Bayesian Learning of Dynamic Multilayer Networks

Daniele Durante   DURANTE@STAT.UNIPD.IT
Department of Statistical Sciences
University of Padova
Padova, 35121, Italy

Nabanita Mukherjee   NM168@DUKE.EDU
Department of Statistical Science
Duke University
Durham, NC 27708-0251, USA

Rebecca C. Steorts   RCS46@STAT.DUKE.EDU
Departments of Statistical Science and Computer Science
Duke University
Durham, NC 27708-0251, USA

Editor: Amos Storkey

Abstract

A plethora of networks is being collected in a growing number of fields, including disease transmission, international relations, social interactions, and others. As data streams continue to grow, the complexity associated with these highly multidimensional connectivity data presents novel challenges. In this paper, we focus on the time-varying interconnections among a set of actors in multiple contexts, called layers. Current literature lacks flexible statistical models for dynamic multilayer networks, which can enhance quality in inference and prediction by efficiently borrowing information within each network, across time, and between layers. Motivated by this gap, we develop a Bayesian nonparametric model leveraging latent space representations. Our formulation characterizes the edge probabilities as a function of shared and layer-specific actors positions in a latent space, with these positions changing in time via Gaussian processes. This representation facilitates dimensionality reduction and incorporates different sources of information in the observed data. In addition, we obtain tractable procedures for posterior computation, inference, and prediction. We provide theoretical results on the flexibility of our model. Our methods are tested on simulations and infection studies monitoring dynamic face-to-face contacts among individuals in multiple days, where we perform better than current methods in inference and prediction.

Keywords: Dynamic multilayer network, edge prediction, face-to-face contact network, Gaussian process, latent space model

1. Introduction

Data on social interaction processes are rapidly becoming highly multidimensional, increasingly complex and inherently dynamic, providing a natural venue for machine learning as seen in applications such as disease transmission (Marathe and Ramakrishnan, 2013), international relations (Schein et al., 2016), social interactions (Guo et al., 2015), and others.

While modeling of a single network observation is still an active area of research, the increasing availability of multidimensional networks from World Wide Web architectures (Leetaru and Schrodt, 2013), telecommunication infrastructures (Blondel et al., 2015) and human sensing devices (Cattuto et al., 2010), have motivated more flexible statistical models. This endeavor is particularly relevant in the dynamic multilayer network field, providing information on the time-varying connectivity patterns among a set of actors in different contexts. Notable examples include dynamic relationships between individuals according to multiple forms of social interactions (Snijders et al., 2013), and time-varying connectivity structures among countries based on different types of international relations (Hoff, 2015).

In modeling these highly complex networks, it is of paramount interest to learn the wiring processes underlying the observed data and to infer differences in network structures across layers and times. Improved estimation of the data's generating mechanism can refine the understanding of social processes and enhance the quality in prediction of future networks. In order to successfully accomplish these goals, it is important to define statistical models which can incorporate the different sources of information in the observed data, without affecting flexibility. However, current literature lacks similar methods, to our knowledge.

Motivated by this gap, we develop a Bayesian nonparametric model for dynamic multilayer networks which efficiently incorporates dependence within each network, across time, and between the different layers, while preserving flexibility. Our formulation borrows network information by defining the edge probabilities as a function of pairwise similarities between actors in a latent space. To share information among layers without affecting flexibility in characterizing layer-specific structures, we force a subset of the latent coordinates of each actor to be common across layers, and let the remaining coordinates to vary between layers. Finally, we accommodate network dynamics by allowing these actors coordinates to change in time, and incorporate time information by modeling the dynamic actors coordinates via Gaussian processes (e.g., Rasmussen and Williams, 2006). Our model is tractable and has a theoretical justification, while providing simple procedures for inference and prediction. In addition, we find that our procedures outperform current methods in inference and out-of-sample prediction on both simulated and real data.

The paper proceeds as follows. Section 1.1 reviews recent contributions in the literature, and Section 1.2 provides a motivational example of dynamic multilayer networks. Section 2 describes our Bayesian latent space model, while Section 3 discusses its properties. Methods for posterior computation and prediction are provided in Section 4. Section 5 highlights the performance gains of our model compared to other possible competitors in a simulation study, whereas Section 6 illustrates on infection studies, monitoring face-to-face contacts, our success in inference and prediction, compared to possible competitors. Future directions of research and additional fields of application are discussed in Section 7.

1.1 Related Work

Current literature for multidimensional network data considers settings in which the multiple networks are either dynamic or multilayer. Statistical modeling of dynamic networks has focused on stochastic processes which are designed to borrow information between edges and across time (e.g., Hanneke et al., 2010; Snijders et al., 2010; Xing et al., 2010; Yang et al., 2011; Durante and Dunson, 2014; Crane, 2016), whereas inference for multilayer net-

works has motivated formulations which can suitably induce dependence between edges and across the different types of relationships—characterizing the multiple layers (e.g., Gollini and Murphy, 2016; Han et al., 2015; Heaney, 2014). These contributions have generalized exponential random graph models (Holland and Leinhardt, 1981; Frank and Strauss, 1986) and latent variables models (Nowicki and Snijders, 2001; Hoff et al., 2002; Airoldi et al., 2008) for a single network to allow inference in multidimensional frameworks, when the multiple networks arise either from dynamic or multilayer studies. These methods are valuable building blocks for more flexible models, but fall far short of the goal of providing efficient procedures in more complex settings when the networks are both dynamic and multilayer.

The routine collection of dynamic multilayer networks is a recent development and statistical modeling of such data is still in its infancy. For example, Lee and Monge (2011) considered a generalization of exponential random graph models for multilayer networks, but performed a separate analysis for each time point. Oselio et al. (2014) focused instead on a dynamic stochastic block model which borrows information across time and within each network, but forces the underlying block structures to be shared between layers. Dynamic multilayer networks are complex objects combining homogenous structures shared between actors, layers, and smoothly evolving across time, with layer-specific patterns and across-actor heterogeneity. Due to this, any procedure that fails to incorporate the different sources of information in the observed data (e.g., Lee and Monge, 2011) is expected to lose efficiency, whereas models focusing on shared patterns (e.g., Oselio et al., 2014) may lack flexibility.

More general formulations are the multilayer stochastic actor-oriented model (Snijders et al., 2013) and the multilinear tensor regression (Hoff, 2015). Snijders et al. (2013) allowed for dynamic inference on network properties within and between layers, but failed to incorporate across-actor heterogeneity. This may lead to a lack of flexibility in prediction. Hoff (2015) considered autoregressive models with the vector of parameters having a tensor factorization representation. This formulation allows for across-actor heterogeneity, but forces the model parameters to be constant across time. In addition, the parameterization of the interdependence between layers relies on homogeneity assumptions. Consistent with these methods, our representation incorporates the different types of dependencies in the observed data, but crucially preserves flexibility to avoid restrictive homogeneity assumptions.

1.2 Motivating Application

Our motivation is drawn from epidemiologic studies monitoring hourly face-to-face contacts among individuals in a rural area of Kenya, during three consecutive days. Data are available from the human sensing platform SocioPatterns (`http://www.sociopatterns.org`) and have been collected using wearable devices that exchange low-power radio packets when two individuals are located within a sufficiently close distance to generate a potential occasion of contagion. Leveraging this technology it is possible to measure for each hour in the three consecutive days which pairs of actors had a face-to-face proximity contact. These information are fundamental to monitor the spread of diseases and learn future patterns. Refer also to Cattuto et al. (2010) for a detailed description of the data collection technology, and Kiti et al. (2016) for a more in-depth discussion of our motivating application.

As illustrated in Figure 1, the observed data can be structured as a dynamic multilayer network characterized by a sequence of $V \times V$ symmetric adjacency matrices $Y_{t_i}^{(k)}$ collected

3
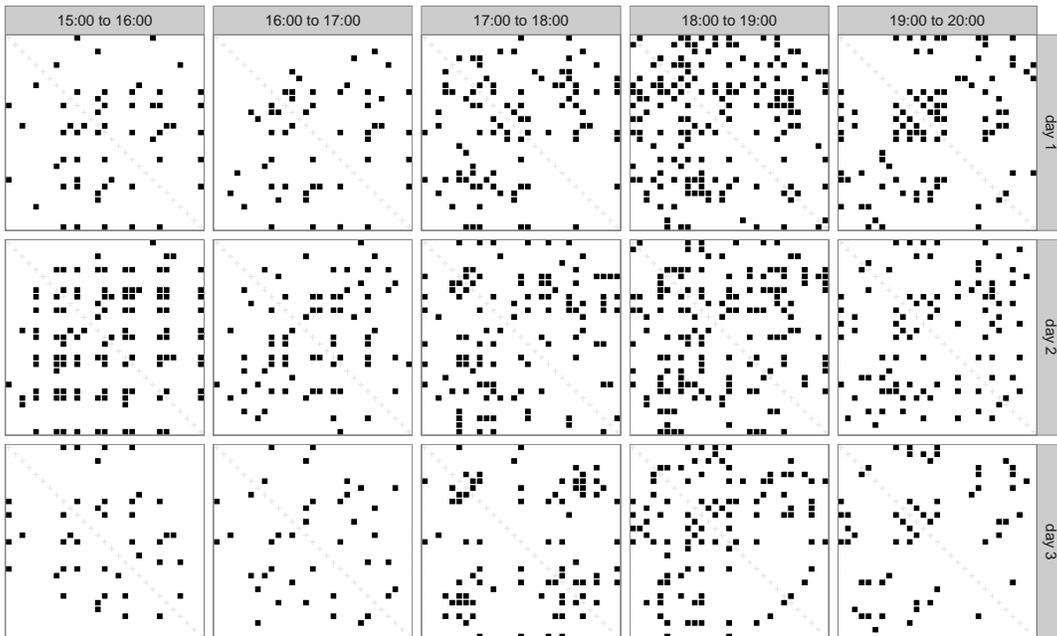
Figure 1: For five consecutive hours in the three days, adjacency matrices $Y_{t_i}^{(k)}$ representing the observed contact networks. Black refers to an edge, and white to a non-edge.

on a common time grid $t_1, \ldots, t_n$ in different days $k = 1, \ldots, K$. The matrices have entries $Y_{t_i[vu]}^{(k)} = Y_{t_i[uv]}^{(k)} = 1$ if actors v and u interact at time t_i in day k, and $Y_{t_i[vu]}^{(k)} = Y_{t_i[uv]}^{(k)} = 0$ otherwise, for every $v = 2, \ldots, V$ and $u = 1, \ldots, v - 1$. Dynamic multilayer proximity data provide key information on infections spread (Cauchemez et al., 2011), facilitating the design of outbreak prevention policies (Vanhems et al., 2013). The importance of this endeavor has motivated a wide variety of descriptive studies summarizing the information provided by face-to-face proximity data in several environments, including hospitals (Vanhems et al., 2013), schools (Stehlé et al., 2011) and households (Kiti et al., 2016).

Although the above analyses provide key insights on the connectivity processes underlying the observed proximity data, explicitly accounting for variability in network structures via carefully tailored statistical models can lead to improved learning of the connectivity patterns and properties, while providing methodologies for formal inference in the network framework, including estimation techniques, uncertainty quantification and prediction. In Section 6, we apply our methodology to this motivating application, showing how a careful statistical model for dynamic multilayer network data can provide substantial performance gains in learning underlying social processes and predicting future contact patterns.

2. Dynamic Multilayer Latent Space Model

Let $Y_{t_i}^{(k)}$ denote the $V \times V$ adjacency matrix characterizing the undirected network observed at time $t_i \in \Re^+$ in layer k, for any $t_i = t_1, \ldots, t_n$ and $k = 1, \ldots, K$. Each matrix has binary elements $Y_{t_i[vu]}^{(k)} = Y_{t_i[uv]}^{(k)} \in \{0, 1\}$ measuring the absence or presence of an edge among actors

4

Type of information	Illustrative example
Network information	If individual v had a face-to-face contact with both actors u and w at time t_i in day k, this information may be relevant to learn the face-to-face contact behavior between u and w at time t_i in day k.
Layer information	If individuals v and u had a face-to-face contact at time t_i in day k, this information may be relevant to learn the contact behavior between v and u at the same time t_i in other days.
Time information	If individuals v and u had a face-to-face contact at time t_i in day k, this information may be relevant to learn the contact behavior between v and u at the next time t_{i+1} in the same day.

Table 1: Relevant sources of information characterizing a dynamic multilayer network.

v and u at time t_i in layer k, for every $v = 2, \ldots, V$ and $u = 1, \ldots, v-1$. When modeling the data $Y = \{Y_{t_i}^{(k)} : t_i = t_1, \ldots, t_n, \ k = 1, \ldots, K\}$, we assume that Y is a realization, on a finite time grid, of the stochastic process $\mathcal{Y} = \{\mathcal{Y}^{(k)}(t) : t \in \mathbb{T} \subset \Re^+, \ k = 1, \ldots, K\}$, and seek a provably flexible representation for the probabilistic generative mechanism associated with the stochastic process $\mathcal{Y}$. Note that, as each random adjacency matrix $\mathcal{Y}^{(k)}(t)$ is symmetric and the diagonal elements—denoting self-relations—are not of interest, it is sufficient to model its lower triangular elements $\{\mathcal{Y}_{vu}^{(k)}(t) : v = 2, \ldots, V, \ u = 1, \ldots, v - 1\}$.

One major modeling objective is to carefully incorporate dependence among edges, between layers and across time, without affecting flexibility. Recalling the motivating application in Section 1.2 and Figure 1, it is reasonable to expect three main sources of information in the dynamic multilayer face-to-face contact data, summarized in Table 1. Incorporating such information can substantially improve the quality of inference and prediction, while facilitating dimensionality reduction. However, in reducing dimensionality and borrowing of information, it is important to avoid restrictive formulations that lead to inadequate characterizations of dynamic patterns, layer-specific structures, and across-actor heterogeneity.

We accomplish the aforementioned goals via a flexible dynamic latent bilinear model, combining shared and layer-specific actors coordinates which are allowed to change in time via Gaussian process priors. Section 2.1 describes our model formulation with a focus on sharing network and layer information, whereas Section 2.2 clarifies how time information is effectively incorporated via Gaussian processes.

2.1 Dynamic Bilinear Model with Shared and Layer-Specific Coordinates

In order to incorporate network and layer information, while preserving flexibility, we adapt latent bilinear models for a single network (Hoff, 2005) to our dynamic multilayer framework. In particular, we characterize the edges as conditionally independent Bernoulli random variables given their corresponding edge probabilities, and borrow information within each network by defining these probabilities as a function of the actors coordinates in a latent space. In order to incorporate dependence among the different layers, we force a subset of these coordinates to be common across layers, and allow the remaining coordinates to change

between layers. This last choice facilitates a flexible characterization of more irregular events and incorporates differences between layers in specific contact patterns.

Consistent with the aforementioned statistical model, and letting $\pi_{vu}^{(k)}(t) \in (0,1)$ denote the probability of an edge between actors v and u at time $t \in \mathbb{T}$ in layer k, we define

$$\mathcal{Y}_{vu}^{(k)}(t) \mid \pi_{vu}^{(k)}(t) \sim \mathrm{Bern}\{\pi_{vu}^{(k)}(t)\}, \tag{1}$$

independently for every time $t \in \mathbb{T}$, layer $k = 1, \ldots, K$ and pair of actors $[vu]$, $v = 2, \ldots, V$, $u = 1, \ldots, v - 1$. To incorporate our multilayer bilinear representation having shared and layer-specific latent coordinates, we express the log-odds of each edge probability as

$$\mathrm{logit}\{\pi_{vu}^{(k)}(t)\} = \mu(t) + \sum_{r=1}^{R} \bar{x}_{vr}(t)\bar{x}_{ur}(t) + \sum_{h=1}^{H} x_{vh}^{(k)}(t)x_{uh}^{(k)}(t), \tag{2}$$

where $\bar{x}_{vr}(t) \in \Re$ is the rth coordinate of actor v at time t shared across the different layers, whereas $x_{vh}^{(k)}(t) \in \Re$ denotes the hth coordinate of actor v at time t specific to layer k, for every $t \in \mathbb{T}$, $k = 1, \ldots, K$, $v = 1, \ldots, V$, $r = 1, \ldots, R$ and $h = 1, \ldots, H$. Finally, $\mu(t) \in \Re$ represents a time-varying baseline parameter centering the log-odds processes to improve computational and mixing performance.

In equation (2) the probability of an edge between actors v and u at time $t \in \mathbb{T}$ in layer k, increases with $\sum_{r=1}^{R} \bar{x}_{vr}(t)\bar{x}_{ur}(t)$ and $\sum_{h=1}^{H} x_{vh}^{(k)}(t)x_{uh}^{(k)}(t)$. Note that $\sum_{r=1}^{R} \bar{x}_{vr}(t)\bar{x}_{ur}(t) \in \Re$ characterizes a similarity measure between actors v and u at time t common to all the layers, whereas $\sum_{h=1}^{H} x_{vh}^{(k)}(t)x_{uh}^{(k)}(t) \in \Re$ defines a layer-specific deviation from this shared similarity, which enhances flexibility in modeling network structures specific to layer k. These similarity measures are defined as the dot product of shared and layer-specific actors coordinates in a latent space, allowing actors with coordinates in the same direction to have an higher chance of contact than actors with coordinates in opposite directions.

The dot product characterization in equation (2) is in the same spirit as the factorization in Hoff (2005) for a single network, and allows dimensionality reduction from $n \times K \times V \times (V-1)/2$ time-varying log-odds to $n \times \{1 + V \times (R + H \times K)\}$ dynamic latent coordinates, where typically $R \ll V$ and $H \ll V$. Although it is possible to consider other concepts of similarity when relating the latent coordinates to the log-odds (e.g., Hoff et al., 2002), the dot product representation facilitates the definition of simple algorithms for posterior inference and has been shown to efficiently borrow information within the network, while accommodating topological structures of interest such as homophily, transitivity, and others.

2.1.1 INTERPRETATION AND IDENTIFIABILITY

Recall the motivating application described in Section 1.2, our representation has an intuitive interpretation. In fact, the latent coordinates of each actor may represent his propensity towards $R + H$ latent interests or tasks. According to the factorization in equation (2), actors having propensities for the different interests in the same directions are expected to be more similar than actors with interests in opposite directions. Therefore, these actors are more likely to interact according to equation (1). Part of these actors interests or tasks may be associated with a common routine or constrained by daytime schedules. Therefore, the actors propensities towards these interests are expected to remain constant across days

providing a subset of R shared coordinates. Finally, these propensities are allowed to change across time inducing dynamic variations in the contacts within each day.

Although inference for the latent coordinates is potentially of interest, the factorization in equation (2) is not unique. In fact, it is easy to show that there exist many different latent coordinates values leading to the same collection of edge probabilities under the factorization in equation (2). However, such an overcomplete representation does not lead to identifiability issues if inference focuses on identified functionals of the latent coordinates, and has beneficial effects in terms of posterior computation and borrowing of information (Bhattacharya and Dunson, 2011; Ghosh and Dunson, 2009). Our focus is not on estimating the latent coordinates, but on leveraging the factorization in equation (2) to enhance efficiency in learning of identified functionals of the edge probabilities and improve prediction of future networks. Therefore, we avoid identifiability constraints on the latent coordinates in equation (2) as they are not required to ensure identifiability of the edge probabilities.

2.2 Gaussian Process Priors for the Time-Varying Latent Coordinates

Equation (2) facilitates borrowing of information among edges and between layers. This is obtained by leveraging the shared dependence on a common set of latent coordinates. We additionally incorporate across-layer heterogeneity via a set of layer-specific latent positions. Moreover, these coordinates are allowed to vary across time in order to accommodate flexible dynamic variations in the multilayer network structure. While it is possible to estimate the edge probabilities separately for every time point, this approach may be suboptimal in ruling out relevant temporal dependence. As seen in Figure 1, it is reasonable to expect a degree of dependence between the contact networks observed at close times, with this dependence decreasing with the time lag. Therefore, a formulation sharing information across time may provide substantial benefits in terms of efficiency, uncertainty propagation, and prediction.

Motivated by dynamic modeling of a single network, Durante and Dunson (2014) addressed a related goal by considering Gaussian process priors for the actors latent coordinates, with an additional shrinkage effect to facilitate automatic adaptation of the latent space dimensions and avoid overfitting. In the spirit of their proposed methods, we define

$$\bar{x}_{vr}(\cdot) \sim \mathrm{GP}(0, \tau_r^{-1} c_{\bar{x}}), \qquad \text{with} \quad c_{\bar{x}}(t_i, t_j) = \exp\{-\kappa_{\bar{x}}(t_i - t_j)^2\}, \kappa_{\bar{x}} > 0, \qquad (3)$$

independently for $v = 1, \ldots, V$, $r = 1, \ldots, R$ and

$$x_{vh}^{(k)}(\cdot) \sim \mathrm{GP}(0, \tau_h^{(k)-1} c_x), \quad \text{with} \quad c_x(t_i, t_j) = \exp\{-\kappa_x(t_i - t_j)^2\}, \kappa_x > 0, \qquad (4)$$

independently for $v = 1, \ldots, V$, $h = 1, \ldots, H$ and $k = 1, \ldots, K$. In the prior specification given in equations (3)–(4), the quantities $c_{\bar{x}}(t_i, t_j)$ and $c_x(t_i, t_j)$ denote the squared exponential correlation functions of the Gaussian processes for the shared and layer-specific latent coordinates, respectively. The quantities $\tau_1^{-1}, \ldots, \tau_R^{-1}$ and $\tau_1^{(k)-1}, \ldots, \tau_H^{(k)-1}$, for each $k = 1, \ldots, K$, are instead positive shrinkage parameters controlling the concentration of the latent coordinates around the zero constant mean function. Focusing on the prior for the trajectories of the rth shared coordinates $\bar{x}_{vr}(\cdot)$, $v = 1, \ldots, V$, a value of τ_r^{-1} close to zero forces these trajectories to be concentrated around the zero constant mean function, reducing the effect of the rth shared dimension in defining the edge probabilities in equation (2).

In order to penalize highly parameterized representations and facilitate adaptive deletion of unnecessary dimensions, the above shrinkage effects are designed to be increasingly strong as the latent coordinates indices $r = 1, \ldots, R$ and $h = 1, \ldots, H$ increase. This goal is accomplished by considering the following multiplicative inverse gamma priors (Bhattacharya and Dunson, 2011) for the shrinkage parameters:

$$\frac{1}{\tau_r} = \prod_{m=1}^{r} \frac{1}{\delta_m}, \quad r = 1, \ldots, R, \text{ with } \delta_1 \sim \text{Ga}(a_1, 1), \ \delta_{m \geq 2} \sim \text{Ga}(a_2, 1), \tag{5}$$

$$\frac{1}{\tau_h^{(k)}} = \prod_{l=1}^{h} \frac{1}{\delta_l^{(k)}}, \quad h = 1, \ldots, H, \ k = 1, \ldots, K, \text{ with } \delta_1^{(k)} \sim \text{Ga}(a_1, 1), \ \delta_{l \geq 2}^{(k)} \sim \text{Ga}(a_2, 1). \tag{6}$$

According to Bhattacharya and Dunson (2011), the multiplicative inverse gamma prior in equations (5)–(6) induces prior distributions for the shrinkage parameters with a cumulative shrinkage effect. In particular, these priors are increasingly concentrated close to zero as the indices $r = 1, \ldots, R$ and $h = 1, \ldots, H$ increase, for appropriate $a_1 \in \Re^+$ and $a_2 \in \Re^+$, facilitating adaptive dimensionality reduction and reducing overfitting issues. Refer also to Durante (2017) for an additional discussion on the multiplicative inverse gamma prior and its shrinkage properties. To conclude the prior specification, we let $\mu(\cdot) \sim \text{GP}(0, c_\mu)$, with $c_\mu(t_i, t_j) = \exp\{-\kappa_\mu(t_i - t_j)^2\}$, $\kappa_\mu > 0$.

The Gaussian process prior provides an accurate choice in our setting which incorporates time dependence, allowing the amount of information shared between networks to increase as the time lag decreases. Beside improving efficiency, the Gaussian process prior can deal with multilayer networks observed at potentially unequally spaced time grids, and is closely related to the multivariate Gaussian distribution, providing substantial benefits in terms of computational tractability and interpretability. In fact, following Rasmussen and Williams (2006), equations (3)–(4) imply the following prior for the shared and layer-specific latent coordinates at the finite time grid $t_1, \ldots, t_n$ on which data are observed:

$$\{\bar{x}_{vr}(t_1), \ldots, \bar{x}_{vr}(t_n)\}^\intercal \sim \text{N}_n(0, \tau_r^{-1} \Sigma_{\bar{x}}), \tag{7}$$

independently for $v = 1, \ldots, V$, $r = 1, \ldots, R$ and

$$\{x_{vh}^{(k)}(t_1), \ldots, x_{vh}^{(k)}(t_n)\}^\intercal \sim \text{N}_n(0, \tau_h^{(k)-1} \Sigma_x), \tag{8}$$

independently for $v = 1, \ldots, V$, $h = 1, \ldots, H$, and $k = 1, \ldots, K$. In equations (7)–(8), the $n \times n$ variance and covariance matrices $\Sigma_{\bar{x}}$ and Σ_x have elements $\Sigma_{\bar{x}[ij]} = \exp\{-\kappa_{\bar{x}}(t_i - t_j)^2\}$ and $\Sigma_{x[ij]} = \exp\{-\kappa_x(t_i - t_j)^2\}$, respectively. The same holds for the baseline process obtaining $\{\mu(t_1), \ldots, \mu(t_n)\}^\intercal \sim \text{N}_n(0, \Sigma_\mu)$.

3. Model Properties

In order to ensure accurate learning and prediction, it is important to guarantee that our factorization in equation (2) along with the Gaussian process priors for its components are sufficiently flexible to approximate a broad variety of dynamic multilayer edge probability processes. These properties are stated in Propositions 1 and 2. See Appendix A for detailed proofs. In particular, Proposition 1 guarantees that equation (2) is sufficiently flexible to

characterize any collection $\pi(t) = \{\pi_{vu}^{(k)}(t) \in (0,1) : k = 1,\ldots,K, \ v = 2,\ldots,V, \ u = 1,\ldots,v-1\}$ for every time $t \in \mathbb{T}$.

Proposition 1 *Let $\pi(t) = \{\pi_{vu}^{(k)}(t) \in (0,1) : k = 1,\ldots,K, \ v = 2,\ldots,V, \ u = 1,\ldots,v-1\}$ denote the collection of edge probabilities between the actors in each layer at time $t \in \mathbb{T}$. Then every collection $\pi(t)$ can be represented as in equation (2) for some R and H.*

Although Proposition 1 guarantees that factorization in equation (2) is sufficiently general, it is important to ensure that the same flexibility is maintained when defining priors on the parameters in equation (2). Proposition 2 guarantees that our model and prior choices induce a prior on the dynamic multilayer edge probability process with full support.

Proposition 2 *If $\mathbb{T}$ is compact, then for every $\pi^0 = \{\pi_{vu}^{0(k)}(t) \in (0,1) : t \in \mathbb{T}, \ k = 1,\ldots,K, \ v = 2,\ldots,V, \ u = 1,\ldots,v-1\}$ and $\epsilon > 0$,*

$$\mathrm{pr}\left(\sup_{t\in\mathbb{T}}\left[\sum_{k=1}^{K}\sqrt{\sum_{v=2}^{V}\sum_{u=1}^{v-1}\{\pi_{vu}^{(k)}(t) - \pi_{vu}^{0(k)}(t)\}^2}\right] < \epsilon\right) > 0,$$

with $\pi_{vu}^{(k)}(t)$ factorized as in equation (2) with Gaussian process priors on the latent coordinates.

Full prior support is a key property in ensuring that our Bayesian formulation assigns a positive probability to a neighborhood of every possible true dynamic multilayer edge probability process, avoiding flexibility issues that may arise if zero mass is assigned to a subset of the parameters space where the truth may be.

In order to highlight the dependence structures induced by our model and priors, we additionally study the prior variances and covariances associated with the log-odds processes $z = \{z_{vu}^{(k)}(t) = \mathrm{logit}\{\pi_{vu}^{(k)}(t)\} \in \Re : t \in \mathbb{T}, \ k = 1,\ldots,K, \ v = 2,\ldots,V, \ u = 1,\ldots,v-1\}$. By conditioning on the shrinkage parameters, and leveraging equations (7) and (8), we obtain

$$\mathrm{var}\{z_{vu}^{(k)}(t) \mid \tau,\tau^{(k)}\} = 1 + \sum_{r=1}^{R}\tau_r^{-2} + \sum_{h=1}^{H}\tau_h^{(k)-2}, \quad \mathrm{cov}\{z_{vu}^{(k)}(t), z_{pq}^{(k)}(t) \mid \tau,\tau^{(k)}\} = 1,$$

for every layer $k = 1,\ldots,K$, time $t \in \mathbb{T}$ and pairs of actors $[vu]$, $v > u$ and $[pq]$, $p > q$ with $p \neq v$ or $q \neq u$. The covariance between different layers, at the same time $t \in \mathbb{T}$ is

$$\mathrm{cov}\{z_{vu}^{(k)}(t), z_{vu}^{(g)}(t) \mid \tau,\tau^{(k)},\tau^{(g)}\} = 1 + \sum_{r=1}^{R}\tau_r^{-2}, \quad \mathrm{cov}\{z_{vu}^{(k)}(t), z_{pq}^{(g)}(t) \mid \tau,\tau^{(k)},\tau^{(g)}\} = 1,$$

for every pair of layers k and g with $k \neq g$, time $t \in \mathbb{T}$ and pairs of actors $[vu]$, $v > u$ and $[pq]$, $p > q$ with $p \neq v$ or $q \neq u$. Finally, the covariances at different times are

$$\mathrm{cov}\{z_{vu}^{(k)}(t_i), z_{vu}^{(k)}(t_j) \mid \tau,\tau^{(k)}\} = e^{-\kappa_\mu(t_i-t_j)^2} + \sum_{r=1}^{R}\tau_r^{-2}e^{-2\kappa_{\bar{x}}(t_i-t_j)^2} + \sum_{h=1}^{H}\tau_h^{(k)-2}e^{-2\kappa_x(t_i-t_j)^2},$$

$$\mathrm{cov}\{z_{vu}^{(k)}(t_i), z_{vu}^{(g)}(t_j) \mid \tau,\tau^{(k)},\tau^{(g)}\} = e^{-\kappa_\mu(t_i-t_j)^2} + \sum_{r=1}^{R}\tau_r^{-2}e^{-2\kappa_{\bar{x}}(t_i-t_j)^2},$$

$$\mathrm{cov}\{z_{vu}^{(k)}(t_i), z_{pq}^{(k)}(t_j) \mid \tau,\tau^{(k)}\} = \mathrm{cov}\{z_{vu}^{(k)}(t_i), z_{pq}^{(g)}(t_j) \mid \tau,\tau^{(k)},\tau^{(g)}\} = e^{-\kappa_\mu(t_i-t_j)^2},$$

9

for every pair of layers k and g with $k \neq g$, times $t_i \in \mathbb{T}$, $t_j \in \mathbb{T}$ and pairs of actors $[vu]$, $v > u$ and $[pq]$, $p > q$ with $p \neq v$ or $q \neq u$.

It is easy to show that a priori the log-odds of the edge probabilities have mean zero, whereas, based on the above results, their variances increase with the sum of the shared and layer-specific shrinkage parameters. Allowing a subset of the coordinates to be common to all layers, has the effect of introducing dependence between the log-odds of edge probabilities in different layers. The Gaussian process priors provide an efficient choice to incorporate dependence between edge probabilities in different times. The strength of this dependence is regulated by the time lag and the positive smoothing parameters κ_μ, $\kappa_{\bar{x}}$ and κ_x. The lower these quantities, the stronger the dependence is between coordinates in different times.

4. Posterior Computation and Prediction

Equations (1)–(2) along with the Gaussian process priors for the latent coordinates can be seen as a nonlinear Bayesian logistic regression with the parameters entering in a bilinear form. Although performing posterior computation in this setting is apparently a cumbersome task, leveraging the Pólya-gamma data augmentation for Bayesian logistic regression (Polson et al., 2013) and adapting derivations in Durante and Dunson (2014), it is possible to obtain a simple and tractable Gibbs sampler having conjugate full conditionals. In fact, the Pólya-gamma data augmentation allows recasting the problem from a logistic regression to a multiple linear regression having transformed Gaussian response data, while derivations in Durante and Dunson (2014) provide results to linearize the factorization in equation (2).

Joining the above procedures and exploiting equations (7)–(8), the updating of the latent coordinates processes at every step simply relies on standard Bayesian linear regression. Algorithm 1 in Appendix B provides derivations and guidelines for step-by-step implementation of our Gibbs sampler. We summarize below the main steps of the MCMC routine.

Step [1]: For each time $t_i = t_1, \ldots, t_n$, layer $k = 1, \ldots, K$ and pair of actors $[vu]$, $v = 2, \ldots, V$, $u = 1, \ldots, v - 1$, sample the corresponding Pólya-gamma augmented data.

Step [2]: Update the baseline process $\mu = \{\mu(t_1), \ldots, \mu(t_n)\}^\intercal$ from its full conditional multivariate Gaussian distribution. This is obtained by recasting the logistic regression for μ in terms of a multiple linear regression having transformed Gaussian response.

Step [3]: Update the shared coordinates. In performing this step we block-sample in turn the coordinates trajectories $\bar{x}_{(v)} = \{\bar{x}_{v1}(t_1), \ldots, \bar{x}_{v1}(t_n), \ldots, \bar{x}_{vR}(t_1), \ldots, \bar{x}_{vR}(t_n)\}^\intercal$ of each actor $v = 1, \ldots, V$ conditionally on the others $\{\bar{x}_{(u)} : u \neq v\}$. This choice allows us to linearize equation (2), with $\bar{x}_{(v)}$ acting as a coefficients vector and $\{\bar{x}_{(u)} : u \neq v\}$ representing appropriately selected regressors. Leveraging the Pólya-gamma data augmentation also this step relies on a multiple linear regression with transformed Gaussian response data, providing Gaussian full conditionals for each $\bar{x}_{(v)}$, $v = 1, \ldots, V$.

Step [4]: Update the layer-specific coordinates. For each layer $k = 1, \ldots, K$, this step relies on the same strategy considered for the shared coordinates, providing again Gaussian full conditionals for each $x_{(v)}^{(k)} = \{x_{v1}^{(k)}(t_1), \ldots, x_{v1}^{(k)}(t_n), \ldots, x_{vH}^{(k)}(t_1), \ldots, x_{vH}^{(k)}(t_n)\}^\intercal$, $v = 1, \ldots, V$ and $k = 1, \ldots, K$. Moreover the updating can be performed separately for each layer, allowing this step to be efficiently implemented in parallel.

Step [5] and [6]: The updating of the gamma parameters characterizing the priors in equations (5)–(6) follows conjugate analysis, proving gamma full conditionals.

Step [7]: Update the dynamic multilayer edge probabilities simply by applying equation (2) to the samples of the baseline process, the shared and the layer-specific coordinates.

In performing posterior computation we set R and H at conservative upper bounds, allowing the multiplicative inverse gamma priors for the shrinkage parameters to delete redundant latent space dimensions not required to characterize the data. Hence, posterior inference is not substantially affected by the choice of R and H, unless these bounds are fixed at excessively low values compared to the complexity of the data analyzed. We additionally assess the goodness of these bounds via in-sample and out-of-sample predictive performance.

4.1 Edge Prediction

Edge prediction is an important topic in dynamic modeling of multilayer networks. For example some networks may have unobserved edges due to inability to monitor certain types of relationships or actors at specific times. Likewise, some layers may be available in time earlier than others, facilitating prediction of those yet missing. The availability of efficient procedures that are able to reconstruct partially unobserved connectivity structures or forecast future networks can have important consequences in many applications, such as destabilization of terrorists networks or epidemic prevention (e.g., Tan et al., 2016).

Our statistical model for dynamic multilayer networks in equations (1)–(2) facilitates the definition of simple procedures for formal edge prediction relying on the expectation $\mathrm{E}\{\mathcal{Y}_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}\}$ of the posterior predictive distribution for $\mathcal{Y}_{vu}^{(k)}(t_i)$, with Y_{obs} denoting the observed data. In fact, under equations (1)–(2), this functional is easily available as

$$\mathrm{E}\{\mathcal{Y}_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}\} = \mathrm{E}_{\pi_{vu}^{(k)}(t_i)}[\mathrm{E}_{\mathcal{Y}_{vu}^{(k)}(t_i)}\{\mathcal{Y}_{vu}^{(k)}(t_i) \mid \pi_{vu}^{(k)}(t_i)\} \mid Y_{\mathrm{obs}}] = \mathrm{E}\{\pi_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}\}, \quad (9)$$

for every time $t_i \in \mathbb{T}$, layer $k = 1, \ldots, K$ and actors $v = 2, \ldots, V$, $u = 1, \ldots, v-1$, where $\mathrm{E}\{\pi_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}\}$ simply coincides with the posterior mean of $\pi_{vu}^{(k)}(t_i)$. Hence prediction requires the posterior distribution of the edge probabilities. These quantities are available also for unobserved edges by adding the following data augmentation step in Algorithm 1:

Step [8]: Impute the missing edges given the current state of $\pi_{vu}^{(k)}(t_i)$ from a $\mathrm{Bern}\{\pi_{vu}^{(k)}(t_i)\}$, for all the combinations of times $t_i \in \mathbb{T}$, layers $k = 1, \ldots, K$ and actors $v = 2, \ldots, V$, $u = 1, \ldots, v-1$ corresponding to unobserved edges.

5. Simulation Study

We consider a simulation study to evaluate the performance of our methodology in a simple scenario partially mimicking the size and structure of the face-to-face dynamic multilayer networks. Consistent with our aim we assess performance in the tasks outlined in Table 2.

To highlight the benefits of our methods (Joint Analysis), we compare the performance with two competitors. The first (Collapsed Analysis) assumes the layers arise from a common latent space model, forcing all the actors coordinates to be shared among layers. This formulation is a special case of our model holding out in (2) the layer-specific similarities.

Quantity	Focus of the performance assessment
Edge probabilities	Assess accuracy in learning the true edge probability process underlying the simulated network data. As already discussed, accurate modeling of the edge probabilities $\pi_{vu}^{(k)}(t_i)$ is fundamental for inference and prediction.
Unobserved edges	Assess out-of-sample predictive performance in forecasting future unobserved networks leveraging the expectation $E\{\mathcal{Y}_{vu}^{(k)}(t_i) \mid Y_{\text{obs}}\}$ of the posterior predictive distribution for the unobserved edges.
Network density	Accuracy in modeling the expected overall chance of connectivity $\theta^{(k)}(t_i) = E[\sum_{v=2}^{V}\sum_{u=1}^{v-1}\mathcal{Y}_{vu}^{(k)}(t_i)/\{V(V-1)/2\}] = \sum_{v=2}^{V}\sum_{u=1}^{v-1}\pi_{vu}^{(k)}(t_i)/\{V(V-1)/2\}$, for every time $t_i \in \mathbb{T}$ and layer $k = 1,\ldots,K$.
Actors degrees	Accuracy in learning the total number of different individuals $\mathrm{d}_v^{(k)}(t_i) = E\{\sum_{u\neq v}\mathcal{Y}_{vu}^{(k)}(t_i)\} = \sum_{u\neq v}\pi_{vu}^{(k)}(t_i)$, each actor v is expected to interact with, for every $v = 1,\ldots,V$, time $t_i \in \mathbb{T}$ and layer $k = 1,\ldots,K$.

Table 2: Quantities of interest, which are considered in performance assessments. Although it is possible to provide inference on several properties, from an epidemiological perspective, the network density and actors degrees are of interest in characterizing the overall risk of contagion and the infectivity of each actor, respectively.

Hence, posterior analysis requires minor modifications of Algorithm 1. The second competitor (Separate Analysis) estimates the dynamic latent space model of Durante and Dunson (2014) separately for each layer. Although both methods are reasonable choices, modeling only shared structures lacks flexibility in capturing layer-specific patterns, whereas separate analyses lose efficiency in inference and prediction when the layers share common patterns. We describe the simulation setting in Section 5.1 and discuss results in Section 5.2.

5.1 Outline of the Simulation Settings

We consider $V = 30$ actors whose connections are monitored for $K = 2$ days on the same time grid $t_1,\ldots,t_{17}$, with $t_i \in \mathbb{T} = \{1,\ldots,17\}$. In simulating the networks, we sample the edges from conditionally independent Bernoulli variables given their corresponding edge probabilities as in equation (1), with these probabilities defined to mimic possible scenarios in our application. To accomplish this goal, we consider, for each day, five representative edge probability matrices $\pi^{0(k)}(t_i)$ at times $t_i \in \{1, 5, 9, 13, 17\}$, displayed in Figure 2, and define the matrices at the remaining times as a convex combination of these representative matrices. Focusing, for example, on the times t_{14}, t_{15} and t_{16} between t_{13} and t_{17}, we let

$$\begin{aligned}
\pi^{0(k)}(t_{14}) &= 0.75\pi^{0(k)}(t_{13}) + 0.25\pi^{0(k)}(t_{17}), \\
\pi^{0(k)}(t_{15}) &= 0.50\pi^{0(k)}(t_{13}) + 0.50\pi^{0(k)}(t_{17}), \\
\pi^{0(k)}(t_{16}) &= 0.25\pi^{0(k)}(t_{13}) + 0.75\pi^{0(k)}(t_{17}).
\end{aligned}$$

The same construction holds for all the other time windows, except for times t_4 and t_6, where we induce a more rapid variation by letting $\pi^{0(k)}(t_4) = \pi^{0(k)}(t_6) = \pi^{0(k)}(t_5)$, instead
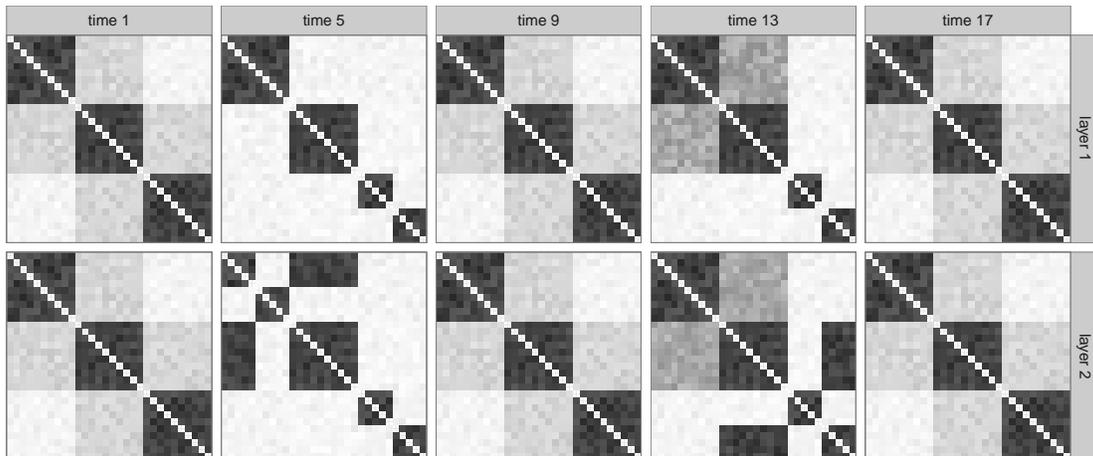
Figure 2: True representative edge probability matrices $\pi^{0(k)}(t_i)$. The color in each cell goes from white to black as the corresponding edge probability goes from 0 to 1.

of $\pi^{0(k)}(t_4) = 0.25\pi^{0(k)}(t_1) + 0.75\pi^{0(k)}(t_5)$, and $\pi^{0(k)}(t_6) = 0.75\pi^{0(k)}(t_5) + 0.25\pi^{0(k)}(t_9)$. We incorporate this setting to study the effects of the Gaussian process' constant smoothness assumption in scenarios having time-varying smoothness. In order to empirically evaluate Proposition 1, the edge probability matrices in Figure 2 are not generated from equation (2), but are instead constructed to characterize possible face-to-face contact scenarios.

Focusing on the first day, which corresponds to layer $k = 1$, the edge probability matrices at times t_1, t_9 and t_{17} characterize contacts during breakfast, lunch and dinner, respectively. These favor block-structures due to homophily by age and gender. In fact, it is reasonable to expect that young individuals, adult women and adult men, corresponding to actors in the first, second, and third block, may have a higher chance of contact with individuals in the same socio-demographic group. We additionally allow adult women to have a moderate chance of contact with young individuals and adult men during breakfast, lunch, and dinner times. The edge probability matrix at time t_5 characterizes instead block-structures due environmental restrictions during the morning, with young individuals attending school, adult women sharing the house, and two sub-groups of adult men working in two different places. Finally, the edge probability matrix at time t_{13} is similar to the one at time t_5 with exception of an increased chance of contact between young individuals and adult women who are expected to share the same environment in the afternoon after school.

The dynamic contact networks in the different days share similar patterns, as shown in Figure 1. We maintain this property by considering the same representative edge probability matrices in the second day, which corresponds to layer $k = 2$, with exception of t_5 and t_{13}. In the morning of the second day, we assume five young individuals contract a disease and remain at home. This increases their chance of contact with the adult women taking care of them, and reduces the probability of a contact with the other young individuals. The edge probability matrix in the afternoon of the second day characterizes a similar scenario, but focuses on five adult men. Including different edge probability matrices at times t_5 and t_{13}, allows us to assess performance in learning layer-to-layer differences in contact patterns.

	Squared Posterior Bias			Posterior Variance			Posterior Concentration		
	$n\downarrow$	$V\downarrow$		$n\downarrow$	$V\downarrow$		$n\downarrow$	$V\downarrow$	
Joint	**0.009**	0.014	0.012	**0.011**	0.018	0.016	**0.020**	0.032	0.028
Collapsed	**0.016**	0.019	0.019	**0.007**	0.011	0.010	**0.023**	0.030	0.029
Separate	**0.016**	0.025	0.019	**0.018**	0.027	0.023	**0.034**	0.052	0.042

Table 3: For our methodology and the two competitors, measures of posterior concentration around the true dynamic multilayer edge probability process, at varying n and V. Bold numbers are the measures associated with our initial simulation scenario.

In order to assess predictive performance, we perform posterior analysis under our model and the competing methods, holding out from the observed data the networks from time t_{13} to t_{17} in the second day. This choice provides a scenario of interest in our application. For example a subset of actors may contract a disease at time t_{12} in the second day, motivating the design of prevention policies relying on the forecasted contact patterns at future times.

5.2 Results in the Simulation

We perform posterior computation under our model with $\kappa_\mu = \kappa_{\bar{x}} = \kappa_x = 0.05$ to favor smooth trajectories a priori and $a_1 = 2$, $a_2 = 2.5$ to facilitate adaptation of the latent spaces dimensions. Although our model can be easily modified to learn these hyperparameters from the data as in Murray and Adams (2010) and Bhattacharya and Dunson (2011), respectively, borrowing of information within the dynamic multilayer networks has the effect of reducing sensitivity to the choice of the hyperparameters. In fact, we found results robust to moderate changes in these quantities, and therefore prefer to elicit them to improve convergence and mixing. We consider 5000 Gibbs iterations with $R = H = 5$, and set a burn-in of 1000. Traceplots for the time-varying edge probabilities in the two layers show no evidence against convergence, and mixing is good in our experience with most of the effective sample sizes for the quantities of interest being around 1500 out of 4000 samples. Posterior computation for our competitors is performed with the same smoothing parameters and $a_1 = 2$, $a_2 = 2.5$, using $R + H = 10$ dimensional latent spaces, to improve comparison with our results.

5.2.1 PERFORMANCE IN LEARNING THE TRUE EDGE PROBABILITY PROCESS

Flexible modeling of $\pi^{(k)}(t)$ is fundamental for learning and prediction. Table 3 summarizes the concentration of the posterior distributions for the dynamic multilayer edge probabilities using the posterior mean of the squared difference between $\pi_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}$ and $\pi_{vu}^{0(k)}(t_i)$. To provide empirical insights on posterior consistency we study this property also at varying n and V. In the first case we focus on the dynamic multilayer network studied at odd times $t_i \in \{1, 3, 5, \ldots, 17\}$, whereas in the second case we consider the reduced networks where four actors within each block are not analyzed. To highlight the contribution of the squared posterior bias and the posterior variance, we factorize the concentration measure as

$$\mathrm{E}([\{\pi_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}\} - \pi_{vu}^{0(k)}(t_i)]^2) = [\mathrm{E}\{\pi_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}\} - \pi_{vu}^{0(k)}(t_i)]^2 + \mathrm{var}\{\pi_{vu}^{(k)}(t_i) \mid Y_{\mathrm{obs}}\}.$$

To assess the overall performance, these measures are averaged across the edge probabilities
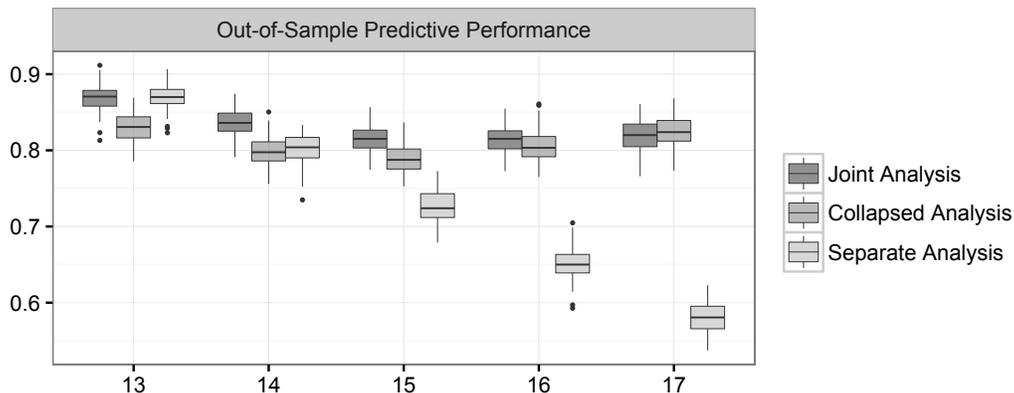
Figure 3: For times from t_{13} to t_{17} —held out in the second day from posterior computation —boxplots of the areas under the ROC curves (AUC) when predicting the unobserved edges with our method and the two competitors, from 100 simulations.

characterizing the dynamic multilayer network process.

According to Table 3, borrowing of information across layers has the effect of reducing both bias and posterior variance compared to the separate analyses, improving accuracy and efficiency in inference and prediction. Enhancing flexibility via layer-specific latent coordinates additionally facilitates bias reduction compared to collapsed analyses, but provides posterior distributions with slightly higher variance. This result is not surprising provided that the collapsed analyses assume the edge probability process to be common across layers, reducing the number of parameters. Moreover, although low variance is a desired feature when bias is small, this property may lead to poor inference and prediction when the posterior distribution is centered far from the truth. In our simulation—having few differences between the true edge probability processes in the two layers—the bias is low also for the collapsed analyses, but this bias is expected to grow when the layers have more differences.

The above results are maintained at varying n and V, while showing improved concentration for higher n or V. Although it would be interesting to prove posterior consistency, state-of-the-art literature provide results for the single binary outcome case (e.g., Ghosal and Roy, 2006), and it is not clear how to adapt this theory to our much complex network–valued data. However, results in Table 3 provide positive empirical support for this property.

5.2.2 Out-of-sample Predictive Performance

As discussed in Sections 2 and 4, developing models characterized by lower bias is important to improve inference and prediction. In fact, our method has an improved ability to predict unobserved edges compared to the competing models, as shown in Figure 3.[1] In predicting held-out edges we leverage the procedures discussed in Section 4.1. Since equation (1) is also valid for our competitors, results in Section 4.1 can be used for the collapsed and separate

---

1. In assessing predictive performance in Figure 3, we generate for each time from t_{13} to t_{17}—held out in the second day from posterior computation—100 networks from (1) with edge probabilities $\pi_{vu}^{0(2)}(t_i)$, and compute for each network the area under the ROC curve when predicting its edges with the expectation of the posterior predictive distribution associated with each method—according to Section 4.1.
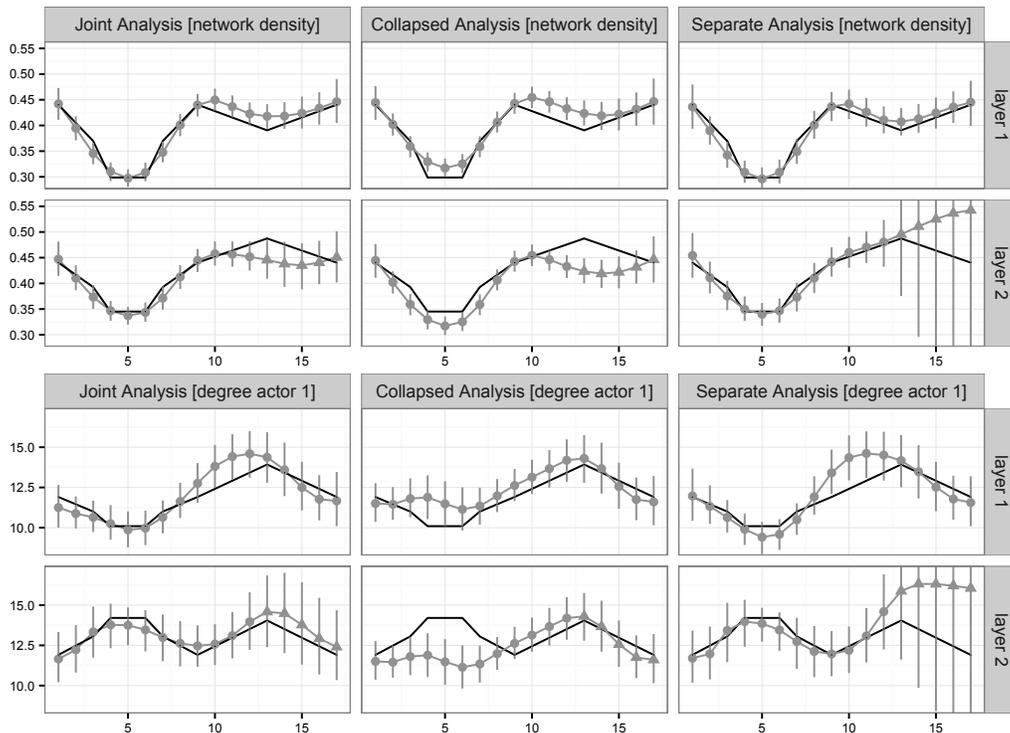
Figure 4: For the three methods, posterior mean (grey lines) and pointwise 0.95 credibility intervals (grey segments) for the dynamic expected network density (upper panels) and the dynamic expected degree of the first actor (lower panels) in the two layers. Black lines are the true trajectories, and triangles the held-out networks.

analyses after replacing $E\{\pi_{vu}^{(k)}(t_i) \mid Y_{\text{obs}}\}$ with $E\{\pi_{vu}(t_i) \mid Y_{\text{obs}}\}$ and $E\{\pi_{vu}^{(k)}(t_i) \mid Y_{\text{obs}}^{(k)}\}$, respectively.

According to Figure 3 the separate analysis obtains good predictive performance for the times subsequent to the last observed network, but this accuracy decays with increasing time lag. As the underlying edge probability process is smooth, it is reasonable to expect good predictive performance at time t_{13} exploiting information at time t_{12}, whereas the reduced accuracy for later times is likely due to inefficient exploitation of the information from the first layer. As expected, the collapsed analysis provides accurate predictions when there is a high degree of similarity between the edge probabilities in the two layers, such as at time t_{17}, but the accuracy decreases in correspondence of more evident differences. Differently from the above procedures, our model incorporates dependence between layers without assuming a single edge probability process common to all of them, providing predictions which are less sensitive to the underlying data generating mechanism. In fact, as shown in Figure 3, we obtain an improved prediction of unobserved edges compared to the competing methods.

5.2.3 Performance in Learning Network Properties of Interest

Accurate modeling of the dynamic multilayer edge probabilities has key benefits in improving inference on time-varying network structures in the different layers. Figure 4 summarizes

the posterior distribution of the dynamic expected network density, and the time-varying expected degree of a selected actor in the two layers.

As shown in Figure 4, our method outperforms the competing approaches also in providing inference for the selected network properties. The collapsed analysis forces the trajectories of these networks summary measures to be equal across layers, causing poor performance in capturing differences among layers in specific times. Modeling the dynamic edge probability processes underlying each layer separately improves flexibility, but the inefficient exploitation of the information shared between layers leads to poor performance in modeling network properties associated with unobserved edges. Our formulation in equations (1)–(2) incorporates dependence between layers and preserves flexibility via a set of layer-specific latent coordinates, which reduces in-sample and out-of-sample bias. Although our model has accurate performance, the constant smoothness assumption characterizing the Gaussian process priors leads to a slight over-smoothing around the more rapid variation in times t_4 and t_6. One possibility to avoid this behavior is to replace the Gaussian process priors with nested Gaussian processes (Zhu and Dunson, 2013) having time-varying smoothness.

6. Application to Dynamic Multilayer Face-to-Face Contact Data

We now turn to applying the methods proposed in Sections 2 and 4 to the face-to-face contact networks described in Section 1.2. Raw contact data are available for 75 individuals belonging to five different households in rural Kenya. Each household has different families living in the same compound and reporting to one head. Face-to-face contacts for the individuals in each household are collected hourly from 06:00 to 20:00 during three consecutive days, with these three days windows differing for most of the households. Hence, contact data are unavailable for almost all the pairs corresponding to actors in different households, making it impossible for any methodology to infer across-households connectivity structures. In order to avoid complications due to non-overlapping days, we focus on the face-to-face contact networks for the individuals in the most populated household comprising $V = 29$ actors. This choice is further supported by the fact that within households contacts play a fundamental role in the speed of contagion, motivating primary interests in household-based analyses; see e.g., House and Keeling (2009) and the references cited therein.

6.1 Results in the Application

The face-to-face proximity data can be structured as a dynamic multilayer network recording for each day $k \in \{1, 2, 3\}$ the time-varying face-to-face contacts among the $V = 29$ actors on the same hourly basis $t_1, \ldots, t_{14}$, where $t_i \in \{1, \ldots, 14\}$ is simply a discrete time index denoting the current hour. We apply our statistical model and the competing methods to the aforementioned data, where we hold out the contact network at the last time in the third day $Y_{t_{14}}^{(3)}$ to assess out-of-sample predictive performance.[2]

---

2. Posterior computation under the three models uses the same default hyperparameters as in the simulation study. Similarly to the simulation study, we consider 5000 Gibbs iterations and obtain no evidence against convergence after a burn-in of 1000. The effective sample sizes for the quantities of interest for inference are around 1200 out of 4000 samples. Although this provides a slightly reduced result compared to the simulation studies, the mixing is still good given the complexity of our dynamic multilayer network data.

		Area under the curve			Accuracy with cutoff 0.5		
Number of latent coordinates		8	10	12	8	10	12
Joint	In–sample	0.97	0.97	0.97	95%	95%	94%
	Out–of–sample	0.95	0.95	0.95	94%	94%	95%
Collapsed	In–sample	0.94	0.94	0.95	92%	92%	92%
	Out–of–sample	0.95	0.95	0.95	94%	94%	94%
Separate	In–sample	0.97	0.97	0.97	94%	94%	94%
	Out–of–sample	0.82	0.83	0.82	89%	89%	89%

Table 4: For our method and the two competitors, in-sample and out-of-sample predictive performance measured via the AUC, and the percentage of correctly predicted edges—using a cutoff probability of 0.5—at varying $R + H$, with $R = H$.

6.1.1 In-sample and Out-of-sample Predictive Performance

As one assessment of our proposed methodology, Table 4 compares our performance for in-sample and out-of-sample edge prediction with those provided by the competing procedures. Consistent with the strategy outlined in equation (9), the ROC curves for in-sample edge prediction are constructed using the observed data Y_{obs} and the posterior mean of their corresponding edge probabilities, estimated under the three different methods. The ROC curves for assessing out-of-sample predictive performance are defined in a similar way, but focusing on the held-out network at time t_{14} in day 3. To evaluate sensitivity to the choice of the upper bounds for the latent space dimensions, we study predictive performance also under other settings of R and H, including $R = H = 4$ and $R = H = 6$.

Edge prediction is clearly more accurate in-sample than out-of-sample. This is because Y_{obs} is part of the posterior computation, whereas the network at time t_{14} in day 3 is held out. According to Table 4, the separate analysis achieves comparable results to our model for in-sample edge prediction in accommodating layer-specific patterns, but provides poor predictions of future edges due to inability to incorporate information from the previous days. The collapsed analysis allows instead borrowing of information between the three days via shared coordinates, but does not incorporate differences in contact patters across these days. As a result, the collapsed analysis provides comparable performance to our model for out-of-sample edge prediction, but has a reduced in-sample predictive ability. Our dynamic multilayer representation efficiently incorporates both shared structures and layer-specific differences, facilitating good predictive performance both in-sample and out-of-sample. As discussed in Section 4, these results are not substantially affected by moderate variations in R and H, and therefore we perform posterior inference for $R = H = 5$, as in the simulation.

6.1.2 Learning Network Properties of Interest

The above results motivate additional posterior analyses of the dynamic multilayer stochastic process estimated under our model. Figure 5 summarizes the posterior distribution for the trajectory of the expected network density in the three days. The expected frequency
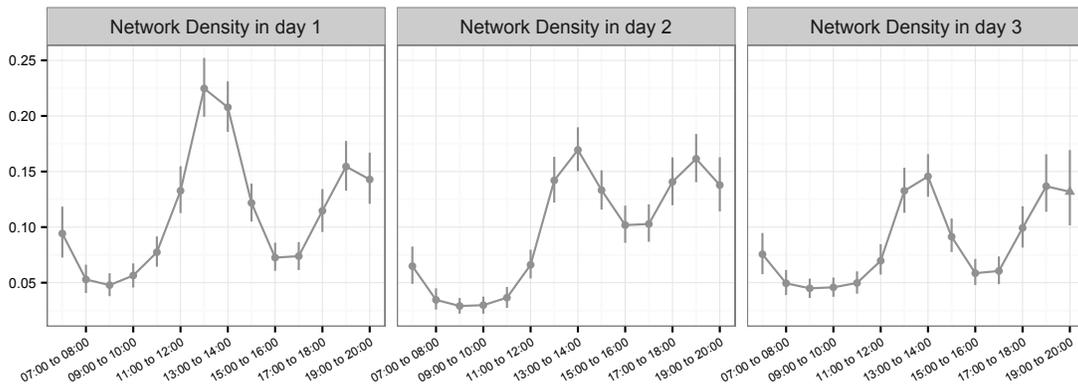
Figure 5: Posterior mean (grey lines) and pointwise 0.95 credibility intervals (grey segments) for the time-varying expected network density in the three days. The triangle for the last time in the third day denotes the held-out contact network.

of proximity contacts evolves on similar patterns during the three days, while showing some day-specific variations at certain hours. This result further confirms the importance of analyzing such data with statistical models that are able to borrow information across days, while maintaining flexibility in capturing day-specific patterns. Consistent with Kiti et al. (2016), the expected network density remains in general on low levels during the morning and afternoon when the actors are in different environments such as school and workplace, and peaks in correspondence of lunch and dinner times when the households members congregate. We learn a similar pattern when considering the dynamic expected frequencies of contact within and between groups of actors having similar traits, such as gender and age.[3] These trajectories provide key additional information compared to the expected network density, highlighting how infectious diseases can spread within and between groups of actors. The posterior mean of these trajectories is visualized in Figure 6.

According to Figure 6, all the trajectories inherit the pattern discussed for the expected network density, with clear peaks during lunch and dinner times. Although there are not evident differences in the dynamic expected frequencies of face-to-face contact within and between gender groups, we notice how contacts among males are typically less likely than contacts between females, whereas face-to-face interactions among actors of different genders become notable during lunch. Therefore, these hours might be at risk of spreading diseases from a gender group to the other. Consistent with the static analyses considered in Kiti et al. (2016), we observe more differences in the trajectories by age groups, with a substantially high chance of contact between young actors, whereas the dynamic expected frequency of interaction among adults remains on low levels. In fact, young individuals potentially have less restrictions from environment or work schedule than adults, and therefore have more chances of interaction, especially with actors of their same age. Hence, diseases may spread more easily among young individuals and reach adults during lunch times when the chance of contact between the two age groups is higher.

---

3. We can easily derive the posterior mean of these trajectories by averaging, for each time t_i and day k, all the estimated edge probabilities corresponding to pairs of actors having the same combination of traits, e.g., both males, both females, one male and one female.
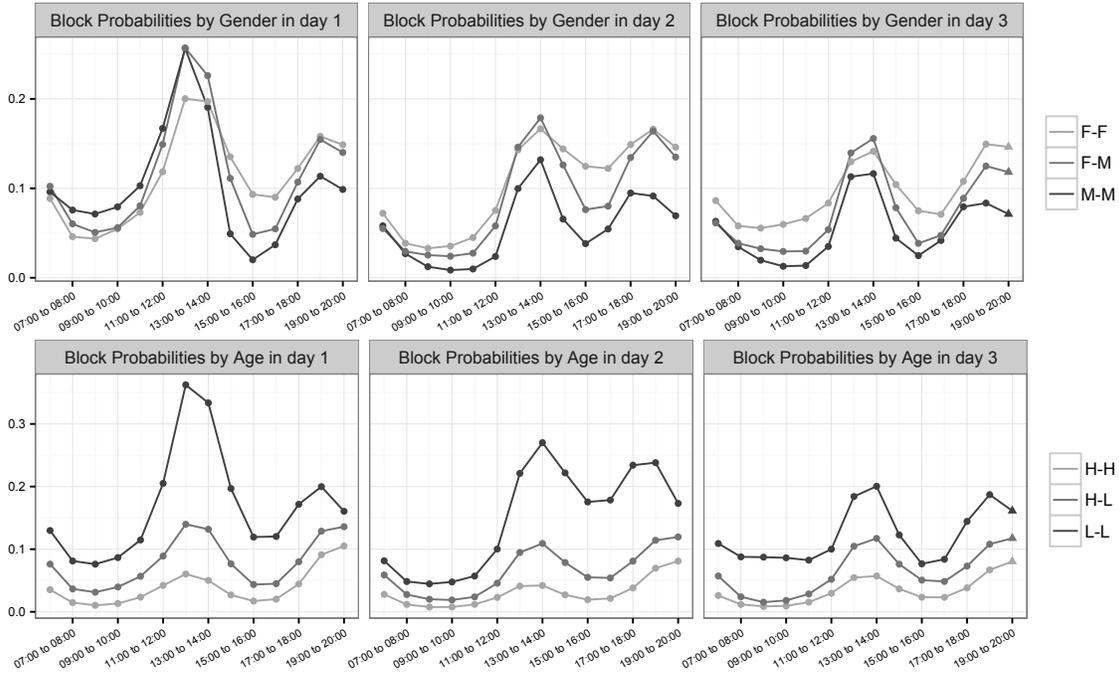
Figure 6: Upper panels: posterior mean of the dynamic expected frequencies of contact for actors having same gender—both males or both females—and different gender, in each day. Lower panels: same quantities with respect to high (≥ 15) and low (<15) age groups.

Figures 5 and 6 provide global measures for the dynamic risk of contagion in the network and for groups of actors. In order to develop more refined prevention policies it is of key interest to monitor the infectivity of each actor, and how this measure evolves with time and days. Consistent with this goal, Figure 7 provides a graphical analysis of selected contact networks, with the actors positions depending on the estimated edge probabilities averaged, for each day, over the two time windows having the more dense networks which may lead to a faster disease spread. As shown in Figure 7, the contact patterns in the same time window share a similar configuration across the three days, while displaying some day-to-day difference in the connectivity behavior of subsets of actors. For example, the community structure formed by actors 6, 17 and 28 during lunch times is less separated from the rest of the network in the first day, compared to the second and third day. This result provides additional support for our methodology which can simultaneously incorporate shared patterns and possible day-specific deviations from these common structures. Consistent with Figures 5 and 6, young individuals—represented by circles—are typically characterized by more frequent contacts, whereas the adult actors—corresponding to squares—have a more peripheral position in the network. As a result the degree of the young individuals is typically higher, making these actors particularly relevant in the economy of disease contagion and transmission. In this respect, actors 9 and 15 may represent dangerous hubs.

20
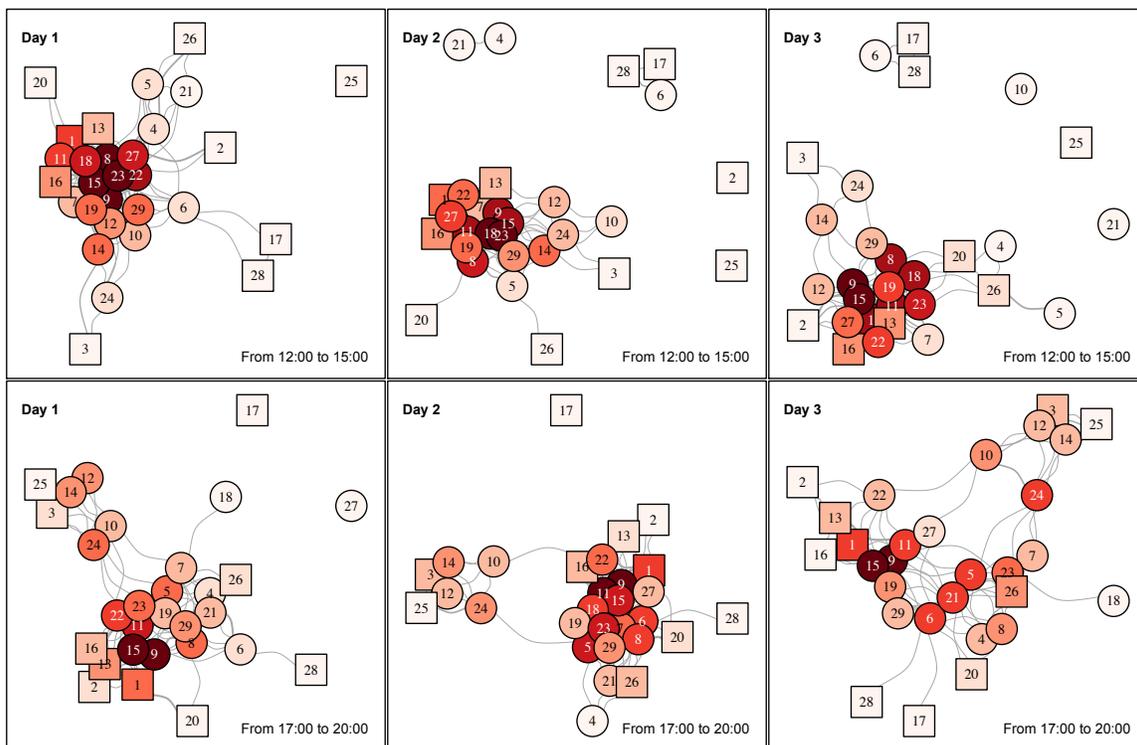
Figure 7: Weighted network visualization with weights obtained by averaging the posterior mean of the edge probabilities in the two time windows having the more dense networks, within each of the three days. Squares and circles represent the actors having high (≥ 15) and low (< 15) age, respectively. The color of each node goes from white to dark red as the estimated expected degree of the corresponding actor increases, relative to the others. Actors positions are obtained by applying the Fruchterman and Reingold (1991) force–directed placement algorithm, whereas the width of each edge is proportional to its averaged edge probability.

We can additionally notice some community structure within the network, particularly during dinner times. For example, actors 3, 10, 12, 14, 24 and 25 are tightly connected, but separated from the others, with exception of the young subjects 10 and 24 which may play a key role in the transmission of a disease from their community to the rest of the network. Other two communities can be more clearly observed during dinner times in day three. It is interesting to notice how these groups of actors comprise both young and adult individuals, and therefore may represent different families within the household under analysis.

7. Discussion

The increasing availability of multidimensional, complex, and dynamic information on social interaction processes motivates a growing demand for novel statistical models. In order to successfully enhance quality in inference and prediction, these models need to efficiently incorporate the complex set of dependencies in the observed data, without affecting flexi-

bility. Motivated by this consideration and by epidemiological studies monitoring disease transmission via dynamic face-to-face interactions, we have developed a Bayesian nonparametric model for dynamic multilayer networks leveraging latent space representations. In order to preserve flexibility and borrow information across layers and time, we modeled the edge probabilities as a function of shared and layer-specific latent coordinates which evolve in time via Gaussian process priors. We provided theoretical support for our model, and developed simple procedures for posterior computation and formal prediction. Finally, we illustrated on both simulated data and on infection studies monitoring face-to-face contacts that our methods perform better than competitors in terms of inference and prediction.

Although we focus on face-to-face interaction networks collected at multiple times and days, our methods have a broad range of applications. Notable examples include dynamic cooperations among countries with respect to different types of international relations, time-varying interactions between researchers according to multiple forms of academic collaborations, and dynamic contacts between terrorists in relation to different types of dark interactions. In all these relevant applications, our flexible methodology can provide an appealing direction in accurately learning and predicting hidden wiring mechanisms and their implication in several environments and phenomena.

In addition, our methods motivate further directions of research. An important one is to facilitate scaling to larger dynamic multilayer network data. Currently, the computational complexity of our Gibbs sampler—as a function of the dimensions of the input data—is of order $O(VKn^3)$, and corresponds to the most intensive step updating the layer-specific latent coordinates trajectories for the V actors. Although the latent space representation reduces computational complexity from quadratic in the number of actors V to linear, the cubic complexity in the number of time points n associated with the Gaussian process prior, may still represent a computational barrier when data are monitored for wide time windows. A possible strategy to successfully address this issue is to consider more scalable processes such as the low-rank approximations to the Gaussian process (Banerjee et al., 2013) or state-space models. Another important generalization is accommodating the more informative contact counts instead of just a binary variable indicating presence or absence of face-to-face interactions. In accomplishing this goal one possibility is to adapt the methodology proposed in Canale and Dunson (2011) to our framework and assume the weighted edges are realizations from a rounded Gaussian whose mean is factorized as in equation (2). Finally, additional studies on the theoretical properties of these stochastic processes, such as those recently considered in Crane (2017), are of relevant interest.

Acknowledgments

Appendix A. Proofs of Propositions 1 and 2

Proof of Proposition 1. As the logistic mapping is one-to-one continuous, Proposition 1 is valid if and only if—for each time $t \in \mathbb{T}$—every possible collection of log-odds $z(t) = \{z_{vu}^{(k)}(t) \in \Re : k = 1, \ldots, K, \ v = 2, \ldots, V, \ u = 1, \ldots, v-1\}$ can be factorized as

$$z_{vu}^{(k)}(t) = \mu(t) + \bar{s}_{vu}(t) + s_{vu}^{(k)}(t) = \mu(t) + \bar{x}_v(t)^\intercal \bar{x}_u(t) + x_v^{(k)}(t)^\intercal x_u^{(k)}(t),$$

for all $k = 1, \ldots, K, \ v = 2, \ldots, V, \ u = 1, \ldots, v-1$, where $x_v^{(k)}(t) = \{x_{v1}^{(k)}(t), \ldots, x_{vH}^{(k)}(t)\}^\intercal$ and $\bar{x}_v(t) = \{\bar{x}_{v1}(t), \ldots, \bar{x}_{vR}(t)\}^\intercal$, for every $v = 1, \ldots, V$.

Assume without loss of generality $\mu(t) = 0$ for every $t \in \mathbb{T}$ and let $\bar{X}(t)$ and $X^{(k)}(t)$ for $k = 1, \ldots, K$, denote the $V \times R$ and $V \times H$ matrices containing the shared and layer-specific latent coordinates, respectively, at time $t \in \mathbb{T}$. Since we are not interested in the diagonal elements of the $V \times V$ edge probabilities and log-odds matrices, it is always possible to write

$$Z^{(k)}(t) = \bar{S}(t) + S^{(k)}(t), \quad k = 1, \ldots, K,$$

where $Z^{(k)}(t)$ is the $V \times V$ symmetric matrix having the log-odds of the edges at time t in layer k as off-diagonal elements, whereas $\bar{S}(t)$ and $S^{(k)}(t)$ are $V \times V$ positive semidefinite symmetric matrices having quantities $\bar{s}_{vu}(t), \ v = 2, \ldots, V, \ u = 1, \ldots, v-1$ and $s_{vu}^{(k)}(t)$, $v = 2, \ldots, V, \ u = 1, \ldots, v-1$ as off-diagonal elements. Since we are not interested in self-relations, there is no loss of generality in assuming $\bar{S}(t)$ and $S^{(k)}(t)$ are positive semidefinite, since for any configuration of shared and layer-specific similarities there exist infinitely many positive semidefinite matrices having these quantities as off-diagonal elements.

Since $\bar{S}(t)$ and $S^{(k)}(t)$ are positive semidefinite, they admit the eigendecompositions $\bar{S}(t) = \bar{U}(t)\bar{\Lambda}(t)\bar{U}(t)^\intercal$ and $S^{(k)}(t) = U^{(k)}(t)\Lambda^{(k)}(t)U^{(k)}(t)^\intercal$, where $\bar{U}(t)$ and $U^{(k)}(t)$ denote the $V \times R_t$ and $V \times H_t^{(k)}$ matrices of eigenvectors, whereas $\bar{\Lambda}(t)$ and $\Lambda^{(k)}(t)$ are the corresponding diagonal matrices with the positive eigenvalues. Therefore, letting $R \geq R_t$ for every $t \in \mathbb{T}$ and $H \geq H_t^{(k)}$, for every $k = 1, \ldots, K$ and $t \in \mathbb{T}$, Proposition 1 follows after defining $\bar{X}(t)$ and $X^{(k)}(t)$ as the block matrices $\bar{X}(t) = \{\bar{U}(t)\bar{\Lambda}(t)^{1/2}, 0_{V \times (R-R_t)}\}$ and $X^{(k)}(t) = \{U^{(k)}(t)\Lambda^{(k)}(t)^{1/2}, 0_{V \times (H-H_t^{(k)})}\}$ for every $k = 1, \ldots, K$. ∎

Proof of Proposition 2. Leveraging proof of Corollary 2 in Durante and Dunson (2014), to prove Proposition 2 it suffices to show that

$$\mathrm{pr}\left(\sup_{t \in \mathbb{T}}\left[\sum_{k=1}^{K}\sqrt{\sum_{v=2}^{V}\sum_{u=1}^{v-1}\{z_{vu}^{(k)}(t) - z_{vu}^{0(k)}(t)\}^2}\right] < \epsilon\right) > 0,$$

where $z_{vu}^{(k)}(t)$ and $z_{vu}^{0(k)}(t)$ are the log-odds of $\pi_{vu}^{(k)}(t)$ and $\pi_{vu}^{0(k)}(t)$, respectively. Recalling the proof of Proposition 1, the above probability can be factorized as

$$\mathrm{pr}\left(\sup_{t \in \mathbb{T}}\left[\sum_{k=1}^{K}\sqrt{\sum_{v=2}^{V}\sum_{u=1}^{v-1}\{\mu(t) + \bar{s}_{vu}(t) + s_{vu}^{(k)}(t) - \mu^0(t) - \bar{s}_{vu}^0(t) - s_{vu}^{0(k)}(t)\}^2}\right] < \epsilon\right),$$

with $\bar{s}_{vu}(t) = \bar{x}_v(t)^\intercal \bar{x}_u(t)$, $s_{vu}^{(k)}(t) = x_v^{(k)}(t)^\intercal x_u^{(k)}(t)$, $\bar{s}_{vu}^0(t) = \bar{x}_v^0(t)^\intercal \bar{x}_u^0(t)$ and $s_{vu}^{0(k)}(t) = x_v^{0(k)}(t)^\intercal x_u^{0(k)}(t)$. Exploiting the triangle inequality and the independence of the Gaussian

process priors for the different trajectories, a lower bound for the above probability is

$$
\text{pr}\left(\sup_{t\in\mathbb{T}}\left[\sqrt{\sum_{v=2}^{V}\sum_{u=1}^{v-1}\{\mu(t)-\mu^0(t)+\bar{s}_{vu}(t)-\bar{s}_{vu}^0(t)\}^2}\right]<\frac{\epsilon}{2K}\right)\times
$$

$$
\times\prod_{k=1}^{K}\text{pr}\left(\sup_{t\in\mathbb{T}}\left[\sqrt{\sum_{v=2}^{V}\sum_{u=1}^{v-1}\{s_{vu}^{(k)}(t)-s_{vu}^{0(k)}(t)\}^2}\right]<\frac{\epsilon}{2K}\right).
$$

Applying Theorem 2 in Durante and Dunson (2014) to each term in this factorization, it is easy to show that all the above probabilities are strictly positive, proving Proposition 2. ∎

Appendix B. Pseudocode for Posterior Computation

Algorithm 1 provides guidelines for step-by-step implementation of our Gibbs sampler.

Algorithm 1 Gibbs sampler for the dynamic multilayer latent space model

---

[1] **Generate the Pólya-gamma augmented data**
for each $t_i = t_1, \ldots, t_n$, $k = 1, \ldots K$, $v = 2, \ldots, V$ and $u = 1, \ldots, v-1$ **do**
  Sample the augmented data $\omega_{vu}^{(k)}(t_i)$ from the full conditional Pólya-gamma

$$
\omega_{vu}^{(k)}(t_i)\mid - \sim \text{PG}\left\{1, \mu(t_i)+\bar{x}_v(t_i)^{\mathsf{T}}\bar{x}_u(t_i)+x_v^{(k)}(t_i)^{\mathsf{T}}x_u^{(k)}(t_i)\right\},
$$

  where $\text{PG}(a,b)$ is the Pólya-gamma random variable with parameters $a > 0$ and $b \in \Re$.
end for

---

[2] **Update the baseline trajectory** $\mu = \{\mu(t_1), \ldots, \mu(t_n)\}^{\mathsf{T}}$ from $\mu \mid - \sim \text{N}_n(\mu_\mu, \Psi_\mu)$
where $\Psi_\mu = \left[\text{diag}\left\{\sum_{k=1}^{K}\sum_{v=2}^{V}\sum_{u=1}^{v-1}\omega_{vu}^{(k)}(t_1), \ldots, \sum_{k=1}^{K}\sum_{v=2}^{V}\sum_{u=1}^{v-1}\omega_{vu}^{(k)}(t_n)\right\}+\Sigma_\mu^{-1}\right]^{-1}$
and $\mu_\mu = \Psi_\mu \eta_\mu$ with

$$
\eta_\mu = \left[\begin{array}{c}\sum_{k=1}^{K}\sum_{v=2}^{V}\sum_{u=1}^{v-1}\{Y_{t_1[vu]}^{(k)}-1/2-\omega_{vu}^{(k)}(t_1)[\bar{x}_v(t_1)^{\mathsf{T}}\bar{x}_u(t_1)+x_v^{(k)}(t_1)^{\mathsf{T}}x_u^{(k)}(t_1)]\} \\ \vdots \\ \sum_{k=1}^{K}\sum_{v=2}^{V}\sum_{u=1}^{v-1}\{Y_{t_n[vu]}^{(k)}-1/2-\omega_{vu}^{(k)}(t_n)[\bar{x}_v(t_n)^{\mathsf{T}}\bar{x}_u(t_n)+x_v^{(k)}(t_n)^{\mathsf{T}}x_u^{(k)}(t_n)]\}\end{array}\right].
$$

---

[3] **Sample the vectors of shared coordinates** $\bar{x}_v(t_1), \ldots, \bar{x}_v(t_n)$ **for** $v = 1, \ldots, V$
for each actor $v = 1, \ldots, V$ **do**
  Block-sample $\{\bar{x}_v(t_1), \ldots, \bar{x}_v(t_n)\}$ given the others $\{\bar{x}_u(t_i) : u \neq v, t_i = t_1, \ldots t_n\}$.

  [a] Let $\bar{x}_{(v)} = \{\bar{x}_{v1}(t_1), \ldots, \bar{x}_{v1}(t_n), \ldots, \bar{x}_{vR}(t_1), \ldots, \bar{x}_{vR}(t_n)\}^{\mathsf{T}}$

  [b] Define a Bayesian logistic regression with $\bar{x}_{(v)}$ acting as coefficient vector and having prior, according to equation (7), $\bar{x}_{(v)} \sim \text{N}_{n\times R}\left\{0, \text{diag}(\tau_1^{-1}, \ldots, \tau_R^{-1})\otimes\Sigma_{\bar{x}}\right\}$

---

[c] For every $k = 1, \ldots, K$, the Bayesian logistic regression for the updating of $\bar{x}_{(v)}$ is

$$Y_{(v)}^{(k)} \sim \text{Bern}(\pi_{(v)}^{(k)}), \quad \text{logit}(\pi_{(v)}^{(k)}) = 1_{V-1} \otimes \mu + \bar{X}_{(-v)}\bar{x}_{(v)} + X_{(-v)}^{(k)}x_{(v)}^{(k)},$$

with

- $Y_{(v)}^{(k)}$ obtained by stacking vectors $\{Y_{t_1[vu]}^{(k)}, \ldots, Y_{t_n[vu]}^{(k)}\}^{\mathsf{T}}$ for all pairs having v as a one of the two actors
- $\pi_{(v)}^{(k)}$ the corresponding vector of edge probabilities
- $x_{(v)}^{(k)} = \{x_{v1}^{(k)}(t_1), \ldots, x_{v1}^{(k)}(t_n), \ldots, x_{vH}^{(k)}(t_1), \ldots, x_{vH}^{(k)}(t_n)\}^{\mathsf{T}}$
- $\bar{X}_{(-v)}$ and $X_{(-v)}^{(k)}$ the matrices of regressors whose entries are suitably chosen from $\{\bar{x}_u(t_i) : u \neq v, t_i = t_1, \ldots t_n\}$ and $\{x_u^{(k)}(t_i) : u \neq v, t_i = t_1, \ldots t_n\}$

According to the above specification, and letting $\Omega_{(v)}^{(k)}$ denote the diagonal matrix with the corresponding Pólya-gamma augmented data, we obtain

$$\bar{x}_{(v)} \mid - \sim \text{N}_{n \times R}\left(\mu_{\bar{x}_{(v)}}, \Psi_{\bar{x}_{(v)}}\right),$$

with

$$\Psi_{\bar{x}_{(v)}} = \{\bar{X}_{(-v)}^{\mathsf{T}}(\Omega_{(v)}^{(1)} + \ldots + \Omega_{(v)}^{(K)})\bar{X}_{(-v)} + \text{diag}(\tau_1, \ldots, \tau_R) \otimes \Sigma_{\bar{x}}^{-1}\}^{-1},$$
$$\mu_{\bar{x}_{(v)}} = \Psi_{\bar{x}_{(v)}}(\bar{X}_{(-v)}^{\mathsf{T}}[\textstyle\sum_{k=1}^{K}\{Y_{(v)}^{(k)} - 1_{V-1} \otimes 1_n 0.5 - \Omega_{(v)}^{(k)}(1_{V-1} \otimes \mu + X_{(-v)}^{(k)}x_{(v)}^{(k)})\}]).$$

end for

---

[4] **Sample the layer-specific coordinates** $x_v^{(k)}(t_1), \ldots, x_v^{(k)}(t_n)$ **for** $v = 1, \ldots, V$ **and** $k = 1, \ldots, K$

for each layer $k = 1, \ldots, K$ **do**

  for each actor $v = 1, \ldots, V$ **do**

    Block-sample $\{x_v^{(k)}(t_1), \ldots, x_v^{(k)}(t_n)\}$ given $\{x_u^{(k)}(t_i) : u \neq v, t_i = t_1, \ldots t_n\}$. In particular, letting $x_{(v)}^{(k)} = \{x_{v1}^{(k)}(t_1), \ldots, x_{v1}^{(k)}(t_n), \ldots, x_{vH}^{(k)}(t_1), \ldots, x_{vH}^{(k)}(t_n)\}^{\mathsf{T}}$ and adapting derivations in step [3] to the sampling of the layer-specific coordinates, we obtain

$$x_{(v)}^{(k)} \mid - \sim \text{N}_{n \times H}\left(\mu_{x_{(v)}^{(k)}}, \Psi_{x_{(v)}^{(k)}}\right),$$

    with
$$\Psi_{x_{(v)}^{(k)}} = \{X_{(-v)}^{(k)\mathsf{T}}\Omega_{(v)}^{(k)}X_{(-v)}^{(k)} + \text{diag}(\tau_1^{(k)}, \ldots, \tau_H^{(k)}) \otimes \Sigma_x^{-1}\}^{-1},$$
$$\mu_{x_{(v)}^{(k)}} = \Psi_{x_{(v)}^{(k)}}[X_{(-v)}^{(k)\mathsf{T}}\{Y_{(v)}^{(k)} - 1_{V-1} \otimes 1_n 0.5 - \Omega_{(v)}^{(k)}(1_{V-1} \otimes \mu + \bar{X}_{(-v)}\bar{x}_{(v)})\}].$$

  end for

end for

---

[5] Update the gamma quantities defining the shared shrinkage parameters $\tau_1^{-1}, \ldots, \tau_R^{-1}$

$$\delta_1 \mid - \quad \sim \quad \mathrm{Ga}\left\{ a_1 + \frac{V \times n \times R}{2}, 1 + \frac{1}{2} \sum_{m=1}^{R} \theta_m^{(-1)} \sum_{v=1}^{V} \bar{x}_{vm}^{\intercal} \Sigma_{\bar{x}}^{-1} \bar{x}_{vm} \right\},$$

$$\delta_{r \geq 2} \mid - \quad \sim \quad \mathrm{Ga}\left\{ a_2 + \frac{V \times n \times (R - r + 1)}{2}, 1 + \frac{1}{2} \sum_{m=r}^{R} \theta_m^{(-r)} \sum_{v=1}^{V} \bar{x}_{vm}^{\intercal} \Sigma_{\bar{x}}^{-1} \bar{x}_{vm} \right\},$$

where $\theta_m^{(-r)} = \prod_{f=1, f \neq r}^{m} \delta_f$ for $r = 1, \ldots, R$ and $\bar{x}_{vm} = \{\bar{x}_{vm}(t_1), \ldots, \bar{x}_{vm}(t_n)\}^{\intercal}$.

---

[6] Sample the variables defining the layer-specific parameters $\tau_1^{(k)-1}, \ldots, \tau_H^{(k)-1}$, $k = 1, \ldots, K$

for each layer $k = 1, \ldots, K$ **do**

$$\delta_1^{(k)} \mid - \quad \sim \quad \mathrm{Ga}\left\{ a_1 + \frac{V \times n \times H}{2}, 1 + \frac{1}{2} \sum_{l=1}^{H} \theta_l^{(-1)} \sum_{v=1}^{V} x_{vl}^{(k)\intercal} \Sigma_x^{-1} x_{vl}^{(k)} \right\},$$

$$\delta_{h \geq 2}^{(k)} \mid - \quad \sim \quad \mathrm{Ga}\left\{ a_2 + \frac{V \times n \times (H - h + 1)}{2}, 1 + \frac{1}{2} \sum_{l=h}^{H} \theta_l^{(-h)} \sum_{v=1}^{V} x_{vl}^{(k)\intercal} \Sigma_x^{-1} x_{vl}^{(k)} \right\},$$

where $\theta_l^{(-h)} = \prod_{f=1, f \neq h}^{l} \delta_f^{(k)}$ for $h = 1, \ldots, H$ and $x_{vl}^{(k)} = \{x_{vl}^{(k)}(t_1), \ldots, x_{vl}^{(k)}(t_n)\}^{\intercal}$.

end for

---

[7] Obtain the posterior samples for the edge probabilities $\pi_{vu}^{(k)}(t_i)$ via

$$\pi_{vu}^{(k)}(t_i) = [1 + \exp\{-\mu(t_i) - \bar{x}_v(t_i)^{\intercal} \bar{x}_u(t_i) - x_v^{(k)}(t_i)^{\intercal} x_u^{(k)}(t_i)\}]^{-1},$$

for every $t_i = t_1, \ldots, t_n$, layer $k = 1, \ldots, K$ and actors $v = 2, \ldots, V$, $u = 1, \ldots, v - 1$.

---

[8] Impute the missing edges given the current state of $\pi_{vu}^{(k)}(t_i)$ from a $\mathrm{Bern}\{\pi_{vu}^{(k)}(t_i)\}$, for all the combinations of times $t_i \in \mathbb{T}$, layers $k = 1, \ldots, K$ and actors $v = 2, \ldots, V$, $u = 1, \ldots, v - 1$ corresponding to unobserved edges.

References

E.M. Airoldi, D.M. Blei, S.E. Fienberg, and E.P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

A. Banerjee, D.B. Dunson, and S.T. Tokdar. Efficient Gaussian process regression for large datasets. *Biometrika*, 100(1):75–89, 2013.

A. Bhattacharya and D.B. Dunson. Sparse Bayesian infinite factor models. *Biometrika*, 98 (2):291–306, 2011.

V.D. Blondel, A. Decuyper, and G. Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):1, 2015.

A. Canale and D.B. Dunson. Bayesian kernel mixtures for counts. *Journal of the American Statistical Association*, 106(496):1528–1539, 2011.

C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.F. Pinton, and A. Vespignani. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PloS one*, 5(7):e11596, 2010.

S. Cauchemez, A. Bhattarai, T.L. Marchbanks, R.P. Fagan, S. Ostroff, N.M. Ferguson, D. Swerdlow, and the Pennsylvania H1N1 working group. Role of social networks in shaping disease transmission during a community outbreak of 2009 h1n1 pandemic influenza. *Proceedings of the National Academy of Sciences*, 108(7):2825–2830, 2011.

H. Crane. Combinatorial Levy processes. *arXiv preprint*, arXiv:1612.05746, 2017.

H. Crane. Dynamic random networks and their graph limits. *The Annals of Applied Probability*, 26(2):691–721, 2016.

D. Durante. A note on the multiplicative gamma process. *Statistics & Probability Letters*, 122:198–204, 2017.

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

O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986.

T.M. Fruchterman and E.M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991.

S. Ghosal and A. Roy. Posterior consistency of Gaussian process prior for nonparametric binary regression. *The Annals of Statistics*, 34(5):2413–2429, 2006.

J. Ghosh and D.B. Dunson. Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18(2): 306–320, 2009.

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

F. Guo, C. Blundell, H. Wallach, and K. Heller. The Bayesian echo chamber: Modeling social influence via linguistic accommodation. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pages 315–323. Journal of Machine Learning Research, Workshops & Proceedings, 2015.

Q. Han, K. Xu, and E.M. Airoldi. Consistent estimation of dynamic and multi-layer block models. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1511–1520. Journal of Machine Learning Research, Workshops & Proceedings, 2015.

S. Hanneke, W. Fu, and E.P. Xing. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605, 2010.

M.T. Heaney. Multiplex networks and interest group influence reputation: An exponential random graph model. *Social Networks*, 36(1):66–81, 2014.

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

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

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

P.W. Holland and S. Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981.

T. House and M.J. Keeling. Household structure and infectious disease transmission. *Epidemiology and Infection*, 137(5):654–661, 2009.

M.C. Kiti, M. Tizzoni, T.M. Kinyanjui, D.C. Koech, P.K. Munywoki, M. Meriac, L. Cappa, A. Panisson, A. Barrat, C. Cattuto, and D.J. Nokes. Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors. *EPJ Data Science*, 5 (1):1, 2016.

S. Lee and P. Monge. The coevolution of multiplex communication networks in organizational communities. *Journal of Communication*, 61(4):758–779, 2011.

K. Leetaru and P.A. Schrodt. GDELT: Global data on events, location, and tone, 1979–2012. In *International Studies Association Annual Convention*, 2013.

M. Marathe and N. Ramakrishnan. Recent advances in computational epidemiology. *IEEE Intelligent Systems*, 28(4):96–101, 2013.

I. Murray and R.P. Adams. Slice sampling covariance hyperparameters of latent Gaussian models. In *Advances in Neural Information Processing Systems*, pages 1732–1740, 2010.

K. Nowicki and T.A.B. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

B. Oselio, A. Kulesza, and A.O. Hero. Multi-layer graph analysis for dynamic social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):514–523, 2014.

N.G. Polson, J.G. Scott, and J. Windle. Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.

C.E. Rasmussen and C.K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

A. Schein, M. Zhou, D. Blei, and H. Wallach. Bayesian Poisson Tucker decomposition for learning the structure of international relations. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 2810–2819. Journal of Machine Learning Research, Workshops & Proceedings, 2016.

T.A.B. Snijders, G.G. van de Bunt, and C.E.G. Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60, 2010.

T.A.B. Snijders, A. Lomi, and V.J. Torló. 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.

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

S.Y. Tan, J. Wu, L. Lü, M.J. Li, and X. Lu. Efficient network disintegration under incomplete information: The comic effect of link prediction. *Scientific Reports*, 6:1–9, 2016.

P. Vanhems, A. Barrat, C. Cattuto, J.F. Pinton, N. Khanafer, C. Régis, B. Kim, B. Comte, and N. Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS one*, 8(9):e73970, 2013.

E.P. Xing, W. Fu, and L. Song. A state-space mixed membership blockmodel for dynamic network tomography. *The Annal of Applied Statistics*, 4(2):535–566, 2010.

T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin. Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Machine Learning*, 82(2):157–189, 2011.

B. Zhu and D.B. Dunson. Locally adaptive Bayes nonparametric regression via nested Gaussian processes. *Journal of the American Statistical Association*, 108(504):1445–1456, 2013.