frac{1}{\sigma^2}\Big\rangle\Big)\mathbf{x}_t + \\
&\qquad \Big(\langle A_t^{\mathrm{T}} C_t^{-1}\rangle \mathbf{y}_t - \langle A_t^{\mathrm{T}} C_t^{-1}\mathbf{b}_t\rangle\Big)^{\mathrm{T}}\mathbf{x}_t + \frac{1}{2}\boldsymbol{\mu}_{t-1}^{*\mathrm{T}}\Sigma_{t-1}^{*-1}\boldsymbol{\mu}_{t-1}^*\Bigg) \\
&\propto \exp\Bigg(-\frac{1}{2}\mathbf{x}_t^{\mathrm{T}}\Big(\Big\langle\frac{1}{\sigma^2}\Big\rangle I_d + \langle A_t^{\mathrm{T}} C_t^{-1} A_t\rangle - \Big\langle\frac{1}{\sigma^2}\Big\rangle\Sigma_{t-1}^*\Big\langle\frac{1}{\sigma^2}\Big\rangle\Big)\mathbf{x}_t + \\
&\qquad \Big(\langle A_t^{\mathrm{T}} C_t^{-1}\rangle \mathbf{y}_t - \langle A_t^{\mathrm{T}} C_t^{-1}\mathbf{b}_t\rangle + \Big\langle\frac{1}{\sigma^2}\Big\rangle\Sigma_{t-1}^*\Sigma_{t-1}^{-1}\boldsymbol{\mu}_{t-1}\Big)^{\mathrm{T}}\mathbf{x}_t\Bigg) \\
&\propto N(\mathbf{x}_t \mid \boldsymbol{\mu}_t, \Sigma_t),
\end{aligned}$$

where

$$\begin{aligned}
\Sigma_t &= \Bigg[\Big\langle\frac{1}{\sigma^2}\Big\rangle I_d + \langle A_t^{\mathrm{T}} C_t^{-1} A_t\rangle - \Big\langle\frac{1}{\sigma^2}\Big\rangle^2\Sigma_{t-1}^*\Bigg]^{-1}, \\
\boldsymbol{\mu}_t &= \Sigma_t\Bigg[\langle A_t^{\mathrm{T}} C_t^{-1}\rangle \mathbf{y}_t - \langle A_t^{\mathrm{T}} C_t^{-1}\mathbf{b}_t\rangle + \Big\langle\frac{1}{\sigma^2}\Big\rangle\Sigma_{t-1}^*\Sigma_{t-1}^{-1}\boldsymbol{\mu}_{t-1}\Bigg].
\end{aligned}$$

■

D.2 Variational Kalman Smoother

Property 16 *For the Gaussian state space model specified in Equations (16)—(18), the backwards message variables* $h(\mathbf{y}_{(t+1):T} \mid \mathbf{x}_t) = N(\mathbf{x}_t \mid \boldsymbol{\eta}_t, \Psi_t)$ *with parameters that can be calculated recursively via Algorithm 7.*

Proof Define the backwards message variables $\beta_t(\mathbf{x}_t) = h(\mathbf{y}_{(t+1):T} \mid \mathbf{x}_t)$ and set $\beta_T(\mathbf{x}_T) = 1$. Note that

$$\begin{aligned}
\beta_{t-1}(\mathbf{x}_{t-1}) &= \int d\mathbf{x}_t\, h(\mathbf{y}_{t:T}, \mathbf{x}_t \mid \mathbf{x}_{t-1}) \\
&= \int d\mathbf{x}_t\, h(\mathbf{y}_{(t+1):T} \mid \mathbf{x}_t) h(\mathbf{y}_t \mid \mathbf{x}_t) h(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \\
&= \int d\mathbf{x}_t\, \beta_t(\mathbf{x}_t) h(\mathbf{y}_t \mid \mathbf{x}_t) h(\mathbf{x}_t \mid \mathbf{x}_{t-1}).
\end{aligned}$$

Given the natural parameters $\langle \Gamma^1_{1:T} \rangle = \langle A^{\mathrm{T}}_{1:T} C^{-1}_{1:T} \rangle \mathbf{y}_{1:T} - \langle A^{\mathrm{T}}_{1:T} C^{-1}_{1:T} \mathbf{b}_{1:T} \rangle$, $\langle \Gamma^2_{1:T} \rangle = \langle A^{\mathrm{T}}_{1:T} C^{-1}_{1:T} A_{1:T} \rangle$, $\langle \Psi_0^{-1} \rangle$, and $\langle 1/\sigma^2 \rangle$, calculate the smoothing distribution's moments and cross-covariances as follows:

1. Calculate $\{\boldsymbol{\mu}_{1:T}, \Sigma_{1:T}, \Sigma^*_{1:(T-1)}\}$ as in Algorithm 6.
2. Set $\boldsymbol{\nu}_T = \boldsymbol{\mu}_T$ and $\Upsilon_T = \Sigma_T$.
3. Initialize $\Psi_T^{-1} = 0$ to satisfy $\beta_T(\mathbf{x}_T) = 1$.
4. For $t = T, \ldots, 2$.

   ▷ Calculate backwards message variables

$$\Psi^*_t = \left[\left\langle \frac{1}{\sigma^2} \right\rangle I_d + \langle A^{\mathrm{T}}_t C^{-1}_t A_t \rangle + \Psi_t^{-1} \right]^{-1},$$
$$\Psi_{t-1} = \left[\left\langle \frac{1}{\sigma^2} \right\rangle I_d - \left\langle \frac{1}{\sigma^2} \right\rangle^2 \Psi^*_t \right]^{-1},$$
$$\boldsymbol{\eta}_{t-1} = \Psi_{t-1} \left[\left\langle \frac{1}{\sigma^2} \right\rangle \Psi^*_t \left(\langle A^{\mathrm{T}}_t C^{-1}_t \rangle \mathbf{y}_t - \langle A^{\mathrm{T}}_t C^{-1}_t \mathbf{b}_t \rangle + \Psi_t^{-1} \boldsymbol{\eta}_t \right) \right].$$

   ▷ Calculate smoothing distributions and cross-covariances

$$\Upsilon_{t-1} = \left[\Sigma^{-1}_{t-1} + \Psi^{-1}_{t-1} \right]^{-1},$$
$$\Upsilon_{t-1,t} = \left\langle \frac{1}{\sigma^2} \right\rangle \Sigma^*_{t-1} \left[\Psi^{*\,-1}_t - \left\langle \frac{1}{\sigma^2} \right\rangle^2 \Sigma^*_{t-1} \right]^{-1},$$
$$\boldsymbol{\nu}_{t-1} = \Upsilon_{t-1} \left[\Sigma^{-1}_{t-1} \boldsymbol{\mu}_{t-1} + \Psi^{-1}_{t-1} \boldsymbol{\eta}_{t-1} \right].$$

Output smoothing distributions $\{\boldsymbol{\nu}_{1:T}, \Upsilon_{1:T}\}$ and cross-covariances $\{\Upsilon_{t,t+1}\}_{t=1}^{T-1}$.

Algorithm 7: Variational Kalman smoother. Throughout the text, we refer to this algorithm as `KalmanSmoother`$(\langle \Gamma^1_{1:T} \rangle, \langle \Gamma^2_{1:T} \rangle, \langle \Psi_0^{-1} \rangle, \langle 1/\sigma^2 \rangle)$.

The derivation proceeds sequentially backwards in time. For $t = T-1$, we have

$$\begin{aligned}
\beta_{T-1}(\mathbf{x}_{T-1}) &\propto \int d\mathbf{x}_T \times \\
&\exp\Bigg(-\frac{1}{2}\Big\langle (\mathbf{y}_T - A_T\mathbf{x}_T - \mathbf{b}_T)^{\mathrm{T}} C_T^{-1}(\mathbf{y}_T - A_T\mathbf{x}_T - \mathbf{b}_T) + \\
&\quad (\mathbf{x}_T - \mathbf{x}_{T-1})^{\mathrm{T}} \frac{1}{\sigma^2} I_d (\mathbf{x}_T - \mathbf{x}_{T-1}) \Big\rangle \Bigg) \\
&\propto \int d\mathbf{x}_T \exp\Bigg(-\frac{1}{2}\Big[\mathbf{x}_{T-1}\Big\langle \frac{1}{\sigma^2}\Big\rangle I_d \mathbf{x}_{T-1} + \mathbf{x}_T^{\mathrm{T}}(\Big\langle \frac{1}{\sigma^2}\Big\rangle I_d + \langle A_T^{\mathrm{T}} C_T^{-1} A_T\rangle)\mathbf{x}_T - \\
&\quad 2(\langle A_T^{\mathrm{T}} C_T^{-1}\rangle \mathbf{y}_T - \langle A_T^{\mathrm{T}} C_T^{-1}\mathbf{b}_T\rangle + \Big\langle \frac{1}{\sigma^2}\Big\rangle \mathbf{x}_{T-1})^{\mathrm{T}}\mathbf{x}_T \Big]\Bigg).
\end{aligned}$$

Inside the integral is a $N(\mathbf{x}_T \mid \boldsymbol{\eta}_T^*, \Psi_T^*)$ with

$$\begin{aligned}
\Psi_T^* &= \left[\Big\langle \frac{1}{\sigma^2}\Big\rangle I_d + \langle A_T^{\mathrm{T}} C_T^{-1} A_T\rangle \right]^{-1}, \\
\boldsymbol{\eta}_T^* &= \Psi_T^* \left[\langle A_T^{\mathrm{T}} C_T^{-1}\rangle \mathbf{y}_T - \langle A_T^{\mathrm{T}} C_T^{-1}\mathbf{b}_T\rangle + \Big\langle \frac{1}{\sigma^2}\Big\rangle \mathbf{x}_{T-1} \right].
\end{aligned}$$

Marginalizing over this density, we are left with

$$\begin{aligned}
\beta_{T-1}(\mathbf{x}_{T-1}) &\propto \exp\Bigg(-\frac{1}{2}\mathbf{x}_{T-1}^{\mathrm{T}}\Big\langle \frac{1}{\sigma^2}\Big\rangle \mathbf{x}_{T-1} + \frac{1}{2}\boldsymbol{\eta}_T^{*\mathrm{T}} \Psi_T^{*-1} \boldsymbol{\eta}_T^* \Bigg) \\
&\propto \exp\Bigg(-\frac{1}{2}\mathbf{x}_{T-1}^{\mathrm{T}}(\Big\langle \frac{1}{\sigma^2}\Big\rangle I_d - \Big\langle \frac{1}{\sigma^2}\Big\rangle^2 \Psi_T^*)\mathbf{x}_{T-1} + \\
&\quad (\langle A_T^{\mathrm{T}} C_T^{-1}\rangle \mathbf{y}_T - \langle A_T^{\mathrm{T}} C_T^{-1}\mathbf{b}_T\rangle)^{\mathrm{T}} \Big\langle \frac{1}{\sigma^2}\Big\rangle \Psi_T^* \mathbf{x}T - 1 \Bigg) \\
&\propto N(\mathbf{x}_{T-1} \mid \boldsymbol{\eta}_{T-1}, \Psi_{T-1}),
\end{aligned}$$

where

$$\begin{aligned}
\Psi_{T-1} &= \left[\Big\langle \frac{1}{\sigma^2}\Big\rangle I_d - \Big\langle \frac{1}{\sigma^2}\Big\rangle^2 \Psi_T^* \right]^{-1}, \\
\boldsymbol{\eta}_{T-1} &= \Psi_{T-1} \left[\Big\langle \frac{1}{\sigma^2}\Big\rangle \Psi_T^* \left(\langle A_T^{\mathrm{T}} C_T^{-1}\rangle \mathbf{y}_T - \langle A_T^{\mathrm{T}} C_T^{-1}\mathbf{b}_T\rangle \right) \right].
\end{aligned}$$

Now we proceed inductively. For $2 < t < T-1$, we have

$$\begin{aligned}
\beta_{t-1}(\mathbf{x}_{t-1}) &\propto \int d\mathbf{x}_t \, N(\mathbf{x}_t \mid \boldsymbol{\eta}_t, \Psi_t) \times \\
&\exp\Bigg(-\frac{1}{2}\Big\langle (\mathbf{y}_t - A_t\mathbf{x}_t - \mathbf{b}_t)^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - A_t\mathbf{x}_t - \mathbf{b}_t) + \\
&\quad (\mathbf{x}_t - \mathbf{x}_{t-1})^{\mathrm{T}} \frac{1}{\sigma^2} I_d (\mathbf{x}_t - \mathbf{x}_{t-1}) \Big\rangle \Bigg).
\end{aligned}$$

In fact, the remaining derivation is exactly the same as before, except

$$\Psi_t^* = \left[\left\langle \frac{1}{\sigma^2} \right\rangle I_d + \left\langle A_t^{\mathrm{T}} C_t^{-1} A_t \right\rangle + \Psi_t^{-1}\right]^{-1},$$
$$\boldsymbol{\eta}_t^* = \Psi_t^* \left[\left\langle A_t^{\mathrm{T}} C_t^{-1} \right\rangle \mathbf{y}_t - \left\langle A_t^{\mathrm{t}} C_t^{-1} \mathbf{b}_t \right\rangle + \Psi_t^{-1} \boldsymbol{\eta}_t + \left\langle \frac{1}{\sigma^2} \right\rangle \mathbf{x}_{t-1}\right],$$

so that $\beta_{t-1}(\mathbf{x}_{t-1}) = N(\mathbf{x}_{t-1} \mid \boldsymbol{\eta}_{t-1}, \Psi_{t-1})$ with

$$\Psi_{t-1} = \left[\left\langle \frac{1}{\sigma^2} \right\rangle I_d - \left\langle \frac{1}{\sigma^2} \right\rangle^2 \Psi_T^*\right]^{-1},$$
$$\boldsymbol{\eta}_{t-1} = \Psi_{t-1}\left[\left\langle \frac{1}{\sigma^2} \right\rangle \Psi_t^* \left(\left\langle A_t^{\mathrm{T}} C_t^{-1} \right\rangle \mathbf{y}_t - \left\langle A_t^{\mathrm{T}} C_t^{-1} \mathbf{b}_t \right\rangle + \Psi_t^{-1} \boldsymbol{\eta}_t\right)\right].$$

■

Property 17 *For the Gaussian state space model specified in Equations (16)—(18), the variational smoothing distributions $h(\mathbf{x}_t \mid \mathbf{y}_{1:T}) = N(\mathbf{x}_t \mid \boldsymbol{\nu}_t, \Upsilon_t)$ with parameters that can be calculated recursively via Algorithm 7.*

Proof We define forwards and backwards message variables, $\alpha_t(\mathbf{x}_t)$ and $\beta_t(\mathbf{x}_t)$, as in the previous proofs. For $t = T$, $h(\mathbf{x}_T \mid \mathbf{y}_{1:T}) = \alpha_T(\mathbf{x}_T) = N(\mathbf{x}_T \mid \boldsymbol{\mu}_T, \Sigma_T)$. For $t < T - 1$, we have

$$\begin{aligned} h(\mathbf{x}_t \mid \mathbf{y}_{1:T}) &\propto \alpha_t(\mathbf{x}_t)\beta_t(\mathbf{x}_t) \\ &\propto N(\mathbf{x}_t \mid \boldsymbol{\mu}_t, \Sigma_t) N(\mathbf{x}_t \mid \boldsymbol{\eta}_t, \Psi_t) \\ &\propto N(\mathbf{x}_t \mid \boldsymbol{\nu}_t, \Upsilon_t), \end{aligned}$$

where

$$\Upsilon_t = \left[\Sigma_t^{-1} + \Psi_t^{-1}\right]^{-1},$$
$$\boldsymbol{\nu}_t = \Upsilon_t \left[\Sigma_t^{-1} \boldsymbol{\mu}_t + \Psi_t^{-1} \boldsymbol{\eta}_t\right].$$

■

D.3 Cross-Covariance Matrices

Property 18 *For the Gaussian state space model specified in Equations (16)—(18), the variational joint distributions $h(\mathbf{x}_t, \mathbf{x}_{t+1} \mid \mathbf{y}_{1:T})$ are Gaussian with cross-covariance matrices $\Upsilon_{t,t+1}$ that can be calculated recursively via Algorithm 7.*

Proof We have that

$$\begin{aligned} h(\mathbf{x}_t, \mathbf{x}_{t+1} \mid \mathbf{y}_{1:T}) &\propto h(\mathbf{x}_t \mid \mathbf{x}_{1:t}) h(\mathbf{x}_{t+1} \mid \mathbf{x}_t) h(\mathbf{y}_{t+1} \mid \mathbf{x}_{t+1}) h(\mathbf{y}_{(t+2):T} \mid \mathbf{x}_{t+1}), \\ &\propto \alpha_t(\mathbf{x}_t) h(\mathbf{x}_{t+1} \mid \mathbf{x}_t) h(\mathbf{y}_{t+1} \mid \mathbf{x}_{t+1}) \beta_{t+1}(\mathbf{x}_{t+1}). \end{aligned}$$

To determine $\Upsilon_{t,t+1}$, we identify the cross-terms in the above product. This product is proportional to

$$\begin{aligned}
&\alpha_t(\mathbf{x}_t) \times \\
&\quad \exp\Bigg(-\frac{1}{2}\Big(-\mathbf{x}_t^{\mathrm{T}}\Big\langle\frac{1}{\sigma^2}\Big\rangle I_d\mathbf{x}_{t+1} + \mathbf{x}_t^{\mathrm{T}}\Big\langle\frac{1}{\sigma^2}\Big\rangle I_d\mathbf{x}_t + \mathbf{x}_{t+1}^{\mathrm{T}}\Big\langle\frac{1}{\sigma^2}\Big\rangle I_d\mathbf{x}_{t+1} \\
&\qquad \mathbf{x}_{t+1}^{\mathrm{T}}\langle A_{t+1}^{\mathrm{T}}C_{t+1}^{-1}A_{t+1}\rangle\mathbf{x}_{t+1} - \\
&\qquad 2\big(\langle A_{t+1}^{\mathrm{T}}C_{t+1}^{-1}\rangle\mathbf{y}_{t+1} + \langle A_{t+1}^{\mathrm{T}}C_{t+1}^{-1}\mathbf{b}_{t+1}\rangle\big)^{\mathrm{T}}\mathbf{x}_{t+1}\Big)\Bigg) \times \\
&\quad \beta_{t+1}(\mathbf{x}_{t+1}), \\
&\propto \exp\left\{-\frac{1}{2}\begin{pmatrix}\mathbf{x}_t\\ \mathbf{x}_{t+1}\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\Gamma_t & \Gamma_{t,t+1}\\ \Gamma_{t,t+1}^{\mathrm{T}} & \Gamma_{t+1}\end{pmatrix}\begin{pmatrix}\mathbf{x}_t\\ \mathbf{x}_{t+1}\end{pmatrix}\right\},
\end{aligned}$$

with

$$\begin{aligned}
\Gamma_t &= \Big\langle\frac{1}{\sigma^2}\Big\rangle I_d + \Sigma_t^{-1} = \Sigma_t^{*\,-1}, \\
\Gamma_{t+1} &= \Big\langle\frac{1}{\sigma^2}\Big\rangle I_d + \langle A_{t+1}^{\mathrm{T}}C_{t+1}^{-1}A_{t+1}\rangle + \Psi_{t+1}^{-1} = \Psi_{t+1}^{*\,-1}, \\
\Gamma_{t,t+1} &= -\Big\langle\frac{1}{\sigma^2}\Big\rangle I_d,
\end{aligned}$$

where in the last line we only kept terms quadratic in $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$. Applying Schur's compliment, we obtain the cross-covariance matrix

$$\Upsilon_{t,t+1} = \Sigma_t^*\Big\langle\frac{1}{\sigma^2}\Big\rangle\left[\Psi_{t+1}^{*\,-1} - \Big\langle\frac{1}{\sigma^2}\Big\rangle^2\Sigma_t^*\right]^{-1}.$$

■

Appendix E. Parameter Initialization and Prior Settings

Due to variational inference algorithm's non-convex objective, appropriate initial values of the approximate posterior's parameters can greatly improve convergence. In this section, we outline the initialization scheme and prior settings that we used in all experiments and real data applications. Our initialization method is based on first estimating the probability matrices $P_k^t = \mathrm{logit}^{-1}(\mathbf{\Theta}_t^k)$ using universal singular value thresholding (USVT) proposed by Chatterjee (2015) and then computing the initial estimates of $\boldsymbol{\delta}_{1:K,1:T}, \Lambda_{1:K}, \mathbf{X}_{1:T}$ heuristically by inverting the logit transform. A similar procedure was proposed by Ma et al. (2020) to initialize a bilinear latent space model for a single-layer static network. A step specific to our model involves estimating the shared latent positions by finding the shared column space of the layer specific residual matrices at each time point using a singular value decomposition. Our proposed initialization procedure is summarized in Algorithm 8.

Given numeric precision ε, latent space dimension d, and reference layer r, perform the following steps:

1. For each $k \in \{1, \ldots, K\}$ and $t \in \{1, \ldots, T\}$:
   (a) *(USVT).* Define the threshold $\tau = \sqrt{n\hat{p}}$, where $\hat{p} = \binom{n}{2}^{-1} \sum_{j<i} Y^k_{ijt}$, i.e., the density of $\mathbf{Y}^k_t$. Let $\tilde{P}^k_t = \sum_{s_i \geq \tau} s_i \mathbf{u}_i \mathbf{v}_i^{\mathrm{T}}$ where $\sum_{i=1}^n s_i \mathbf{u}_i \mathbf{v}_i^{\top}$ is the singular value decomposition of $\mathbf{Y}^k_t$. Project $\tilde{P}^k_t$ elementwise to the interval $[\varepsilon, 1-\varepsilon]$ to obtain $\hat{P}^k_t$. Let $\hat{\mathbf{\Theta}}^k_t = \mathrm{logit}((\hat{P}^k_t + \hat{P}^{k\,\mathrm{T}}_t)/2)$.
   (b) *(Sociality parameters).* Let $\hat{\boldsymbol{\delta}}_{k,t} = \arg\min_{\boldsymbol{\delta}_{k,t}} \|\hat{\mathbf{\Theta}}^k_t - \boldsymbol{\delta}_{k,t}\mathbf{1}_n^{\mathrm{T}} - \mathbf{1}_n \boldsymbol{\delta}_{k,t}^{\mathrm{T}}\|_F^2$. Calculate the residual matrix $\mathbf{E}^k_t = \hat{\mathbf{\Theta}}^k_t - \hat{\boldsymbol{\delta}}_{k,t}\mathbf{1}_n^{\mathrm{T}} - \mathbf{1}_n \hat{\boldsymbol{\delta}}_{k,t}^{\mathrm{T}}$.
2. *(Find a common subspace).* For $t \in \{1, \ldots, T\}$, set $\hat{\mathbf{X}}_t$ as the matrix containing the d leading left singular vectors of $(\mathbf{E}^1_t \mid \cdots \mid \mathbf{E}^K_t) \in \mathbb{R}^{n \times Kn}$.
3. *(Align $\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_T$).* Moving sequentially forward in time starting at $t = 2$, project $\hat{\mathbf{X}}_t$ to the locations that are closest to its previous location $\hat{\mathbf{X}}_{t-1}$ through a Procrustes rotation (Hoff et al., 2002).
4. *(Homophily matrices).* For $1 \leq k \leq K$, set $\hat{\Lambda}_k = \arg\min_{\Lambda_k} \sum_{t=1}^T \|\mathbf{E}^k_t - \hat{\mathbf{X}}_t \Lambda_k \hat{\mathbf{X}}_t^{\mathrm{T}}\|_F^2$.
5. *(Set layer r as the reference layer).* For $1 \leq k \neq r \leq K$, set $\hat{\Lambda}_k = \hat{\Lambda}_k |\hat{\Lambda}_r|^{-1}$ and $\hat{\mathbf{X}}_t = \hat{\mathbf{X}}_t |\hat{\Lambda}_r|^{1/2}$.
6. Output initial estimates $\boldsymbol{\mu}^i_t = \hat{\mathbf{X}}^i_t$, $\mu_{\delta^i_{k,t}} = \hat{\delta}^i_{k,t}$, $\boldsymbol{\mu}_{\boldsymbol{\lambda}_k} = \mathrm{diag}(\hat{\Lambda}_k)$, where $\mathrm{diag}(\cdot)$ extracts the diagonal of a matrix.

Algorithm 8: Initialization by universal singular value thresholding.

It remains to initialize the variances of the variational distributions and set the parameters of the priors. For the social trajectories, we initialized the variational distribution's variances and cross-covariances to $\sigma^2_{\delta^i_{k,t}} = 1$ and $\sigma^2_{\delta^i_{k,t,t+1}} = 1$, respectively. Furthermore, we placed broad priors on the state space parameters. To make the prior on τ^2_δ flat, one can set the shape and scale parameters of the inverse-gamma priors to $a_{\tau^2_\delta} = 2(2+\epsilon)$ and $b_{\tau^2_\delta} = 2(1+\epsilon)\mathbb{E}\left[\tau^2_\delta\right]$ for some small $\epsilon > 0$, respectively. We set $\epsilon = 0.05$ and $\mathbb{E}\left[\tau^2_\delta\right] = 10$. For σ^2_δ, we set $c_{\sigma^2_\delta} = 2$ and $d_{\sigma^2_\delta} = 2$. For the latent position's variational distributions, we set the variances and cross-covariances to the identity matrix I_d. We chose uninformative priors for the inverse-Wishart distribution by setting $\nu = 2 + d$ and $V = I_d$. For σ^2, we set $c_{\sigma^2} = 2$ and $d_{\sigma^2} = 2$. Lastly, for the reference layer, we set $\rho = 1/2$ and $\Sigma_{\boldsymbol{\lambda}_k} = 10 I_d$. In addition, we set the prior variance to $\sigma^2_\lambda = 4$. Lastly, we initialized the mean of the Pólya-gamma random variables, $\mu_{\omega^k_{ijt}}$, to zero.

Appendix F. The Mean-Field Algorithm

In this section, we derive the updating equations for the mean-field (MF) approximation

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega}) = \left[\prod_{h=1}^{d} q(\lambda_{1h})\right]\left[\prod_{k=2}^{K} q(\boldsymbol{\lambda}_k)\right]\left[\prod_{k=1}^{K}\prod_{i=1}^{n}\prod_{t=1}^{T} q(\delta^i_{k,t})\right]\left[\prod_{i=1}^{n}\prod_{t=1}^{T} q(\mathbf{X}^i_t)\right]\left[\prod_{k=1}^{K}\prod_{t=1}^{T}\prod_{j<i} q(\omega^k_{ijt})\right] \times q(\Psi_0) q(\sigma^2) q(\tau^2_\delta) q(\sigma^2_\delta). \tag{50}$$

This equation differs from structured mean-field approximation in Equation (3) used throughout the main text in that the variational posterior restricts the latent positions and sociality parameters to be independent across time points. In the remaining sections, we outline the updates for $q(\mathbf{X}^i_t)$ and $q(\delta^i_{k,t})$ under this mean-field approximation. The updates for the remaining variational distributions can be obtained from the previous updates outlined in Appendix C by setting the cross-covariances equal to zero, i.e., $\Sigma^i_{t,t+1} = 0$ and $\sigma^2_{\delta^i_{k,t,t+1}} = 0$. Note that when fitting this model, we used the same hyperparameter settings and initialization method as the dynamic multilayer eigenmodel described in Appendix E.

F.1 Updates for $q(\mathbf{X}^i_t)$

Throughout this section let $A_t, \mathbf{b}_t$, and C_t be defined as in Equation (42) of Proposition 12. Standard manipulations similar to the ones employed in Proposition 12 show that

- For $t = 1$,

$$\begin{aligned} q(\mathbf{X}^i_t) &\propto \exp\left\{\mathbb{E}_{-q(\mathbf{X}^i_t)}\left[-\frac{1}{2}(\mathbf{y}_t - A_t\mathbf{X}^i_t - \mathbf{b}_t)^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - A_t\mathbf{X}^i_t - \mathbf{b}_t)\right.\right. \\ &\qquad \left.\left. - \frac{1}{2}\mathbf{X}^{i\,\mathrm{T}}_t \Psi_0^{-1}\mathbf{X}^i_t + \frac{\|\mathbf{X}^i_t - \mathbf{X}^i_{t+1}\|_2^2}{2\sigma^2}\right]\right\} \\ &\propto \exp\left\{\left(\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle\frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}^i_{t+1}\right)^{\mathrm{T}} \mathbf{X}^i_t \right. \\ &\qquad \left. - \frac{1}{2}\operatorname{tr}\left(\left[\langle A_t^T C_t^{-1} A_t\rangle + \langle\Psi_0^{-1}\rangle + \left\langle\frac{1}{\sigma^2}\right\rangle\right]\mathbf{X}^i_t \mathbf{X}^{i\,\mathrm{T}}_t\right)\right\}, \end{aligned}$$

which we recognize as a Gaussian distribution with parameters

$$\Sigma_t^i = \left(\langle A_t^T C_t^{-1} A_t\rangle + \langle \Psi_0^{-1}\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle\right)^{-1} = \left(\Gamma_t^2 + \langle \Psi_0^{-1}\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle\right)^{-1},$$
$$\boldsymbol{\mu}_t^i = \Sigma_t^i \left[\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t+1}^i\right] = \Sigma_t^i \left[\Gamma_t^1 + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t+1}^i\right],$$

where Γ_t^1 and Γ_t^2 are defined in Equation (43) and Equation (44), respectively.

- For $1 < t < T$,

$$\begin{aligned} q(\mathbf{X}_t^i) &\propto \exp\left\{\mathbb{E}_{-q(\mathbf{X}_t^i)}\left[-\frac{1}{2}(\mathbf{y}_t - A_t\mathbf{X}_t^i - \mathbf{b}_t)^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - A_t\mathbf{X}_t^i - \mathbf{b}_t)\right.\right. \\ &\qquad \left.\left. - \frac{\|\mathbf{X}_t^i - \mathbf{X}_{t-1}^i\|_2^2}{2\sigma^2} - \frac{\|\mathbf{X}_t^i - \mathbf{X}_{t+1}^i\|_2^2}{2\sigma^2}\right]\right\} \\ &\propto \exp\left\{\left(\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t-1}^i + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t+1}^i\right)^{\mathrm{T}} \mathbf{X}_t^i \right. \\ &\qquad \left. - \frac{1}{2}\operatorname{tr}\left(\left[\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{2}{\sigma^2}\right\rangle\right] \mathbf{X}_t^i \mathbf{X}_t^{i\,\mathrm{T}}\right)\right\}, \end{aligned}$$

  which we recognize as a Gaussian distribution with parameters

$$\begin{aligned} \Sigma_t^i &= \left(\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{2}{\sigma^2}\right\rangle\right)^{-1} = \left(\Gamma_t^2 + \left\langle \frac{2}{\sigma^2}\right\rangle\right)^{-1}, \\ \boldsymbol{\mu}_t^i &= \Sigma_t^i \left[\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t-1}^i + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t+1}^i\right] \\ &= \Sigma_t^i \left[\Gamma_t^1 + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t-1}^i + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t+1}^i\right], \end{aligned}$$

  where Γ_t^1 and Γ_t^2 are defined in Equation (43) and Equation (44), respectively.

- For $t = T$:

$$\begin{aligned} q(\mathbf{X}_t^i) &\propto \exp\left\{\mathbb{E}_{-q(\mathbf{X}_t^i)}\left[-\frac{1}{2}(\mathbf{y}_t - A_t\mathbf{X}_t^i - \mathbf{b}_t)^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - A_t\mathbf{X}_t^i - \mathbf{b}_t)\right.\right. \\ &\qquad \left.\left. - \frac{\|\mathbf{X}_t^i - \mathbf{X}_{t-1}^i\|_2^2}{2\sigma^2}\right]\right\} \\ &\propto \exp\left\{\left(\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t-1}^i\right)^{\mathrm{T}} \mathbf{X}_t^i \right. \\ &\qquad \left. - \frac{1}{2}\operatorname{tr}\left(\left[\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle\right] \mathbf{X}_t^i \mathbf{X}_t^{i\,\mathrm{T}}\right)\right\}, \end{aligned}$$

  which we recognize as a Gaussian distribution with parameters

$$\Sigma_t^i = \left(\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle\right)^{-1} = \left(\Gamma_t^2 + \left\langle \frac{1}{\sigma^2}\right\rangle\right)^{-1},$$
$$\boldsymbol{\mu}_t^i = \Sigma_t^i \left[\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t-1}^i\right] = \Sigma_t^i \left[\Gamma_t^1 + \left\langle \frac{1}{\sigma^2}\right\rangle \boldsymbol{\mu}_{t-1}^i\right],$$

  where Γ_t^1 and Γ_t^2 are defined in Equation (43) and Equation (44), respectively.

F.2 Updates for $q(\delta^i_{k,t})$

Throughout this section let $A_t, \mathbf{b}_t$, and C_t be defined as in Equation (31) of Proposition 9. Standard manipulations similar to the ones outlined in Proposition 9 show that

- For $t = 1$,

$$\begin{aligned} q(\delta^i_{k,t}) &\propto \exp\left\{\mathbb{E}_{-q(\delta^i_{k,t})}\left[(\mathbf{y}_t - A_t\delta^i_{k,t} - \mathbf{b}_t)^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - A_t\delta^i_{k,t} - \mathbf{b}_t) - \frac{(\delta^i_{k,t})^2}{2\tau_\delta^2}\right.\right. \\ &\qquad \left.\left. - \frac{(\delta^i_{k,t} - \delta^i_{k,t+1})^2}{2\sigma_\delta}\right]\right\} \\ &\propto \exp\left\{\left(\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t+1}}\right)\delta^i_{k,t} \right. \\ &\qquad \left. - \frac{1}{2}\left[\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{1}{\tau_\delta^2}\right\rangle + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle\right](\delta^i_{k,t})^2\right\}, \end{aligned}$$

  which we recognize as a Gaussian distribution with parameters

$$\begin{aligned} \sigma^2_{\delta^i_{k,t}} &= \left(\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{1}{\tau_\delta^2}\right\rangle + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle\right)^{-1} = \left(\Gamma_t^2 + \left\langle \frac{1}{\tau_\delta^2}\right\rangle + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle\right)^{-1}, \\ \mu_{\delta^i_{k,t}} &= \sigma^2_{\delta^i_{k,t}}\left[\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t+1}}\right] = \sigma^2_{\delta^i_{k,t}}\left[\Gamma_t^1 + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t+1}}\right], \end{aligned}$$

  where Γ_t^1 and Γ_t^2 are defined in Equation (32) and Equation (33), respectively.

- For $1 < t < T$,

$$\begin{aligned} q(\delta^i_{k,t}) &\propto \exp\left\{\mathbb{E}_{-q(\delta^i_{k,t})}\left[(\mathbf{y}_t - A_t\delta^i_{k,t} - \mathbf{b}_t)^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - A_t\delta^i_{k,t} - \mathbf{b}_t)\right.\right. \\ &\qquad \left.\left. - \frac{(\delta^i_{k,t} - \delta^i_{k,t-1})^2}{2\sigma_\delta^2} - \frac{(\delta^i_{k,t} - \delta^i_{k,t+1})^2}{2\sigma_\delta}\right]\right\} \\ &\propto \exp\left\{\left(\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t-1}} + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t+1}}\right)\delta^i_{k,t} \right. \\ &\qquad \left. - \frac{1}{2}\left[\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{2}{\sigma_\delta^2}\right\rangle\right](\delta^i_{k,t})^2\right\}, \end{aligned}$$

  which we recognize as a Gaussian distribution with parameters

$$\begin{aligned} \sigma^2_{\delta^i_{k,t}} &= \left(\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{2}{\sigma_\delta^2}\right\rangle\right)^{-1} = \left(\Gamma_t^2 + \left\langle \frac{2}{\sigma_\delta^2}\right\rangle\right)^{-1}, \\ \mu_{\delta^i_{k,t}} &= \sigma^2_{\delta^i_{k,t}}\left[\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t-1}} + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t+1}}\right] \\ &= \sigma^2_{\delta^i_{k,t}}\left[\Gamma_t^1 + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t-1}} + \left\langle \frac{1}{\sigma_\delta^2}\right\rangle \mu_{\delta^i_{k,t+1}}\right], \end{aligned}$$

  where Γ_t^1 and Γ_t^2 are defined in Equation (32) and Equation (33), respectively.

- For $t = T$,

$$q(\delta^i_{k,t}) \propto \exp\left\{\mathbb{E}_{-q(\delta^i_{k,t})}\left[(\mathbf{y}_t - A_t\delta^i_{k,t} - \mathbf{b}_t)^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - A_t\delta^i_{k,t} - \mathbf{b}_t) - \frac{(\delta^i_{k,t} - \delta^i_{k,t-1})^2}{2\sigma^2_\delta}\right]\right\}$$
$$\propto \exp\left\{\left(\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma^2_\delta}\right\rangle \mu_{\delta^i_{k,t-1}}\right)\delta^i_{k,t} - \frac{1}{2}\left[\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{1}{\sigma^2_\delta}\right\rangle\right](\delta^i_{k,t})^2\right\},$$

which we recognize as a Gaussian distribution with parameters

$$\sigma^2_{\delta^i_{k,t}} = \left(\langle A_t^T C_t^{-1} A_t\rangle + \left\langle \frac{1}{\sigma^2_\delta}\right\rangle\right)^{-1} = \left(\Gamma^2_t + \left\langle \frac{1}{\sigma^2_\delta}\right\rangle\right)^{-1},$$
$$\mu_{\delta^i_{k,t}} = \sigma^2_{\delta^i_{k,t}}\left[\langle A_t^{\mathrm{T}} C_t^{-1}(\mathbf{y}_t - \mathbf{b}_t)\rangle + \left\langle \frac{1}{\sigma^2_\delta}\right\rangle \mu_{\delta^i_{k,t-1}}\right] = \sigma^2_{\delta^i_{k,t}}\left[\Gamma^1_t + \left\langle \frac{1}{\sigma^2_\delta}\right\rangle \mu_{\delta^i_{k,t-1}}\right],$$

where Γ^1_t and Γ^2_t are defined in Equation (32) and Equation (33), respectively.

Appendix G. Variational Inference for Static Multilayer Networks

For the special case that $T = 1$, the eigenmodel for dynamic multilayer networks serves as a model for static multilayer networks. In particular, for $T = 1$, the full Bayesian model becomes

$$Y^k_{ij1} \overset{\text{ind.}}{\sim} \text{Bernoulli}\left[\text{logit}^{-1}\left(\delta^i_{k,1} + \delta^j_{k,1} + \mathbf{X}_1^{i\,\mathrm{T}}\Lambda_{k1}\mathbf{X}^j_1\right)\right], \quad 1 \le j < i \le n, \quad 1 \le k \le K,$$
$$\lambda_{11,h} = 2u_h - 1, \qquad u_h \overset{\text{iid}}{\sim} \text{Bernoulli}(\rho), \qquad 1 \le h \le d,$$
$$\boldsymbol{\lambda}_{k1} \overset{\text{iid}}{\sim} N(\mathbf{0}_d, \sigma^2_\lambda I_d), \qquad 2 \le k \le K,$$
$$\delta^i_{k,1} \overset{\text{iid}}{\sim} N(0, \tau^2_\delta), \qquad 1 \le i \le n, \quad 1 \le k \le K,$$
$$\mathbf{X}^i_1 \overset{\text{iid}}{\sim} N(\mathbf{0}_d, \Psi_0), \qquad 1 \le i \le n,$$
$$\Psi_0 \sim \text{Wishart}^{-1}(\nu, V), \qquad \tau^2_\delta \sim \Gamma^{-1}(c_\delta, d_\delta),$$

where we have kept the subscripts indicating that $t = 1$ to make the connection with the dynamic model introduced in the main text explicit. In this case, the proposed variational posterior becomes

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega}) = \left[\prod_{h=1}^{d} q(\lambda_{1h})\right]\left[\prod_{k=2}^{K} q(\boldsymbol{\lambda}_k)\right]\left[\prod_{k=1}^{K}\prod_{i=1}^{n} q(\delta^i_{k,1})\right]\left[\prod_{i=1}^{n} q(\mathbf{X}^i_1)\right]\left[\prod_{k=1}^{K}\prod_{j<i} q(\omega^k_{ij1})\right] \times q(\Psi_0)q(\tau^2_\delta).$$

In the remaining sections, we outline the updates for $q(\mathbf{X}^i_1)$ and $q(\delta^i_{k,1})$ under this variatonal posterior. The updates for the remaining parameters can be obtained from the previous updates outlined in Appendix C by setting $T = 1$, and so are omitted. Note that when fitting this model, we used the same hyperparameter settings for σ^2_λ, c_δ, d_δ, ν, and V and initialization method (with $T = 1$) as the dynamic multilayer eigenmodel described in Appendix E.

G.1 Updates for $q(\mathbf{X}_1^i)$

Throughout this section let $A_t, \mathbf{b}_t$, and C_t be defined as in Equation (42) of Proposition 12. Standard manipulations similar to the ones employed in Proposition 12 show that

$$\begin{aligned} q(\mathbf{X}_1^i) &\propto \exp\left\{\mathbb{E}_{-q(\mathbf{X}_t^i)}\left[-\frac{1}{2}(\mathbf{y}_1 - A_1\mathbf{X}_1^i - \mathbf{b}_1)^{\mathrm{T}} C_1^{-1}(\mathbf{y}_1 - A_1\mathbf{X}_1^i - \mathbf{b}_1) - \frac{1}{2}\mathbf{X}_1^{i\,\mathrm{T}}\Psi_0^{-1}\mathbf{X}_1^i\right]\right\} \\ &\propto \exp\left\{\left(\langle A_1^{\mathrm{T}} C_1^{-1}(\mathbf{y}_1 - \mathbf{b}_1)\rangle\right)^{\mathrm{T}}\mathbf{X}_1^i - \frac{1}{2}\operatorname{tr}\left(\left[\langle A_1^T C_1^{-1} A_1\rangle + \langle\Psi_0^{-1}\rangle\right]\mathbf{X}_1^i\mathbf{X}_1^{i\,\mathrm{T}}\right)\right\}, \end{aligned}$$

which we recognize as a Gaussian distribution with parameters

$$\begin{aligned} \Sigma_1^i &= \left(\langle A_1^T C_1^{-1} A_1\rangle + \langle\Psi_0^{-1}\rangle\right)^{-1} = \left(\Gamma_1^2 + \langle\Psi_0^{-1}\rangle\right)^{-1}, \\ \boldsymbol{\mu}_1^i &= \Sigma_1^i\left(\langle A_1^{\mathrm{T}} C_1^{-1}(\mathbf{y}_1 - \mathbf{b}_1)\rangle\right) = \Sigma_1^i\Gamma_1^1, \end{aligned}$$

where Γ_1^1 and Γ_1^2 are defined in Equation (43) and Equation (44), respectively.

G.2 Updates for $q(\delta_{k,1}^i)$

Throughout this section let $A_t, \mathbf{b}_t$, and C_t be defined as in Equation (31) of Proposition 9. Standard manipulations similar to the ones outlined in Proposition 9 show that

$$\begin{aligned} q(\delta_{k,1}^i) &\propto \exp\left\{\mathbb{E}_{-q(\delta_{k,1}^i)}\left[-\frac{1}{2}(\mathbf{y}_1 - A_1\delta_{k,1}^i - \mathbf{b}_1)^{\mathrm{T}} C_1^{-1}(\mathbf{y}_1 - A_1\delta_{k,1}^i - \mathbf{b}_1) - \frac{(\delta_{k,1}^i)^2}{2\tau_\delta^2}\right]\right\} \\ &\propto \exp\left\{\langle A_1^{\mathrm{T}} C_1^{-1}(\mathbf{y}_1 - \mathbf{b}_1)\rangle\,\delta_{k,1}^i - \frac{1}{2}\left[\langle A_1^T C_1^{-1} A_1\rangle + \left\langle\frac{1}{\tau_\delta^2}\right\rangle\right](\delta_{k,1}^i)^2\right\}, \end{aligned}$$

which we recognize as a Gaussian distribution with parameters

$$\begin{aligned} \sigma^2_{\delta_{k,1}^i} &= \left(\langle A_1^T C_1^{-1} A_1\rangle + \left\langle\frac{1}{\tau_\delta^2}\right\rangle\right)^{-1} = \left(\Gamma_1^2 + \left\langle\frac{1}{\tau_\delta^2}\right\rangle\right)^{-1}, \\ \mu_{\delta_{k,1}^i} &= \sigma^2_{\delta_{k,1}^i}\langle A_1^{\mathrm{T}} C_1^{-1}(\mathbf{y}_1 - \mathbf{b}_1)\rangle = \sigma^2_{\delta_{k,1}^i}\Gamma_1^1, \end{aligned}$$

where Γ_1^1 and Γ_1^2 are defined in Equation (32) and Equation (33), respectively.

Appendix H. Evaluation of Information Criteria for Dimension Selection

In this simulation, we evaluated different information criterion's ability to select the dimension of the latent space. We compared the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the deviance information criterion (DIC) (Spiegelhalter et al., 2002), and the Watanabe-Akaike information criterion (Watanabe, 2010).

We generated 50 networks from the data-generating process outlined in Section 4 with $n = 100$, $K = 5$, and $T = 10$. We varied the magnitude of the latent variable's transitions from $\sigma \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$ and used a moderate amount of transition dependence with $\rho = 0.4$. Note that the true latent space dimension is $d = 2$ in all simulations. For

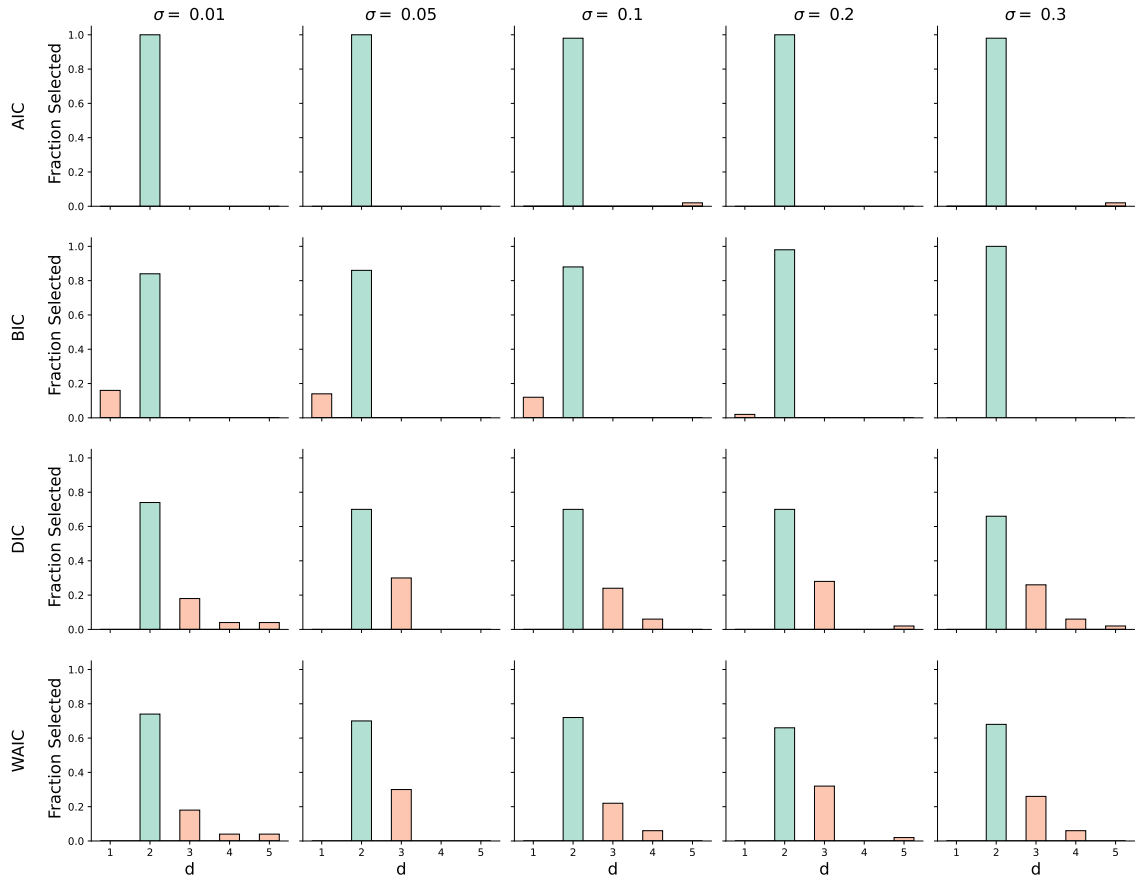


Figure 13: Fraction of the 50 simulations in which $\hat{d}$ equaled a particular value for AIC, BIC, DIC, and WAIC. The green bar corresponds to the true value of $d = 2$.

each network, we fit five models with d ranging from 1 to 5 using the SMF algorithm and calculated the value of each information criterion. The value of d that minimized the criterion was selected as the estimated dimension. Figure 13 displays the results for the four information criteria under consideration.

From these results, we conclude that all four information criteria selected the true latent space dimension the majority of the time. However, both DIC and WAIC overestimated the dimension more often than AIC and BIC. Furthermore, the performance of the AIC estimator is relatively insensitive to σ; however, the BIC estimator occasionally underestimated the dimension for small values of σ. As a result of this simulation study, we recommend using AIC to select the dimension of the latent space.

Appendix I. Additional Simulation: Decreasing σ as T Increases

In this simulation, we studied the SMF algorithm's estimation error as we simultaneously increased the number of time steps T and decreased the transition variance σ under the data

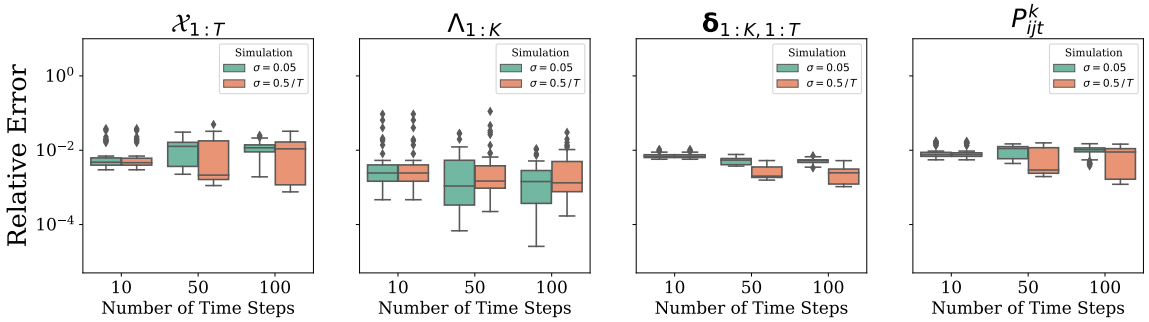


Figure 14: Comparison of the parameter recovery performance of the SMF algorithm for a simulation scenario with a fixed $\sigma = 0.05$ (green) and a decreasing $\sigma = 0.5/T$ (red). The boxplots show the distribution over 50 simulated networks with $n = 100$, $K = 5$, $T \in \{10, 50, 100\}$.

generating process outlined in Section 4. This simulation reflects the scenario of collecting a denser set of network snapshots over a fixed time interval. We varied the network sizes $(n, K, T) \in \{100\} \times \{5\} \times \{10, 50, 100\}$. We set $\sigma = 0.5/T$, so that σ decreased as T increased. We fixed $\rho = 0.4$ and sampled 50 independent networks for each network size.

Figure 14 displays the model parameters' estimation errors for each network size. The results are similar to the original scenario that fixed $\sigma = 0.05$. There is a noticeable improvement in the recovery of the sociality parameters under the new scenario; however, the recovery of the latent positions is relatively unchanged. In contrast, for a fixed network size (n, K, T), the simulations in Section 4.2 showed that parameter recovery improved as σ decreased. Overall, the estimation errors remain low under this scenario.

Appendix J. Additional Figures

In this section, we include the remaining figures for the data analyzed in the main text's real data applications (Section 5). These figures include the AIC selection plots, the remaining social trajectories, and the homophily coefficients for the primary school networks.

We start with the remaining figures from the analysis of the ICEWS data set. The results of the AIC selection are displayed in Figure 15. The model with $d = 3$ minimizes the AIC value. Figure 16 contains the social trajectories for the verbal cooperation networks, Figure 17 contains the social trajectories for the material cooperation networks, and Figure 18 contains the social trajectories for the verbal conflict networks. The shapes of the social trajectories are similar to the material conflict relation. A notable difference is that Ukraine increases its verbal cooperation, material cooperation, and verbal conflict sociality dramatically leading up to the Crimea Crisis.

Next, we present the remaining figures for the primary school face-to-face contact networks. Figure 19 displays the AIC values for models with different latent space dimensions. The model with $d = 2$ minimizes the AIC. Figure 20 and Figure 21 contain the actors' social trajectories on Thursday and Friday, respectively. We highlighted the trajectories

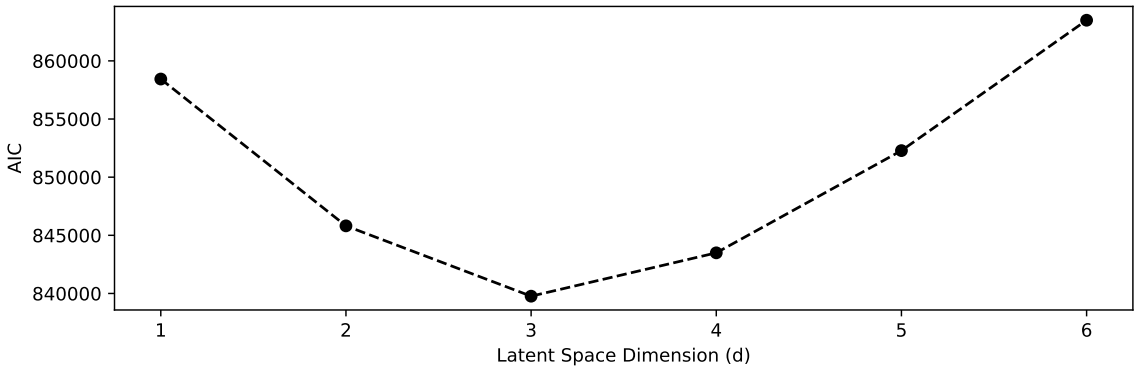


Figure 15: AIC values for the six models fit to the ICEWS data set with latent space dimensions ranging from 1 to 6.

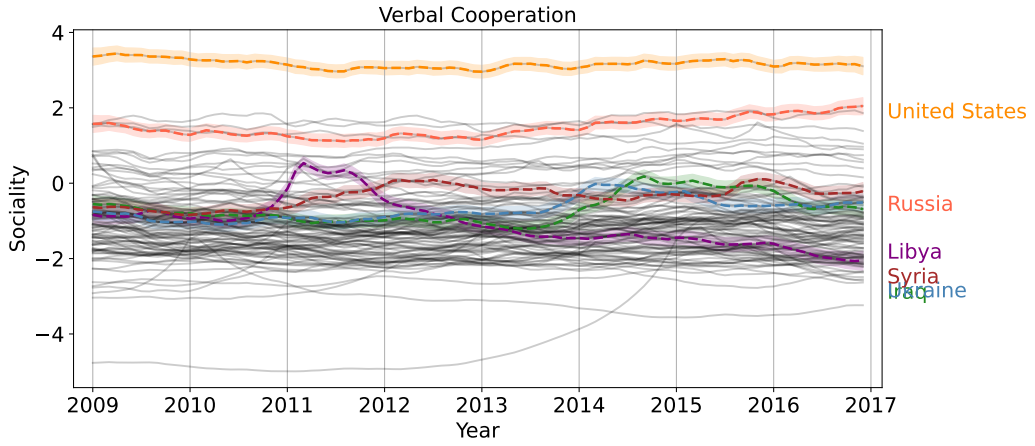


Figure 16: Posterior means of the verbal cooperation social trajectories for the ICEWS network. Select countries are highlighted in color with bands that represent 95% credible intervals. The remaining countries' social trajectories are displayed with gray curves.

of three actors. Actor 148 is a teacher, actor 195 is a student in class 3A, and actor 5 is a student in class 5B. Their social trajectories demonstrate three interesting longitudinal patterns. Actor 148, the teacher, is most socially active during class and least active during lunch. Conversely, actor 195 is most sociable during lunch and less sociable during class. Lastly, actor 5's social trajectory differs between the two days because he/she is absent on Friday. Next, Figure 22 contains the primary school network's homophily coefficients. All homophily coefficients are positive. Also, the magnitude of the homophily coefficients is smaller on Thursday than on Friday.

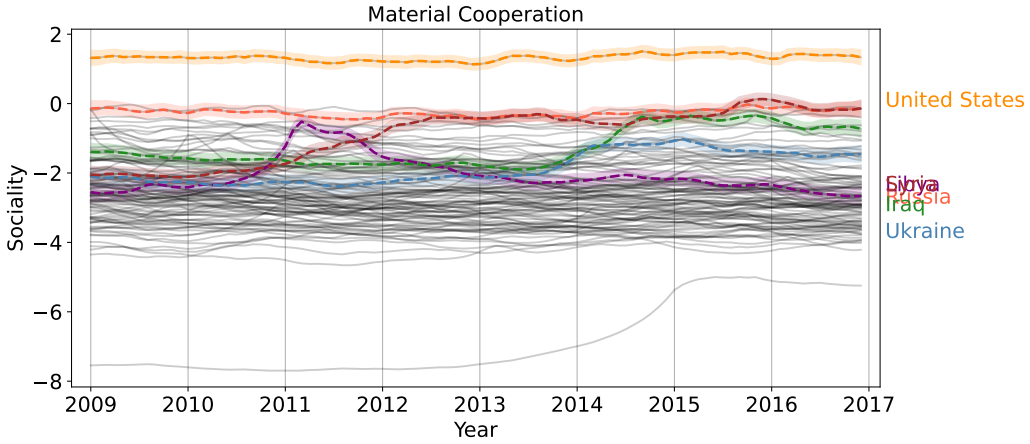


Figure 17: Posterior means of the material cooperation social trajectories for the ICEWS network. Select countries are highlighted in color with bands that represent 95% credible intervals. The remaining countries' social trajectories are displayed with gray curves.

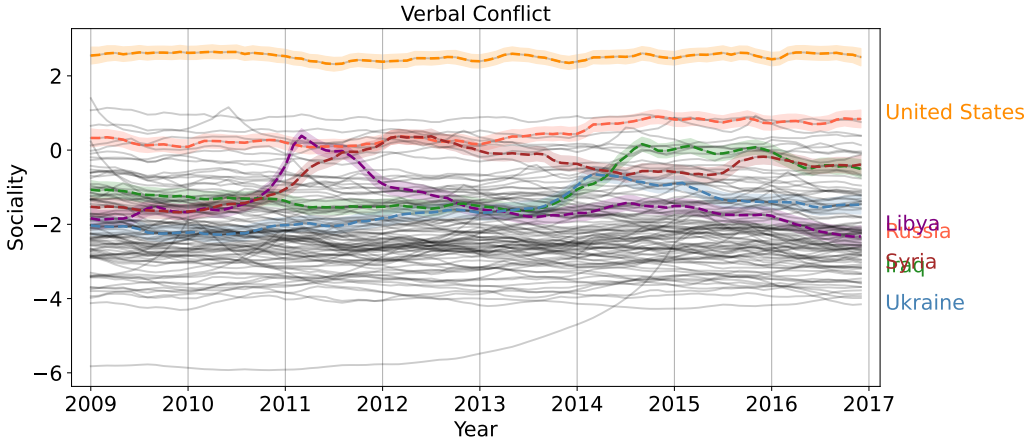


Figure 18: Posterior means of the verbal conflict social trajectories for the ICEWS network. Select countries are highlighted in color with bands that represent 95% credible intervals. The remaining countries' social trajectories are displayed with gray curves.

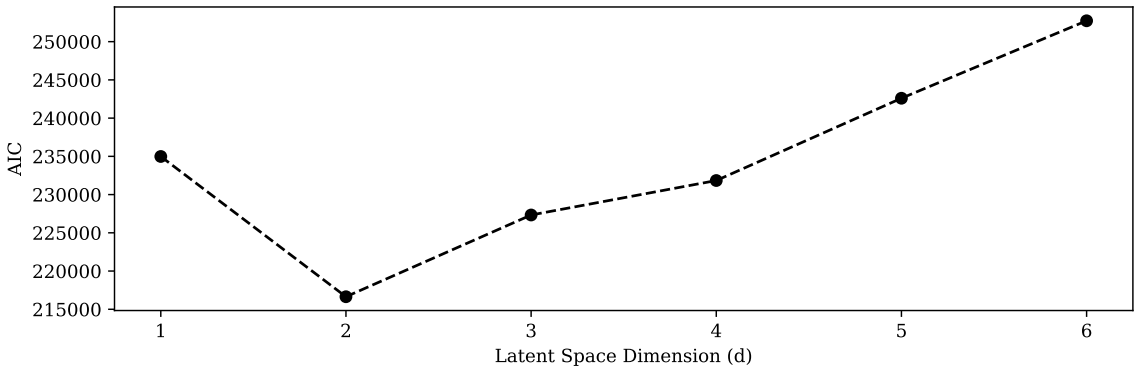


Figure 19: AIC values for the six models fit to the school contact networks data set with latent space dimensions ranging from 1 to 6.

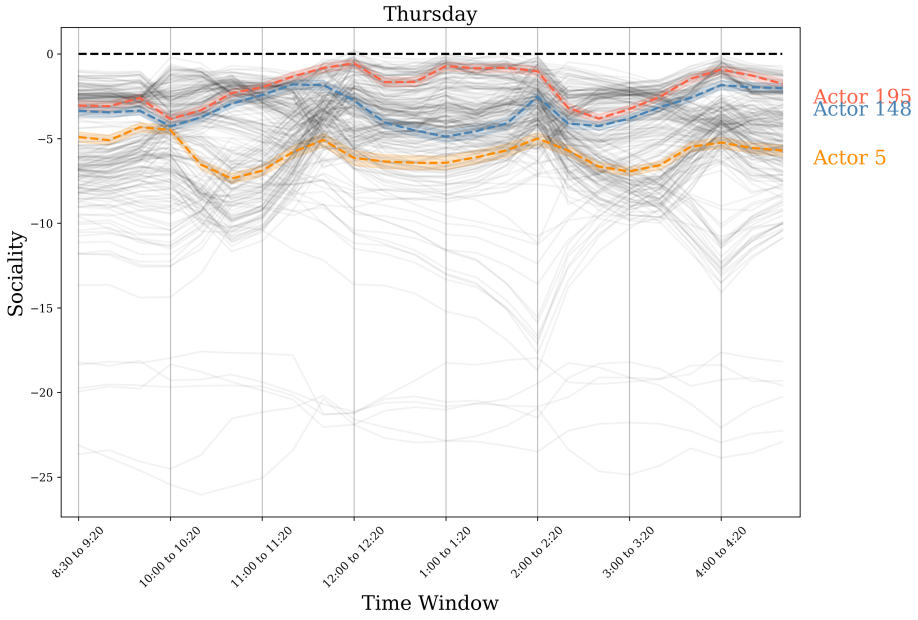


Figure 20: Posterior means of the social trajectories on Thursday for the primary school network. Select actors are highlighted in color with bands that represent 95% credible intervals. The remaining actors' social trajectories are displayed with gray curves.

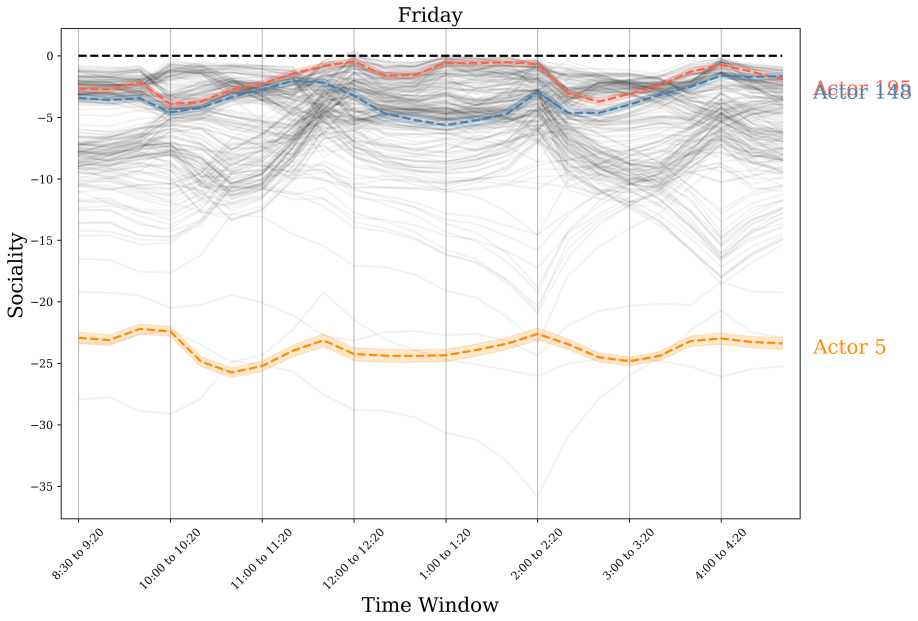


Figure 21: Posterior means of the social trajectories on Friday for the primary school network. Select actors are highlighted in color with bands that represent 95% credible intervals. The remaining actors' social trajectories are displayed with gray curves.

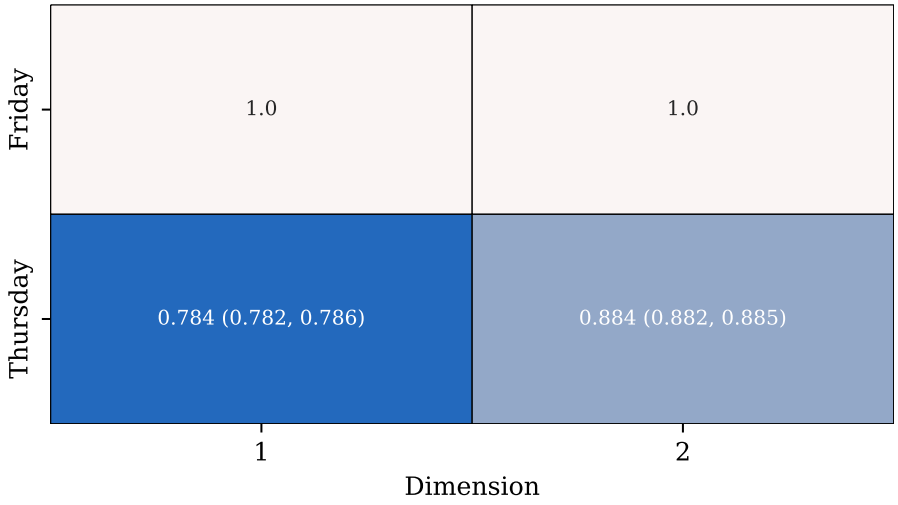


Figure 22: Heatmap of the homophily coefficients' posterior means for the primary school face-to-face contact networks. Each cell contains the coefficient's posterior mean and 95% credible interval. Blue cells indicate values less than the reference layer (Friday).

References

Emanuele Aliverti and Massimiliano Russo. Stratified stochastic variational inference for high-dimensional network factor model. *Journal of Computational and Graphical Statistics*, 31(2):502–511, 2022.

Roy M. Anderson and Robert M. May. *Infectious Diseases of Humans: Dynamics and Control*, volume 1. Oxford University Press, Oxford, 1991.

Håkan Andersson. Epidemics in a population with social structure. *Mathematical Biosciences*, 14(2):79–84, 1997.

Håkan Andersson. Limit theorems for a random graph epidemic model. *The Annals of Applied Probability*, 8(4):1331–1349, 1998.

David Barber and Silvia Chiappa. Unified inference for variational Bayesian linear Gaussian state-space models. In *Advances in Neural Information Processing Systems*, pages 81–88. 2007.

Matthew James Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, University of London, London, 2003.

David M. Blei, Alp Kucukelbir, and Jon D. Mcauliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I. del Genio, Jesus Gómez-Gardeñes, Miguel Romance, Irene Sendiña-Nadal, Zhen Wang, and Massimiliano Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. ICEWS Coded Event Data, 2015. URL `https://doi.org/10.7910/DVN/28075`.

Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE*, 5(7):e11596, 2010.

Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1):177–214, 2015.

Yasuko Chikuse. State space models on special manifold. *Journal of Multivariate Analysis*, 97(6):1284–1294, 2006.

Silvia D'Angelo, Thomas Brendan Murphy, and Marco Altó. Latent space modelling of multidimensional networks with applications to the exchange of votes in Eurovision song contest. *Annals of Applied Statistics*, 13(2):900–930, 2019.

Manlio De Domenico, Antonio Lima, Paul Mougel, and Mirco Musolesi. The anatomy of a scientific rumor. *Scientific Reports*, 3(2980), 2013.

Daniele Durante and David B. Dunson. Nonparametric Bayes dynamic modelling of relational data. *Biometrika*, 101(4):883–898, 2014.

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

Robert D. Duval and William R. Thompson. Reconsidering the aggregate relationship between size, economic development, and some types of foreign policy behavior. *American Journal of Political Science*, 24(3):511–525, 1980.

Deborah J. Gerner, Philip A. Schrodt, and Ömür Yilmaz. Conflict and mediation event observations (CAMEO): An event data framework for a post-Cold War world. In Jacob Bercovitch and Scott Sigmund Gartner, editors, *International Conflict Mediation: New Approaches and Findings*, chapter 13, pages 287–304. Routledge, New York, 2008.

Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010.

Isabella Gollini and Thomas Brendan Murphy. Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265, 2016.

Mark S. Handcock, Adrian E. Raftery, and Jeremy M. Tantrum. Model-based clustering of social networks. *Journal of the Royal Statistical Society, Series A*, 170(2):301–354, 2007.

Yanjun He and Peter D. Hoff. Multiplicative coevolution regression models for longitudinal networks and nodal attributes. *Social Networks*, 57:54–62, 2019.

Peter D. Hoff. Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469):286–295, 2005.

Peter D. Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems*, pages 657–664. 2008.

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

Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14(4):1303–1347, 2013.

Pavel N. Krivitsky, Mark S. Handcock, Adrian E. Raftery, and Peter D. Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31(3):204–213, 2009.

Jun S. Liu and Ying Nian Wu. Parameter expansion for data augmentation. *Journal of the American Statistical Association*, 94(448):1264–1274, 1999.

Yan Liu and Yuguo Chen. Variational inference for latent space models for dynamic networks. *Statistica Sinica*, 32:2147–2170, 2022.

Joshua Daniel Loyal. Replication code for "An eigenmodel for dynamic multilayer networks". `https://github.com/joshloyal/multidynet`, 2023.

Joshua Daniel Loyal and Yuguo Chen. Statistical network analysis: A review with applications to the Coronavirus Disease 2019 pandemic. *International Statistical Review*, 88(2): 419–440, 2020.

Zhuang Ma, Zongming Ma, and Hongsong Yuan. Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research*, 21(4):1–67, 2020.

Peter W. Macdonald, Elizaveta Levina, and Ji Zhu. Latent space models for multiplex networks with shared structure. *Biometrika*, 109(3):683–706, 2022.

Shahryar Minhas, Peter D. Hoff, and Michael D. Ward. Influence networks in international relations. *arXiv preprint arXiv:1706.09072v1*, 2017.

J. Møller, A.N. Pettitt, R. Reeves, and K.K. Berthelsen. An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, 93 (2):451–458, 2006.

Iain Murray, Zoubin Ghahramani, and David J.C. MacKay. MCMC for doubly-intractable distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intellligence*, pages 359–366. 2006.

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

Yuan Qi and Tommi S. Jaakkola. Parameter expanded variational Bayesian methods. In *Advances in Neural Information Processing Systems*, pages 1097–1104. 2006.

Riccardo Rastelli, Nial Friel, and Adrian E. Raftery. Properties of latent variable network models. *Network Science*, 4(4):407–432, 2016.

Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. *Journal of the Royal Statistical Society Series B*, 84(4):1446–1473, 2022.

Michael Salter-Townshend and Tyler H. McCormick. Latent space models for multiview network data. *Annals of Applied Statistics*, 11(3):1217–1244, 2017.

Michael Salter-Townshend and Thomas Brendan Murphy. Variational Bayesian inference for the latent position clustering model for network data. *Computational Statistics and Data Analysis*, 57(1):661–671, 2013.

Purnamrita Sarkar and Andrew W. Moore. Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*, pages 1145–1152, 2005.

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

Daniel K. Sewell and Yuguo Chen. Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, 12(2):351–377, 2017.

Tom A.B. Snijders, Alessandro Lomi, and Vanina Jasmine Torió. 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, 2013.

David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64(4):583–639, 2002.

Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Phillippe Vanhems. High-resolution measurements of face-to-face contact patterns in primary school. *PLoS ONE*, 6(8):e23176, 2011.

David A. van Dyk and Xiao-Li Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50, 2001.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, 164(10):936–944, 2006.

Bo Wang and D. M. Titterington. Lack of consistency of mean field and variational bayes approximations for state space models. *Neural Processing Letters*, 20:151–170, 2004.

Lu Wang, Zhengwu Zhang, and David Dunson. Common and individual structure of brain networks. *Annals of Applied Statistics*, 13(1):85–112, 2019.

Sumio Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:867–897, 2010.

Emilio Zagheni, Francesco C. Billari, Piero Manfredi, Alessia Melegaro, Joel Mossong, and W. John Edmunds. Using time-use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*, 168(9):1082–1090, 2008.

Xuefei Zhang, Songkai Xue, and Ji Zhu. A flexible latent space model for multilayer networks. In *Proceedings of the International Conference on Machine Learning*, pages 8546–8555, 2020.