A Bayesian Nonparametric Latent Space Approach to Modeling Evolving Communities in Dynamic Networks[*]

Joshua Daniel Loyal[†] and Yuguo Chen[‡]

Abstract. The evolution of communities in dynamic (time-varying) network data is a prominent topic of interest. A popular approach to understanding these dynamic networks is to embed the dyadic relations into a latent metric space. While methods for clustering with this approach exist for dynamic networks, they all assume a static community structure. This paper presents a Bayesian nonparametric model for dynamic networks that can model networks with evolving community structures. Our model extends existing latent space approaches by explicitly modeling the additions, deletions, splits, and mergers of groups with a hierarchical Dirichlet process hidden Markov model. Our proposed approach, the hierarchical Dirichlet process latent position cluster model (HDP-LPCM), incorporates transitivity, models both individual and group level aspects of the data, and avoids the computationally expensive selection of the number of groups required by most popular methods. We provide a Markov chain Monte Carlo estimation algorithm and demonstrate its ability to detect evolving community structure in a network of military alliances during the Cold War and a narrative network constructed from the Game of Thrones television series.

Keywords: longitudinal networks, mixture model, nonparametric Bayes, social networks, statistical network analysis.

1 Introduction

Many naturally occurring networks contain discrete changes in community structure. When high school students move across the country for college, old friendship groups often dissolve, which leads to the formation of new collegiate friendship groups. After World War II, the Eastern and Western Blocs emerged and dominated the network of global alliances. However, after the fall of the Soviet Union, these blocs reshuffled into new political alliances. When exposed to external stimuli, regions of the brain activate before becoming dormant. By identifying these community-level phase changes, we can gain valuable insight into the rich processes that generate dynamic (longitudinal or time-varying) networks.

In this work, we address this problem of inferring discrete changes in a network's community structure. See Figure 1 for a concrete example. In this case, a dynamic
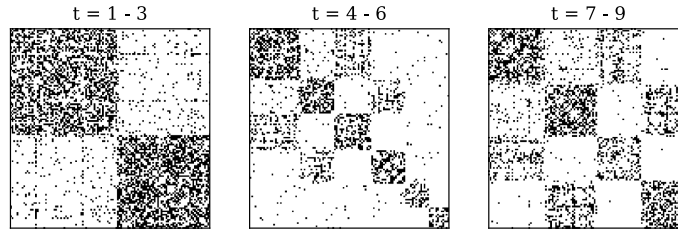
Figure 1: Adjacency matrices from a dynamic sequence of networks that exhibit an evolving community structure. At $t = 4$ the groups split from two into six groups, then at $t = 7$ the groups merge into four communities. The rows and columns of each adjacency matrix have been sorted according to the ground truth labels. These networks were simulated according to the procedure described in Section 5.4.

network's community structure contains two groups during the first three time points, splits into six groups at the fourth time point, and then merges into four groups during the seventh time point. These additions, deletions, splits, and mergers of groups in dynamic networks is our primary object of interest. For brevity, we refer to these group-level dynamics as a network's *evolving community structure.* Note that this does not refer to individual actors (or nodes) moving between existing communities. While we will model such actor-level dynamics, our primary goal is to infer changes in community structure that occur on a larger macro scale.

We adopt the latent space approach to network modeling to infer an evolving community structure in dynamic networks. Initially developed in Hoff et al. (2002), latent space models (LSMs) embed actors within a Euclidean space (the distance model) or a hypersphere (the projection model). Closeness in the latent space increases the probability that the two actors form an edge in the observed network. For this reason, one interprets the proximity of two actors in the latent space as an indication that they have similar characteristics. The LSM's popularity stems from the intuitive meaning of the embeddings and its ability to naturally incorporate desirable sociological features such as homophily, reciprocity, and transitivity. Sarkar and Moore (2005) and Sewell and Chen (2015) extended the distance model to undirected and directed dynamic networks, respectively.

The latent space model was first applied to the community detection problem in the case of a single static network by Handcock et al. (2007). Their proposed latent position cluster model (LPCM) uses a Gaussian mixture model to cluster the latent positions embedded according to the Euclidean distance model. This approach assumes that a cluster of actors in the latent space corresponds to a densely connected community in the network. Sewell and Chen (2017) extended the LPCM to dynamic networks in which the number of communities does not change over time. Their approach allows the actors to transition between these fixed communities by endowing the actor's latent trajectories with autoregressive hidden Markov model (AR-HMM) dynamics.

The following two limitations make the existing LPCM for dynamic networks inadequate for inferring an evolving community structure: (1) the number of communities

must remain fixed over time, and (2) the number of communities must be set a priori. The first shortcoming implies that the existing LPCM cannot allow the number of communities to change over time so that inference over evolving communities is not possible. The second shortcoming means that uncertainty in the community structure is not assessed in a fully Bayesian way, which is essential in most applications where multiple clusterings fit the data well. The inability of LPCMs to handle time-varying groups in dynamic networks motivates our investigation into a suitable Bayesian methodology.

In this work, we propose a Bayesian nonparametric model that addresses the shortcomings outlined in (1) and (2). As in Sewell and Chen (2017), we model the actor's trajectories through the latent space as a collection of independent and identically distributed (iid) AR-HMMs with Gaussian emissions. However, unlike their approach, we remove the restriction that the hidden label Markov chain is homogeneous. Instead, we propose a model with time-inhomogenous transition distributions. This modification provides us with the flexibility to model the additions, deletions, splits, and mergers of communities. Furthermore, we place a hierarchical Dirichlet process (HDP) prior on these time-inhomogenious transition distributions. This nonparametric prior produces fully Bayesian inference over the number of communities. Since our approach combines the advantages of the HDP with the LPCM for modeling dynamic networks, we refer to our model as the hierarchical Dirichlet process latent position cluster model (HDP-LPCM). To the best of our knowledge, there is no other latent space approach that accomplishes our modeling criteria.

The rest of this paper is organized as follows. Section 2 reviews other Bayesian nonparametric approaches for modeling evolving communities. In Section 3, we describe the hierarchical Dirichlet process latent position cluster model (HDP-LPCM), our proposed Bayesian nonparametric method for modeling evolving communities in dynamic networks. In Section 4, we outline our Metropolis-Hastings within Gibbs estimation procedure and our methodology for posterior summarization. We demonstrate the empirical performance of our proposed method through various simulation studies in Section 5. In Section 6, we use our model to analyze two real-world dynamic networks: a network of military alliances during the Cold War (Gibler, 2009b) and a narrative network constructed from the Game of Thrones television series (Beveridge and Chemers, 2018). Section 7 contains a discussion.

2   Related Methods

Although the latent space methodology has not previously incorporated an evolving community structure, the statistical modeling of varying communities has a long history in the broader network literature. For brevity, we focus on Bayesian nonparametric approaches because they are most related to our work. We restrict the discussion to the three most popular statistical network models: stochastic block models (SBMs), mixed-membership stochastic block models (MMSBMs), and latent feature models.

For stochastic block models, Kemp et al. (2006) introduced the infinite relational model (IRM), which uses a Dirichlet process to infer the number of blocks in the traditional stochastic block model. Ishiguro et al. (2010) extended this work to longitudinal

networks in their dynamic infinite relational model (dIRM). Similar to our work, the dIRM assumes the block memberships form a time-inhomogeneous Markov chain. For multiplex networks, Paez et al. (2019) proposed the Hierarchical SBM (HSBM), which uses an HDP to pool community structure across layers. Although the authors apply the HSBM to dynamic networks, the model fails to capture essential autocorrelations in an actor's community assignments by assuming they are independent across time.

In the mixed-membership literature, Fan et al. (2015) introduced the dynamic infinite mixed-membership stochastic block model. Their model also utilizes the hierarchical Dirichlet process framework to extend the original mixed-membership model (Airoldi et al., 2008; Fu et al., 2009; Ho et al., 2011) to the dynamic setting. In contrast to our method, they do not assume an HMM structure and only use the HDP to re-sample the mixed-membership vectors at every time step.

Finally, in the latent feature modeling literature, Kim and Leskovec (2013) introduced the nonparametric multi-group membership model. They utilize a distance-dependent Indian buffet process (dd-IBP) that assigns each actor a latent binary feature vector. This nonparametric prior can infer an evolving number of binary features.

While some of these competing approaches have addressed shortcomings (1) and (2), LPCMs have many appealing advantages that make our extension worthwhile. Foremost is the LPCM's ability to capture network structures on multiple scales. Unlike SBMs, where the same within and between block probabilities are shared by all actors in a group, LPCMs allow for heterogeneous connectivity patterns due to an actor's local position. Furthermore, unlike other methods such as MMSBMs, the latent space provides an interpretable visualization of the entire network. These visualizations allow domain experts to make essential assessments and critiques of the statistical methodology.

3   The HDP Latent Position Cluster Model

We consider binary relational data between n individuals recorded over T time periods. These relations are collected in a sequence of $n \times n$ binary adjacency matrix Y_t, $t = 1, \ldots, T$, where the entries Y_{ijt} indicate the presence ($Y_{ijt} = 1$) or absence ($Y_{ijt} = 0$) of an edge between individuals i and j at time t. For clarity, we only consider undirected random graphs without self-loops, so that Y_t is a symmetric matrix with zeros on the diagonal. The extension of our model to directed networks or networks with weighted edges is straightforward and discussed in Section 7.

To each of the n individuals, we associate a latent (unobserved) position that may vary through time in a p-dimensional Euclidean latent space. We represent the latent position of individual i at time t with the vector-valued random variable $\mathbf{X}_t^i \in \mathbb{R}^p$. In addition, we collect a snapshot of all individual latent positions at time t in the $n \times p$ matrix $\mathcal{X}_t = (\mathbf{X}_t^1, \ldots, \mathbf{X}_t^n)'$. Like the traditional LPCM, we assign each actor to a latent group at each time point. Note that their assignment may change over time. However, unlike the LPCM, we assume that the number of groups changes over time to accommodate the network's evolving community structure. We use G to denote the total number of groups in the network over all observational periods. We refer to a

group that contains at least one actor at time t as an *active group at time t*. We use $G_t \subseteq \{1, \ldots, G\}$ to indicate the set of all active groups at time t. Note that under the assumption of an evolving community structure, G_t is a random set and may grow or shrink over time. We encode the latent group membership of individual i at time t with the integer-valued random variable $Z_t^i \in G_t$. The collection of group assignments for all individuals at time t is summarized by the n-dimensional vector $\mathcal{Z}_t = (Z_t^1, \ldots, Z_t^n)'$.

Common to the latent space literature, we assume that the latent community labels only play a role in the distribution of the latent positions. Formally, we assume that the formation of edges in the dynamic networks $Y_{1:T}$ is conditionally independent of the actor labels $\mathcal{Z}_{1:T}$ given the latent positions $\mathcal{X}_{1:T}$, i.e., $Y_{1:T} \perp \mathcal{Z}_{1:T} \mid \mathcal{X}_{1:T}$. Note that throughout the remainder of this work, we will use the notation $A_{1:K}$ to refer to the sequence $(A_1, A_2, \ldots, A_K)$. This allows us to decompose the joint probability distribution as follows:

$$\mathbb{P}(Y_{1:T}, \mathcal{X}_{1:T}, \mathcal{Z}_{1:T}) = \mathbb{P}(Y_{1:T} \mid \mathcal{X}_{1:T}) \, \mathbb{P}(\mathcal{X}_{1:T}, \mathcal{Z}_{1:T}). \tag{1}$$

This independence assumption says that the probability of a tie is solely determined by the underlying latent positions of the actors. This decomposition is consistent with the latent space clustering idea described earlier. Specifically, the notion that an underlying cluster of actors in the latent space results in observed communities. We believe such a generative model is natural for modeling communities in networks. For example, friend groups often form due to their members having similar interests or characteristics.

3.1   The Euclidean Distance Model

As in Hoff et al. (2002), we posit that the probability of an edge forming between actors increases as the Euclidean distance between actors decreases. Let $d_{ijt} = \|\mathbf{X}_t^i - \mathbf{X}_t^j\|_2$ denote the Euclidean distance between actors i and j at time t. A conventional link between the conditional probability of forming an edge and d_{ijt} is the logistic regression model:

$$\mathbb{P}(Y_{ijt} = 1 \mid \mathcal{X}_t, \boldsymbol{\psi}) = \text{logit}^{-1}(\eta_{ijt}) = \frac{\exp(\eta_{ijt})}{1 + \exp(\eta_{ijt})}, \tag{2}$$

where η_{ijt} is a linear predictor that depends on the distances d_{ijt} and the vector $\boldsymbol{\psi}$ holds any additional parameters. A sequence of conditional independence assumptions results in the full network likelihood. First, we assume that the longitudinal networks are conditionally independent given the latent positions, i.e., $Y_t \perp Y_s \mid \mathcal{X}_{1:T}$ for all $s, t \in \{1, \ldots, T\}$. Second, we posit that edges between actors form independently conditioned on their latent positions at each time point. Under these assumptions, the likelihood of the adjacency matrices factors as a product over the networks at each time point and the set of dyads $\mathcal{D}$:

$$\mathbb{P}(Y_{1:T} \mid \mathcal{X}_{1:T}, \boldsymbol{\psi}) = \prod_{t=1}^{T} \prod_{(i,j) \in \mathcal{D}} \mathbb{P}(Y_{ijt} = y_{ijt} \mid \mathcal{X}_t, \boldsymbol{\psi}) = \prod_{t=1}^{T} \prod_{(i,j) \in \mathcal{D}} \frac{\exp(y_{ijt} \eta_{ijt})}{1 + \exp(\eta_{ijt})}. \tag{3}$$

In this work, we focus on undirected networks without self-loops for which

$$\mathcal{D} = \{(i, j) : 1 \leq i \leq n, j < i\} \tag{4}$$

and the linear predictor takes the form

$$\eta_{ijt} = \beta_0 - d_{ijt} \tag{5}$$

so that $\boldsymbol{\psi} = \{\beta_0\}$. Various versions of this likelihood have been proposed, such as in Sarkar and Moore (2005), Krivitsky et al. (2009), and Sewell and Chen (2015). The intercept parameter β_0 is sometimes referred to as the abundance. This is because higher values of β_0 result in a higher probability of forming edges.

3.2   An AR-HMM for Latent Space Dynamics

Recall that our goal is to endow the dynamics of the latent space with a probabilistic structure that allows for inference over an evolving collection of communities. To accomplish this goal, we must place an adequately flexible joint distribution over the latent space, $\mathbb{P}(\mathcal{X}_{1:T}, \mathcal{Z}_{1:T})$. To begin, we adopt the assumptions taken by Sewell and Chen (2017). Specifically, we assume that an actor's community assignments form a Markov chain, i.e.,

$$Z_t^i \mid Z_1^i, \ldots, Z_{t-1}^i \overset{d}{=} Z_t^i \mid Z_{t-1}^i, \tag{6}$$

where $\overset{d}{=}$ denotes equality in distribution. Similarly, we assume an actor's latent position follows a Markov process with an additional dependence on the current community assignment, i.e.,

$$\mathbf{X}_t^i \mid \mathbf{X}_1^i, \ldots, \mathbf{X}_{t-1}^i, Z_1^i, \ldots, Z_{t-1}^i, Z_t^i \overset{d}{=} \mathbf{X}_t^i \mid \mathbf{X}_{t-1}^i, Z_t^i. \tag{7}$$

These two assumptions allow us to factor the marginal density of an individual actor's trajectory. In particular, we conclude that an actor's trajectory follows an independent autoregressive hidden Markov model (AR-HMM):

$$p(\mathbf{X}_{1:T}^i, Z_{1:T}^i) = p(Z_1^i)p(\mathbf{X}_1^i \mid Z_1^i) \prod_{t=2}^{T} p(Z_t^i \mid Z_{t-1}^i)p(\mathbf{X}_t^i \mid \mathbf{X}_{t-1}^i, Z_t^i), \tag{8}$$

where $p(Z_1^i)$ is the initial distribution over actor i's initial community assignment, $p(Z_t^i \mid Z_{t-1}^i)$ is the transition distribution between actor i's label assignment at time $t-1$ and time t, and $p(\mathbf{X}_t^i \mid \mathbf{X}_{t-1}^i, Z_t^i)$ is the emission distribution for actor i's latent position at time t, which we allow to depend on the previously emitted value $\mathbf{X}_{t-1}^i$.

To properly model evolving communities, we make an important departure from previous dynamic latent space models. In particular, we expand the probabilistic model to include time-inhomogeneous Markov chains where the transition matrix $p(Z_t^i \mid Z_{t-1}^i)$ can vary over time. This structure contrasts with previous methods, which assume that an actor's probability of transitioning from community i to community j is the same at all times. We argue that time-inhomogeneous transitions are essential characteristics of evolving communities. For example, the addition of a group requires the transition matrices to add a non-zero probability of transitioning into that new group. Furthermore, once the new group appears in the network, an actor's probability of transitioning

into that group may approach a different steady-state from when it initially appeared. Although models that utilize homogeneous Markov chains may infer changing group structure by including empty clusters, their inferences will undoubtedly be biased due to smoothing over sudden group structure changes.

Tractable inference over these time-inhomogeneous latent trajectories is made possible by tying together the transition distributions and emission distributions of every actor. We collect the transition probabilities into a single row-stochastic transition matrix Π_t for each time step $t = 2, \ldots, T$. Each entry is defined as $(\Pi_t)_{jk} = \pi_{jk}^t = p(Z_t^i = k \mid Z_{t-1}^i = j)$, which is the same for all n actors. In what follows, it will be useful to isolate the j-th row of Π_t in the vector $\boldsymbol{\pi}_j^t$. Furthermore, we refer to $\{\boldsymbol{\pi}_j^t\}$ as the set of rows of the transition matrix at time t. For simplicity, we assume that all actors share a common initial state $Z_0^i = 0$. This allows us to view the initial distribution of the Markov chain as a special transition distribution denoted by $\boldsymbol{\pi}_0^1 = (\pi_{01}^1, \ldots, \pi_{0\,G_0}^1)$, where $\pi_{0k}^1 = p(Z_1^i = k \mid Z_0^i = 0) = p(Z_1^i = k)$ is equal for all actors.

Conditioned on the initial and transition probability matrices and the parameters of the emission distributions $\boldsymbol{\theta}$, the data generating process of the actors' trajectories is characterized by a collection of iid AR-HMMs. In this work, we model the joint density over all actors' trajectories as

$$p(\mathcal{X}_{1:T}, \mathcal{Z}_{1:T} \mid \Pi_{2:T}, \boldsymbol{\pi}_0^1, \boldsymbol{\theta}) = \prod_{i=1}^n p(\mathbf{X}_{1:T}^i, Z_{1:T}^i \mid \Pi_{2:T}, \boldsymbol{\pi}_0^1, \boldsymbol{\theta}) \tag{9}$$

$$= \prod_{i=1}^n \pi_{0,Z_1^i}^1 \, N(\mathbf{X}_1^i \mid \boldsymbol{\mu}_{Z_1^i}, \sigma_{Z_1^i}^2 I_p) \times \prod_{t=2}^T \pi_{Z_{t-1}^i, Z_t^i}^t \, N(\mathbf{X}_t^i \mid \lambda \boldsymbol{\mu}_{Z_t^i} + (1 - \lambda)\mathbf{X}_{t-1}^i, \sigma_{Z_t^i}^2 I_p),$$

where I_p is the $p \times p$ identity matrix, and $N(\mathbf{X} \mid \boldsymbol{\mu}, \Sigma)$ is the normal density with mean vector $\boldsymbol{\mu}$ and covariance Σ evaluated at the point $\mathbf{X}$. Note that in this case $\boldsymbol{\theta} = \{\{\boldsymbol{\mu}_g, \sigma_g^2\}_{g=1}^G, \lambda\}$. Like the clustering model in Handcock et al. (2007), the communities are modeled as a multivariate normal distribution in the latent space with mean location $\boldsymbol{\mu}_g$, and spherical covariance $\sigma_g^2 I_p$. As in the longitudinal clustering approach of Sewell et al. (2016), the mean position $\mathbf{X}_t^i$ is equal to $\lambda \boldsymbol{\mu}_g + (1 - \lambda)\mathbf{X}_{t-1}^i$ where $\lambda \in (0, 1)$. This is a blend between the actor's previous position and the current assigned group location. Consequently, λ is called the blending coefficient. Note that we exclude $\lambda = 0$ so that latent positions remain clustered around the group centers in latent space. This behavior is necessary to link the latent space's evolving cluster structure to the observed network's evolving community structure. Furthermore, we exclude $\lambda = 1$ so that the model accounts for the usually strong autocorrelation in the latent positions, which improves forecasting performance.

In addition, we kept $\lambda \in (0, 1)$ to allow the model to capture slowly merging (or splitting) communities, which is not possible when $\lambda = 0$ or 1. Specifically, $\lambda \in (0, 1)$ induces a biased random walk that encourages actors to travel along the line connecting their current latent position to their assigned group center $\boldsymbol{\mu}_g$. For example, consider two communities that slowly merge into one, e.g., a dynamic friendship network between students from two middle schools that slowly exhibits more across school friendships

as students begin to attend a common high school. Such dynamics are described by a latent space initially populated by two communities, corresponding to the two middle schools, and the emergence of a third community, corresponding to the high school, at their midpoint. As students are assigned to the high school community, the high school's group center acts as a sink where actors from the two middle school clusters slowly wander over to the third community's center. The size of λ controls the speed of the merge with smaller values resulting in slower merging behavior. Although the main manuscript primarily analyzes abrupt changes in community structure, where λ is near one, we include a slow merging event in Section S.8 of the Supplementary Material (Loyal and Chen, 2022a).

To complete the model, we must specify a prior on the initial distribution $\boldsymbol{\pi}_0^1$, the transition probabilities $\Pi_{2:T}$, and the parameters of the emission distribution $\boldsymbol{\theta} = \{\{\boldsymbol{\mu}_g, \sigma_g^2\}_{g=1}^G, \lambda\}$. Recall that we account for evolving community structure by allowing the transition probabilities' dimensionality to change over time. Furthermore, we refrain from specifying G, the total number of groups, or G_t, the active groups at time t, which means that the prior must quantify their uncertainty. To satisfy these requirements, we turn to the flexibility afforded by a Bayesian nonparametric approach.

3.3   Background: Dirichlet Processes and the HDP-HMM

The Dirichlet process (DP) is a distribution over discrete probability measures F_0:

$$\boldsymbol{\beta} \sim \text{GEM}(\gamma), \qquad \boldsymbol{\theta}_k \overset{\text{iid}}{\sim} H, \qquad F_0 = \sum_{k=1}^{\infty} \beta_k \, \delta_{\boldsymbol{\theta}_k}, \tag{10}$$

where $\boldsymbol{\beta} \sim \text{GEM}(\gamma)$ denotes the following sticking-breaking process (Sethuraman, 1994):

$$\beta_k = w_k \prod_{i=1}^{k-1} (1 - w_i), \qquad w_k \sim \text{Beta}(1, \gamma), \qquad k = 1, 2, \dots, \tag{11}$$

and $\delta_{\boldsymbol{\theta}_k}$ is a point mass at $\boldsymbol{\theta}_k$. In effect, $\boldsymbol{\beta} \sim \text{GEM}(\gamma)$ divides a unit-length stick into pieces with lengths given by the weights β_k where the k-th weight is a random proportion w_k of the remaining stick length $\prod_{i=1}^{k-1}(1 - w_i)$. We use the notation $F_0 \sim \text{DP}(\gamma, H)$ to indicate draws from a DP with concentration parameter γ and base measure H. Since draws from a DP are discrete with probability one, the DP cannot be used as a general nonparametric prior over continuous densities. To extend the DP to continuous density estimation, one uses F_0 as a mixing measure over some parametric class of distributions $f_{\boldsymbol{\theta}}$. This construction is known as the DP mixture model. The sampling process begins by drawing indicator variables $z_i \overset{\text{iid}}{\sim} \boldsymbol{\beta}$, and generating observations as $y_i \sim f_{\boldsymbol{\theta}_{z_i}}$. Note that we adopt the convention that if a K-dimensional vector $\mathbf{v}$ belongs to the K simplex, i.e., $\sum_{j=1}^K v_j = 1$ and each entry $v_j \geq 0$, then $z \sim \mathbf{v}$ means $z \sim \sum_{j=1}^K v_j \, \delta(z, j)$, where $\delta(i, j)$ is the Kronecker delta. Furthermore, the same convention applies when K is infinite, which is the case for $\boldsymbol{\beta}$.

The hierarchical Dirichlet process (HDP) is a distribution over a collection of discrete probability measures, $\{F_j : j = 1, \ldots, J\}$, that share a common set of atoms:

$$\boldsymbol{\beta} \sim \text{GEM}(\gamma),$$

$$\boldsymbol{\pi}_j \stackrel{\text{iid}}{\sim} \text{DP}(\alpha, \boldsymbol{\beta}), \qquad \boldsymbol{\theta}_k \stackrel{\text{iid}}{\sim} H, \qquad F_j = \sum_{k=1}^{\infty} \pi_{jk}\, \delta_{\boldsymbol{\theta}_k}, \qquad j = 1, \ldots, J. \tag{12}$$

The group-specific distributions of the HDP, $\boldsymbol{\pi}_j$, are often used as priors over the rows of an infinite-dimensional transition matrix, i.e., $(\Pi)_{jk} = \pi_{jk}$ for $j \in \mathbb{N}$ so that number of states J is infinite. This formulation allows for the construction of the hierarchical Dirichlet process hidden Markov model (HDP-HMM) (Teh et al., 2006), which is a natural Bayesian nonparametric extension of the Bayesian HMM (Robert et al., 2000). The sampling mechanism of the HDP-HMM proceeds as follows: one samples the hidden states sequentially as $z_t \sim \boldsymbol{\pi}_{z_{t-1}}$, and the observations are linked to the global parameters via $y_t \sim f_{\boldsymbol{\theta}_{z_t}}$. Note that the sharing of atoms induced by the HDP prior allows the Markov chain to utilize a single global set of parameters, $\boldsymbol{\theta}_k$, at all time points.

Despite its popularity the original HDP-HMM struggles to model Markov chains with long state durations. To remedy this issue Fox et al. (2011b) introduced the sticky hierarchical Dirichlet process (sticky HDP):

$$\boldsymbol{\beta} \sim \text{GEM}(\gamma), \tag{13}$$

$$\boldsymbol{\pi}_j \stackrel{\text{iid}}{\sim} \text{DP}(\alpha + \kappa, \frac{\alpha\boldsymbol{\beta} + \kappa\boldsymbol{\delta}_j}{\alpha + \kappa}), \qquad \boldsymbol{\theta}_k \stackrel{\text{iid}}{\sim} H, \qquad F_j = \sum_{k=1}^{\infty} \pi_{jk}\, \delta_{\boldsymbol{\theta}_k}, \qquad j = 1, \ldots, J\,,$$

where $\boldsymbol{\delta}_j$ is a vector of zeros except for a single one at the j-th index. Analogous to the HDP, the sticky HDP can be used as a prior over the transition matrices of an HMM. In this case, the extra stickiness parameter $\kappa > 0$ biases the process towards self-transitions. As a result, Fox et al. (2011b) found that the corresponding sticky HDP-HMM better models the longer state durations found in real-world applications. For this reason, we use the sticky HDP as a prior over the transition matrices in Equation (9).

3.4 An HDP Prior for Time-Inhomogeneous Markov Chains

Now, we present our extension to the latent position cluster model that can infer an evolving community structure in dynamic networks. To accomplish this goal, we place the following Bayesian nonparametric prior on the transition matrices in Equation (9):

$$\boldsymbol{\beta} \sim \text{GEM}(\gamma),$$

$$\boldsymbol{\pi}_0^1 \stackrel{\text{iid}}{\sim} \text{DP}(\alpha_0, \boldsymbol{\beta}), \tag{14}$$

$$\boldsymbol{\pi}_g^t \stackrel{\text{iid}}{\sim} \text{DP}(\alpha + \kappa, \frac{\alpha\boldsymbol{\beta} + \kappa\boldsymbol{\delta}_g}{\alpha + \kappa}), \qquad (\boldsymbol{\mu}_g, \sigma_g^2) \stackrel{\text{iid}}{\sim} H_{\boldsymbol{\mu}} \times H_{\sigma^2}, \qquad g = 1, 2, \ldots, \quad t = 2, \ldots, T.$$

In this work, we take $H_{\boldsymbol{\mu}}$ to be a multivariate normal distribution and H_{σ^2} to be an inverse gamma distribution, i.e., $\boldsymbol{\mu}_g \stackrel{\text{iid}}{\sim} N(\boldsymbol{\mu}_0, \tau^2 I_p)$ and $\sigma_g^2 \stackrel{\text{iid}}{\sim} \Gamma^{-1}(a/2, b/2)$. For the initial distribution, we set $\kappa = 0$ and allocate it a separate concentration parameter α_0 to

distinguish it from the other transition distributions. Note that, unlike the traditional HDP-HMM, our model re-samples the rows of the transition matrix, $\boldsymbol{\pi}_g^t$, at each time step. This difference allows us to infer time-inhomogeneous Markov chains. As we previously argued, this time-inhomogeneity is crucial for modeling evolving communities. In short, our model extends the dynamic Euclidean distance model proposed in Sewell and Chen (2017) by incorporating time-inhomogeneous transition matrices that are inferred through a nonparametric HDP prior. For this reason, we refer to our model as the hierarchical Dirichlet process latent position cluster model (HDP-LPCM). A graphical representation of the HDP-LPCM is depicted in Figure 2.
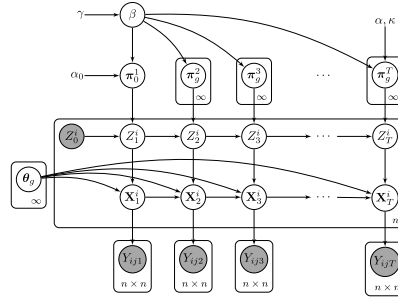

Figure 2: The graphical model for the HDP-LPCM. For clarity, we only display the dependence on α and κ for the last transition matrix, although it is present in all transition matrices from $t = 2, \ldots, T$.

Although our prior borrows heavily from the sticky HDP-HMM formulation, there are three key differences between the state-space model utilized in the HDP-LPCM and the traditional sticky HDP-HMM. First, the emitted values $\mathbf{X}_t^i$ have an autoregressive dependence. This modification was originally explored outside the network setting by Fox et al. (2011a) to model switching linear dynamical systems with Gaussian observations. However, in our case, the observations are a sequence of binary adjacency matrices. The second difference is that the HDP-HMM contains a single state sequence $Z_{1:T}$, while we infer multiple state sequences $Z_{1:T}^i$ for $i = 1, \ldots, n$ from a single observed sequence of networks. The final departure from the HDP-HMM is the re-sampling of the rows of the transition matrix $\{\boldsymbol{\pi}_g^t\}$ at every time step. This difference introduces time-inhomogeneity into the Markov chain, which is crucial for modeling the additions, deletions, splits, and mergers of groups found in evolving communities.

In addition to properly modeling an evolving community structure, the other important accomplishment of our model is its ability to infer the number of communities from the data in a fully Bayesian way. A typical parametric approach (specifying G_t at each time step) would require comparing a combinatorial amount of models. This task is computationally infeasible for even a small number of groups and time steps. Through a nonparametric prior, we naturally incorporate model selection into our inference procedure, which avoids the computationally expensive model selection step found in many LPCMs (Handcock et al., 2007; Sewell and Chen, 2017).

3.5 The Weak-Limit Approximation

For inference, we utilize the weak-limit approximation to the HDP proposed in Fox et al. (2011b). The approximation replaces the infinite dimensional DPs with finite L-dimensional Dirichlet distributions as follows:

$$\boldsymbol{\beta} \sim \text{Dirichlet}(\gamma/L, \ldots, \gamma/L),$$

$$\boldsymbol{\pi}_0^1 \overset{\text{iid}}{\sim} \text{Dirichlet}(\alpha_0\beta_1, \ldots, \alpha_0\beta_L), \quad \boldsymbol{\pi}_g^t \overset{\text{iid}}{\sim} \text{Dirichlet}(\alpha\beta_1, \ldots, \alpha\beta_g + \kappa, \ldots, \alpha\beta_L), \quad (15)$$

$$\boldsymbol{\mu}_g \overset{\text{iid}}{\sim} N(\mathbf{0}, \tau^2 I_p), \qquad \sigma_g^2 \overset{\text{iid}}{\sim} \Gamma^{-1}(a/2, b/2), \quad g = 1, \ldots, L, \quad t = 2, \ldots, T.$$

Practically, the weak limit approximation transforms the infinite transition matrices into finite $L \times L$ matrices. The parameter L gives us control over the approximation with the accuracy increasing as L goes to infinity (Ishwaran and Zarepour, 2000). In practice, one initially sets L larger than some a priori upper limit on the number of communities and verifies this choice by checking that the posterior puts negligible mass on models with L groups.

4 Estimation

We take a Bayesian approach to estimation. In what follows, we describe a Markov chain Monte Carlo (MCMC) method to sample from the HDP-LPCM's posterior. We implement a Metropolis-Hastings within Gibbs MCMC scheme with the goal of identifying an evolving community structure consistent with the inferred posterior distribution.

4.1 Blocked Metropolis-Hastings within Gibbs Sampler

As outlined in Section 3, the joint distribution over all the variables (Equation (1)) factors as

$$\overbrace{\underbrace{p(Y_{1:T} \mid \mathcal{X}_{1:T}, \beta_0)}_{\text{Equation (3)}} \cdot \overbrace{p(\mathcal{X}_{1:T}, \mathcal{Z}_{1:T} \mid \Pi_{2:T}, \boldsymbol{\pi}_0^1, \boldsymbol{\mu}_{1:L}, \sigma_{1:L}^2, \lambda)}^{\text{Equation (9)}}}$$

$$\cdot \underbrace{p(\Pi_{2:T}, \boldsymbol{\pi}_0^1 \mid \boldsymbol{\beta}) \cdot p(\boldsymbol{\mu}_{1:L}, \sigma_{1:L}^2) \cdot p(\boldsymbol{\beta})}_{\text{Equation (15)}}. \qquad (16)$$

To complete the model we assign the following priors:

$$\beta_0 \sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2), \qquad \lambda \sim N_{(0,1)}(\mu_\lambda, \sigma_\lambda^2), \qquad (17)$$

where $N_{(0,1)}(\mu, \sigma^2)$ indicates the normal distribution with mean μ and variance σ^2 truncated to the range $(0, 1)$.

We realize these samples by following a Gibbs sampling algorithm in which we iteratively sample from the appropriate conditional distributions of $\mathcal{X}_{1:T}$, $\mathcal{Z}_{1:T}$, $\boldsymbol{\pi}_0^1$, $\Pi_{2:T}$, $\boldsymbol{\beta}$,

$\boldsymbol{\mu}_{1:L}, \sigma^2_{1:L}, \lambda$, and β_0. Due to our choice of priors, most conditional distributions are conjugate so that the Gibbs updates are derived in standard fashion. See Section S.1 of the Supplementary Material for the details on these conditional distributions. Metropolis-Hastings (MH) steps are necessary for the latent positions and the intercept. In both cases, we use a normal random walk proposal. In all experiments, we tune the proposal step sizes using an initial tuning phase so that the proposed moves have a 25%–40% acceptance rate. The only samplers that need special care are the block sampler for the actor labels $\mathcal{Z}_{1:T}$, the sampler for the global prior $\boldsymbol{\beta}$, as well as the additional hyperparameter sampling schemes. We fully derive these samplers in the Supplementary Material. In addition, we provide details of our parameter initialization scheme in Section S.7 of the Supplementary Material.

The full Metropolis-Hastings within Gibbs sampler is outlined in Algorithm 1. When describing the samplers, we use dot notation to indicate summations over an index, e.g., for a random variable w_{ab}, $w_{\cdot b} = \sum_a w_{ab}$, $w_{a \cdot} = \sum_b w_{ab}$ and $w_{\cdot\cdot} = \sum_a \sum_b w_{ab}$. In addition, we use n_{kjt} to denote the number of actors that transitioned from group k to group j at time t and n_{kt} to denote the number of actors in group k at time t.

4.2   Posterior Summarization

The selection of a single point estimate that adequately summarizes the full posterior is a challenging and open problem in Bayesian nonparametric statistics. A simple solution uses the posterior mode. However, in Bayesian nonparametric models, the MCMC chain often only visits a single partition once, which makes frequency estimates unreliable. Another approach is to choose the partition corresponding to the MAP (maximum a posteriori) estimate. A downside of the MAP approach is that it does not marginalize over the uncertainty in the partitions. As a result, MAP clustering estimates tend to over-fit the data. An ideal summarization methodology should take into account the clustering uncertainty suggested by the posterior samples.

In this work, we take a decision theoretic approach to posterior summarization. Specifically, we select the partition that minimizes the posterior expectation of an appropriately chosen loss function over dynamic clusterings. In the case of static clustering, two popular choices of loss functions are Binder's loss (Binder, 1978) and the variation of information (Meilă, 2007). Their use for posterior summarization was advocated by Lau and Green (2012) and Wade and Ghahramani (2018), respectively. Note that minimizing Binder's loss is equivalent to maximizing the Rand index, another popular measure of clustering performance. Both loss functions have the desirable property that they are metrics over the space of clusterings. Furthermore, the optimization of the expected losses only depends on the posterior co-occurrence probabilities, so avoids complications due to label switching.

To extend this approach to dynamic clusterings, we minimize the posterior expected time-averaged variation of information (VI) which averages the static VI's at each time step as follows:

$$\underset{\hat{\mathcal{Z}}_{1:T}}{\arg\min} \; \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{VI}(\mathcal{Z}_t, \hat{\mathcal{Z}}_t) \;\middle|\; Y_{1:T} \right],$$

Given the previous initial distribution $(\boldsymbol{\pi}_0^1)^{(\ell-1)}$, a set of state-specific transition probabilities $\Pi_{2:T}^{(\ell-1)}$, the global transition distribution $\boldsymbol{\beta}^{(\ell-1)}$, group parameters $(\boldsymbol{\mu}_{1:L}, \sigma_{1:L}^2)^{(\ell-1)}$, node labels $\mathcal{Z}_{1:T}^{(\ell-1)}$, latent positions $\mathcal{X}_{1:T}^{(\ell-1)}$, the likelihood specific parameters $\beta_0^{(\ell-1)}$, and blending coefficient $\lambda^{(\ell-1)}$, update the current parameters as follows:

1. Initialize current parameters to the values of the $(\ell-1)$th iteration.
2. Update latent positions $\mathcal{X}_{1:T}$ via MH with a normal random walk proposal.
3. Update β_0 via MH with a normal random walk proposal.
4. Update node labels $\mathcal{Z}_{1:T}$ using the forward-backward algorithm detailed in Algorithm S.1 in Section S.2 of the Supplementary Material.
5. Sample the auxiliary variables $\mathbf{m}, \bar{\mathbf{m}}, \mathbf{w}$ as in Algorithm S.2 in Section S.4 of the Supplementary Material.
6. Update the global transition distribution by sampling

$$\boldsymbol{\beta} \sim \text{Dirichlet}(\gamma/L + \bar{m}_{\cdot 1 \cdot}, \ldots, \gamma/L + \bar{m}_{\cdot L \cdot}).$$

7. For each $k \in \{1, \ldots, L\}$, sample a new initial distribution based on the initial assignments:
$$\boldsymbol{\pi}_0^1 \sim \text{Dirichlet}(\alpha_0 \beta_1 + n_{011}, \ldots, \alpha_0 \beta_L + n_{0L1}).$$

8. For each $k \in \{1, \ldots, L\}$ and $t \in \{2, \ldots, T\}$, sample a new transition distribution based on the sample assignments:

$$\boldsymbol{\pi}_k^t \sim \text{Dirichlet}(\alpha \beta_1 + n_{k1t}, \ldots, \alpha \beta_k + \kappa + n_{kkt}, \ldots, \alpha \beta_L + n_{kLt}).$$

9. Update cluster parameters. For each $k \in \{1, \ldots, L\}$:

$$\boldsymbol{\mu}_k \sim N(\bar{\boldsymbol{\mu}}_k, \bar{\sigma}_k^2 I_p),$$
$$\sigma_k^2 \sim \Gamma^{-1}((n_k.p + a)/2, \bar{b}/2),$$

 where $\bar{\boldsymbol{\mu}}_k, \bar{\sigma}_k^2$ are defined in Equation (S.1) and $\bar{b}$ is defined in Equation (S.2) in Section S.1 of the Supplementary Material.
10. Update blending coefficient λ:

$$\lambda \sim N_{(0,1)}(\bar{\mu}_\lambda, \bar{\sigma}_\lambda^2),$$

 where $\bar{\mu}_\lambda, \bar{\sigma}_\lambda^2$ are defined in Equation (S.3) in Section S.1 of the Supplementary Material.
11. Update remaining hyperparameters $\{\tau^2, b, \gamma, \alpha_0, \alpha, \kappa, \rho\}$ as in Algorithm S.4 in Section S.6 of the Supplementary Material.

Algorithm 1: Blocked Metropolis-Hastings within Gibbs sampler for the HDP-LPCM.

where $\mathrm{VI}(\mathcal{Z}_t, \hat{\mathcal{Z}}_t) = 2H(\mathcal{Z}_t, \hat{\mathcal{Z}}_t) - H(\mathcal{Z}_t) - H(\hat{\mathcal{Z}}_t)$ and $H(\mathcal{Z}_t, \hat{\mathcal{Z}}_t), H(\mathcal{Z}_t)$, and $H(\hat{\mathcal{Z}}_t)$ are entropies calculated using the joint and marginal distributions of the labels. Similar to Wade and Ghahramani (2018), we instead minimize the following calculable lower-bound:

$$\underset{\hat{\mathcal{Z}}_{1:T}}{\arg\min} \left\{ \sum_{t=1}^{T} \sum_{i=1}^{n} \log(\sum_{j=1}^{n} \mathbb{1}_{\{\hat{Z}_t^j = \hat{Z}_t^i\}}) - 2 \sum_{t=1}^{T} \sum_{i=1}^{n} \log(\sum_{j=1}^{n} p(Z_t^j = Z_t^i \mid Y_{1:T}) \mathbb{1}_{\{\hat{Z}_t^j = \hat{Z}_t^i\}}) \right\}.$$

Notice that this expression only depends on the posterior co-occurrence probabilities, which can be estimated from the posterior samples. We perform this optimization by searching over all posterior samples from the MCMC chain (after an appropriate burn-in). Possible ties are broken by selecting the sample with the highest likelihood (Equation (3)). This minimization does not require the partitions to be visited by the Markov chain; however, by restricting the search to the sampled partitions, we have access to the associated latent space for later visualization of the network.

4.3 Non-Identifiability of the Latent Positions

Since latent position models depend on the distance between actors, it is clear that they are invariant to translations, rotations, and reflections of the latent space. Any posterior inference that utilizes these positions must correct for such a non-identifiability. We take the approach of Procrustes matching commonly employed in the literature. This involves post-processing the samples by rotating and translating them to match a reference layout. For a detailed description of this procedure, see the original work by Hoff et al. (2002). We use the sample chosen by the procedure in Section 4.2 as the reference layout.

5 Simulation Study

We designed a simulation study to assess the HDP-LPCM's performance on synthetic networks with known community dynamics that mimic those found in real-world networks. We considered two scenarios: a sequence of networks with (1) a single static number of groups active over all time steps and (2) a number of groups that changes over time because the groups have either merged or split. We labeled these scenarios the time-homogeneous and time-inhomogeneous simulations, respectively. The Supplementary Material includes an additional simulation study that investigated how group separation and stability affect the HDP-LPCM's performance (Section S.9) and a sensitivity analysis that showed that the estimation method is unaffected by our hyperparameter choices (Section S.10).

5.1 Measuring Performance

For each simulation, we evaluated the following three tasks: (1) assigning nodes to the correct community at each time step while penalizing label switching across time steps, (2) selecting the correct number of communities at each time step, and (3) predicting

in-sample, missing, and one-step ahead dyads. We emphasize that the primary goals of the HDP-LPCM are the first two tasks which involve clustering and community number selection.

To evaluate the clustering performance outlined in task (1), we used two metrics: the adjusted Rand index (ARI) and the variation of information (VI) introduced in Section 4.2. The ARI takes values in $[0, 1]$ and measures the amount of agreement between two cluster assignments. Values close to one indicate near-identical assignments, while values near zero are roughly expected when making assignments at random. On the other hand, the VI is a true metric that takes values in $[0, \, 2\log_2(K^*)]$, where K^* is the maximum number of communities, e.g., $K^* = L$ for the HDP-LPCM. Values near zero indicate that the two cluster assignments are near identical. Unlike the ARI, the VI satisfies the triangle inequality and is symmetric in its arguments. We included the VI because we used it for posterior summarization. In the dynamic setting, we can compute at least two versions of these metrics. The first version computes the metric at each time step and averages the results. We refer to these as the average ARI and average VI. The second version considers the labels for all time steps together and computes a single metric called the global ARI and global VI. Achieving a good value of the global metrics is more challenging since it requires consistent labeling within each time step and across time steps. As such, we are primarily concerned with good global performance.

We measured predictive performance in three ways. First, to determine the goodness-of-fit of the models to the training data, we calculated the area under the receiver operating characteristic curve (AUC) based on the model's predictions of in-sample dyads. Next, we measured out-of-sample performance by evaluating the model's ability to impute missing dyads and make one-step ahead predictions. To do this, in each simulation, we held out the last network and removed 10% of the dyads from each proceeding network. Then, we obtained edge predictions for the missing dyads and one-step ahead predictive probabilities and calculated the AUC values based on the dyads involved in each task.

5.2   Competing Methods

We compared our proposed HDP-LPCM to three competing methods: two stochastic block models (SBMs) and a parametric LPCM. The first method used a single SBM fit at each time step using the variational EM algorithm with the number of communities selected by maximizing the integrated classification likelihood (ICL) criterion. To apply this method to dynamic networks, we needed to account for label switching between time steps. To do so, we combined the output sequentially using the label switching algorithm proposed in Papastamoulis and Iliopoulos (2010). The second method is the dynamic SBM introduced in Matias and Miele (2017) with the number of clusters chosen by maximizing the ICL criterion. This model assumes a fixed number of communities, so we only applied it to the homogeneous simulation. We used the `blockmodels` and `dynsbm` `R` packages to estimate the SBM and dynamic SBM, respectively. However, since these packages do not provide one-step ahead predictions or perform missing edge imputation, we did not remove any dyads from the simulated networks and only calculated the AUC based on predicting in-sample dyads for the block models.

The last competing method is the parametric LPCM proposed in Sewell and Chen (2017), which only applies to the homogeneous simulations. We replaced the directed likelihood with the undirected likelihood of this work and set the number of communities to the ground truth. Lastly, we extracted the following two estimators of the community assignments: LPCM (MAP) and LPCM (VI). As their names imply, LPCM (MAP) is the original MAP estimator proposed in Sewell and Chen (2017), and LPCM (VI) used the VI summarization procedure we introduced in Section 4.2. All LPCMs imputed missing edges and calculated one-step ahead predictions using the augmented Gibbs sampler and Monte Carlo marginalization scheme described in Sewell and Chen (2015). Note that the HDP-LPCM's forecasts assume that $\Pi_{T+1} = \Pi_T$ for simplicity.

The HDP-LPCM and the LPCMs were estimated using a Markov chain consisting of 50,000 samples. We used the initial 5,000 samples to tune the step sizes of the Metropolis-Hastings samplers. We discarded the following 10,000 samples as burn-in leaving 35,000 samples for inference. For the HDP-LPCM, we fixed the truncation level of the weak-limit approximation (Equation (15)) to $L = 10$ in all experiments.

5.3 Time-Homogeneous Simulation

This simulation contains a single set of groups and transition matrices common to all time points. Our goal is to demonstrate that the HDP-LPCM performs comparably to the parametric LPCM (Sewell and Chen, 2017) while also providing a proper assessment of uncertainty in the network's community structure. The networks were generated according to the sampling mechanisms described in Sections 3.1 and 3.2. We fixed the total number of groups to $G = 6$ at each time point. We chose the blending coefficient $\lambda = 0.8$, the dimension of the latent space $p = 2$, and the intercept $\beta_0 = 1.0$. We set the group locations to

$$(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_6) = \begin{pmatrix} -4 & 4 & -2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5 & 5 \end{pmatrix}. \tag{18}$$

We drew the group shapes σ_g^2 from a $\Gamma^{-1}(6, 2)$ distribution. These parameters induce substantial overlap between groups 1 and 3 and groups 2 and 4 in the latent space. See the second column of Figure 1 for an adjacency matrix drawn from this group structure. Next, the initial distribution $\boldsymbol{\pi}_0^1$ was drawn from a six-dimensional Dirichlet$(10, \ldots, 10)$. The six rows of the transition matrix, $\boldsymbol{\pi}_g^t$ for $g = 1, \ldots, 6$ and $t = 2, \ldots, 6$, were chosen proportional to

$$\left(\frac{1}{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_g\|}, \ldots, \frac{1}{\|\boldsymbol{\mu}_{g-1} - \boldsymbol{\mu}_g\|}, \max_{k \neq g} \left\{ \frac{\text{const}}{\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_g\|} \right\}, \frac{1}{\|\boldsymbol{\mu}_{g+1} - \boldsymbol{\mu}_g\|}, \ldots, \frac{1}{\|\boldsymbol{\mu}_G - \boldsymbol{\mu}_g\|} \right).$$

We set the constant in the above equation equal to 20, which yields self-transition probabilities ranging from 0.83 to 0.9. The latent actor positions, $\mathcal{X}_{1:T}$, and group assignments, $\mathcal{Z}_{1:T}$, were drawn from Equation (9). Finally, we generated the adjacency matrices according to Equation (2) with the linear predictor η_{ijt} given by Equation (5).

Simulation	Model	AUC (in-sample)	AUC (missing)	AUC (one-step ahead)
Homogeneous	HDP-LPCM	0.911 (0.008)	0.897 (0.011)	0.810 (0.018)
Homogeneous	LPCM (VI)	0.911 (0.008)	0.896 (0.011)	0.809 (0.018)
Homogeneous	LPCM (MAP)	0.911 (0.008)	0.896 (0.010)	0.809 (0.018)
Homogeneous	Dynamic SBM	0.889 (0.010)	–	–
Homogeneous	SBM	0.892 (0.012)	–	–
Homogeneous	Naive	–	–	0.617 (0.013)
Inhomogeneous	HDP-LPCM	0.871 (0.004)	0.856 (0.007)	0.782 (0.021)
Inhomogeneous	SBM	0.849 (0.006)	–	–
Inhomogeneous	Naive	–	–	0.603 (0.014)

Table 1: Median AUC values for all 50 time-homogeneous / time-inhomogeneous simulations. Standard deviations are displayed in parentheses. The Naive model corresponds to using the dyads in Y_T to forecast Y_{T+1}.
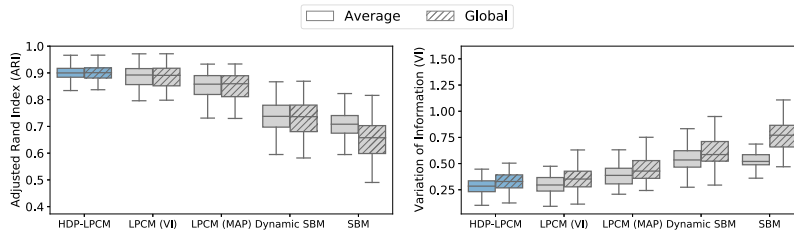
Figure 3: Boxplots of ARI (left) and VI (right) of the five competing methods on the time-homogeneous simulation. The non-hatched and hatched boxes denote the average and global metrics, respectively. The proposed method is highlighted in blue.

We simulated 50 dynamic networks with $T = 7$ time points and $n = 120$ actors each. Note that we only used the last network to evaluate one-step ahead predictions.

Table 1 contains the AUC values from this simulation. As a result of the prevalent community structure in the simulated networks, all the models have comparable in-sample predictive performances. Also, the HDP-LPCM's predictions of missing and one-step ahead dyads are on par with those of the parametric LPCMs that directly model the true data-generating process. However, the ARI and VI values in Figure 3 reveal that the block models are doing a poor job capturing the underlying community assignments. Furthermore, the static SBMs have a significantly worse global ARI and VI due to their inability to properly match the community labels across time steps. Lastly, the LPCMs that used the VI estimator substantially outperformed the LPCMs that used the MAP estimator, which validates our method of posterior summarization.

Next, we evaluated the ability of the HDP-LPCM to properly assign uncertainty to the dynamic network's community structure. Figure 4 contains boxplots of the estimated posterior probabilities of $|G_t|$ over the 50 simulations. In most cases, $|G_t| = 6$ is correctly identified with the highest posterior probability. Next, we compared the ability of our posterior summarization method to select a partition with the correct
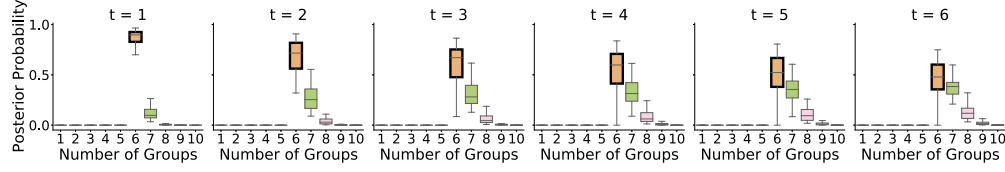
Figure 4: Boxplots showing the posterior distribution of $|G_t|$ over the 50 time-homogeneous simulations with bolded boxes indicating the true number of groups.
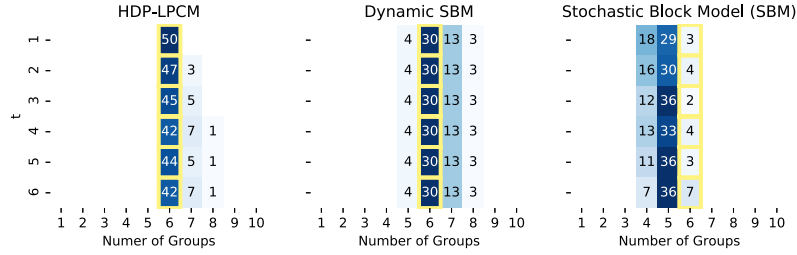

Figure 5: The number of groups estimated at each time step by the HDP-LPCM (left), dynamic SBM (center), and the static SBMs (right). The yellow cells contain the true number of communities, $|G_t| = 6$ at all time steps.

number of communities. Figure 5 displays the number of communities selected by the HDP-LPCM at each time step as well as by the SBMs which are included as a baseline. The HDP-LPCM drastically outperforms the block models, especially the static SBMs which consistently underestimate the number of groups. Overall, the HDP-LPCM has equivalent predictive and clustering performance to the parametric LPCM while also accurately estimating the true number of communities which justifies its use as an efficient drop-in replacement for the parametric LPCM.

5.4   Time-Inhomogeneous Simulation

We designed this simulation to test our model's ability to detect changes in group structure by allowing the number of groups to vary over time. We generated 50 networks with $T = 10$ time steps and $n = 120$ actors each. There are $G = 6$ groups in the networks overall. The simulations begin with two groups, these two groups split into six groups at $t = 4$, and then the six groups merge into four groups at $t = 7$. The introductory example (Figure 1) displays adjacency matrices generated from this procedure.

Once again, the networks were simulated according to the sampling mechanisms described in Sections 3.1 and 3.2 with the last network held out to evaluate one-step ahead forecasts and 10% of the dyads removed at each time step to assess dyad imputation. This simulation differs from the time-homogeneous simulation in how we specified the transition matrices $\boldsymbol{\pi}_g^t$. We set the initial distribution $\boldsymbol{\pi}_0^1 = (1/2, 1/2)^{\mathrm{T}}$. We chose the
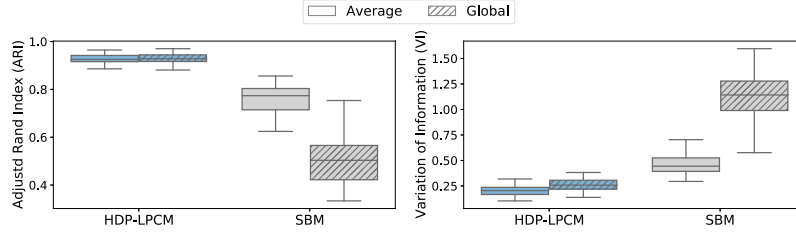
Figure 6: Boxplots of ARI (left) and VI (right) of the two competing methods on the time-inhomogeneous simulation.

rows of the transition matrix, $\boldsymbol{\pi}_g^t$ for $g \in G_{t-1}$ and $t = 2, \ldots, 10$, proportional to

$$
\begin{aligned}
&\left(\frac{\mathbb{1}_{\{1 \in G_t\}}}{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_g\|}, \ldots, \frac{\mathbb{1}_{\{g-1 \in G_t\}}}{\|\boldsymbol{\mu}_{g-1} - \boldsymbol{\mu}_g\|}, \right. \\
&\left. \max_{k \neq g \,:\, k \in G_t} \left\{ \frac{\text{const}}{\|\boldsymbol{\mu}_k - \boldsymbol{\mu}_g\|} \right\} \mathbb{1}_{\{g \in G_t\}}, \frac{\mathbb{1}_{\{g+1 \in G_t\}}}{\|\boldsymbol{\mu}_{g+1} - \boldsymbol{\mu}_g\|}, \ldots, \frac{\mathbb{1}_{\{G \in G_t\}}}{\|\boldsymbol{\mu}_G - \boldsymbol{\mu}_g\|} \right),
\end{aligned}
\tag{19}
$$

where G_t indicates the set of active groups at time t. At $t = 1, 2, 3$ there are two groups with $G_1 = G_2 = G_3 = \{1, 2\}$. From $t = 4, 5, 6$ these two groups split into six groups so that $G_4 = G_5 = G_6 = \{1, 2, 3, 4, 5, 6\}$. Finally, these six groups merge into four groups at $t = 7, 8, 9$ so that $G_7 = G_8 = G_9 = \{1, 2, 3, 4\}$. For all time points, we set the constant in Equation (19) to 20 except for $t = 4$. At $t = 4$, we set the constant equal to 1 so that the nodes were more evenly distributed among the six newly created groups. In addition, we set $\lambda = 0.9$ which resulted in stronger group cohesion and more closely matched the values found in real-world networks.

The competing methods' predictive performance on this simulation are displayed in the second half of Table 1. The in-sample performance is comparable between the HDP-LPCM and the SBM due to the strong group structure. Furthermore, the HDP-LPCM's ability to predict missing and one-step ahead dyads remains good despite the evolving community structure. Once again, the advantage of the HDP-LPCM over the block models is apparent when comparing their clustering performance in Figure 6 where the SBM has an unsatisfactorily low global ARI.

Next, we evaluate the ability of the HDP-LPCM to quantify the changing macro-level community structure. The posterior probabilities for $|G_t|$ are displayed in Figure 7. For most time points the correct number of groups has the highest median posterior probability. The exception is at $t = 3$, which is right before the two groups split into six groups, where no group number has a median posterior probability that exceeds 0.5. As is common in Dirichlet process mixture models, the model tends to create a few small clusters, which can inflate the group count. This is the case here, where the extra clusters at $t = 3$ tend to be composed of less than five nodes. Once again, we assess our posterior summarization method's ability to select a clustering with the correct number of communities. The number of groups selected by this method at each time

step is displayed in Figure 8. For the majority of simulations, our method selects the correct number of clusters. However, the method struggles at $t = 3$ which is where the posterior is the most uncertain. However, the HDP-LPCM selects the true number of clusters, $|G_3| = 2$, the majority of the time despite often assigning $|G_3| = 3$ a higher posterior probability. The HDP-LPCM is able to recover because the VI estimator can select a clustering with the small spurious communities removed when such pruning is supported by the posterior co-occurrence matrix. In contrast, the SBMs fail to estimate the true community structure for $t = 1 - 6$ in the majority of simulations. Overall, we conclude that the HDP-LPCM is able to adequately infer evolving communities.
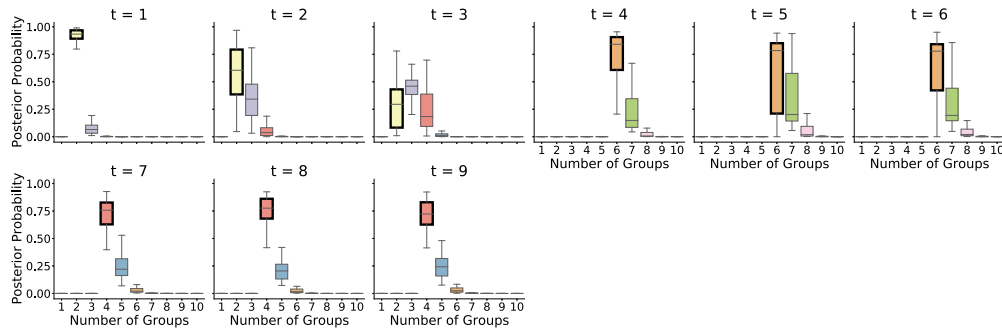
Figure 7: Boxplots showing the posterior distribution of $|G_t|$ over the 50 time-inhomogeneous simulations. Bolded boxes indicate the true number of communities.
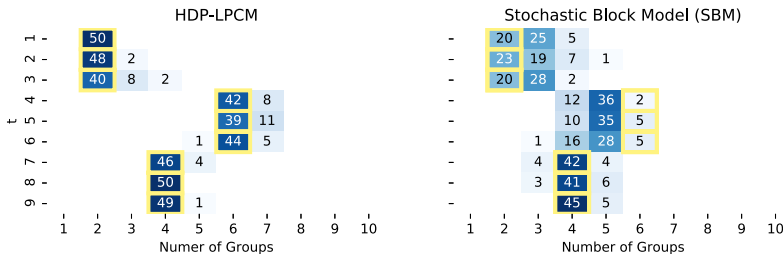

Figure 8: The number of groups estimated at each time step by the HDP-LPCM (left) and the static SBMs (right). The yellow cells contain the true number of communities.

6   Real Data Application

In this section, we demonstrate the utility of our proposed HDP-LPCM on a variety of real-world dynamic networks with an evolving community structure. We include two applications: inferring changing international military alliances during the Cold War and detecting dynamic plotlines in the television series Game of Thrones. Also, in Section S.11 of the Supplementary Material, we show that the HDP-LPCM corroborates

many previous findings concerning Sampson's monastery network (Sampson, 1969), a standard pedagogical example in the network literature.

For each analysis, we estimated the HDP-LPCM with a truncation level $L = 25$ using 50,000 iterations for tuning, 50,000 iterations for burn-in, which left a remaining 400,000 iterations for inference. The trace plots, autocorrelation functions (ACFs), and marginal posterior densities are contained in Section S.12 of the Supplementary Material. Visual inspection of the trace plots indicates the algorithm converged on each dataset. Furthermore, the estimated model's in-sample AUC was 0.99 for each analysis, which indicates a good in-sample fit.

6.1 International Military Alliances

We begin by using the HDP-LPCM to examine changes in international military alliances during the first three decades of the Cold War (1950–1979). We use the Formal Alliances (v4.1) dataset curated as part of the Correlates of War Project (Gibler, 2009b). The raw dataset records all formal alliances – mutual defense pacts, non-aggression treaties, and ententes – among nations between 1816 and 2012. The goal of our analysis is to uncover the competing political blocs that defined the Cold War period in history and to determine any points in time where that alliance structure changed.

For this analysis, we use the yearly dyadic dataset, which records an undirected edge between two nations if there is a formal alliance between them during that year. To simplify the analysis, we discretized the dynamic networks into five year chunks from 1950–1979. We removed the nodes with a degree less than two from each network to focus on the larger political blocs found in the dataset. In addition, we required a nation to have at least one alliance during 1950–1979. We binarized the relations so that a connection between nations i and j at time t means they had at least one active alliance during those five years. This preprocessing resulted in $T = 6$ undirected binary networks that each contains $n = 107$ actors.

The alluvial diagram (Figure 9) of the partition selected with the procedure described in Section 4.2 shows that the HDP-LPCM estimates six communities overall with five communities active during all six time points. Note that the model uses group 2 to collect isolated nodes and small intermittent alliances. This interpretation is supported by the fact that $\hat{\sigma}_2 = 132$, which is 28 times larger than the second largest group shape $\hat{\sigma}_1 = 4.67$, and it encompasses most of the latent space. Thus we consider groups 1, 3, 4, 5, 6, and 7 as the major blocs inferred by the model. Note that these blocs strongly dictate the network's dynamics with an inferred blending coefficient $\lambda = 0.994$. Figure 10 depicts the estimated latent spaces for the years 1950–1954 and 1960–1964. We included the latent spaces for the remaining years in the Supplementary Material.

The static communities (groups 1, 3, 4, 5, and 6) coincide with long term regional alliances during the Cold War. Group 1 corresponds to the Western Bloc, consisting primarily of nations that are a part of the North Atlantic Treaty Organization (NATO) and the Western European Union (WEU). The competing group 5 represents the Eastern Bloc, which consisting of the Soviet Union and its satellite states such as East Germany, Czechoslovakia, and Poland. Group 3 consists of the Latin American coalition of the
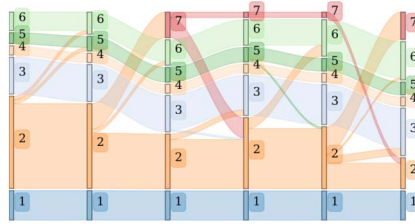
Figure 9: Alluvial diagram for the international military alliances network. Each line is the flux of nodes going from one group to the next at time t to $t+1$. A line's thickness is proportional to the number of nodes and the total height represents the 107 nodes.

Organization of American States (OAS), founded in 1948 to oppose socialism. Group 4 consists of member nations of the Southeast Asia Treaty Organization (SEATO), which is the Asian equivalent of NATO. Group 6 is the Arab League (at the time the League of Arab States) formed in 1945 to protect the interests of Arab countries. As depicted in the alluvial diagram, an interesting finding of the HDP-LPCM is that there is little exchange of nations between these groups over this time period. This finding bolsters the claim that "once the 'cold war' confrontation became apparent ... many nations cast their lot with either the American or Soviet bloc" (Small and Singer, 1969).

The evolving community structure (birth of group 7) is a result of the emergence of Africa as a world power. In particular, seventeen African nations gained their independence in 1960 alone. As a result of these newly independent nations, a large number of regional alliances formed in the early 1960s (Gibler, 2009a). This is reflected in the HDP-LPCM by the emergence of group 7 at $t = 3$, which encompasses the newly independent African nations. Fewer African alliances formed over the next decade. Our model reflects this fact by only including the Union of Central African States, which consists of former French colonies, in group 7 at $t = 4$ and 5. Finally, the Economic Community of West African States (ECOWAS) formed in 1975, which resulted in the re-introduction of many western African states into group 7 at $t = 6$. The dynamic nature of the HDP-LPCM is essential in revealing the emergence of Africa as a world power and demonstrates the importance of incorporating community evolution in latent space network modeling.

6.2 Character Interactions in Game of Thrones

In this section, we study the networks of character interactions in the television series Game of Thrones. The goal of our analysis is to use community detection to pinpoint coherent dynamic plotlines within the series. This is an interesting case study because the series' narrative is known for its many dynamic stories and characters who freely move between them. As such, we expect various groups of characters to form and die out across the series' lifetime and actors to freely move between these groups.

We utilize the networks compiled by Beveridge and Chemers (2018), who parsed fan-generated scripts found on the user curation site Genius. The original dataset consists
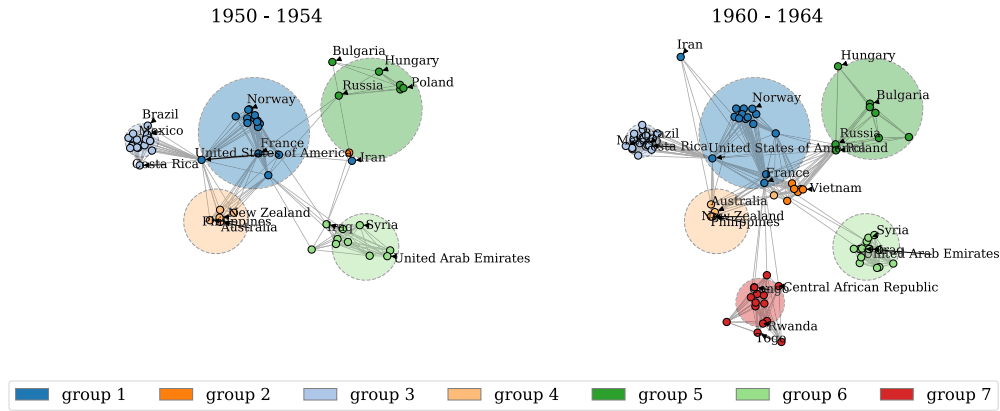
Figure 10: Latent space of international military alliances for the years 1950–1954 (left) and 1960–1964 (right). The group shapes σ_g are displayed as two-standard deviation ellipses centered on the group means $\boldsymbol{\mu}_g$. The names of select nations are annotated. The undirected edges are also displayed. For clarity, all unconnected nodes and group 2's two-standard deviation ellipse are removed.

of weighted character-character interactions split up between the eight seasons of the television series. The weight of an edge equals the number of interactions during a given season. Since this is primarily a pedagogical example, we restrict the dataset to the first four seasons of the show. To remove minor characters, we only keep interactions that occur greater than or equal to 10 times each season. The final result is $T = 4$ binary undirected networks with a total of $n = 165$ actors each.

The alluvial diagram (Figure 11) and associated latent space (Figure 12) reveal a dynamic group structure. The model infers six groups overall; however, only three groups are active during the first season of the show. During the second season, group 2 splits off into groups 3, 4, 6, and 7. In addition, group 5 is created out of characters from group 6. Note that the model uses group 6 to collect inactive characters, so we exclude it from further analysis. After the change in season two, the network's group structure remains constant for seasons three and four. The latent space of seasons three and four are included in the Supplementary Material. Furthermore, since the blending coefficient $\lambda = 0.974$, we conclude that the groups drive the evolution of the network. Overall, the ability of the HDP-LPCM to infer an evolving group structure is crucial for properly understanding the network's dynamics.

The inferred groups and their dynamics are consistent with the storylines in the Game of Thrones series. During season one, there are two active groups. Group 2 consists of all characters on Westeros, while group 1 centers around Deanery Targaryen's story on Essos. Starting at season two, group 2 contains characters revolving around Arya Stark's story arc, House Lannister, and other characters at King's Landing. Group 3 contains Bran Stark's group of companions (Hodor, Rickon, Osha, Meera, and Jojen) as they travel north of the wall. The plotlines related to the remainder of House Stark are
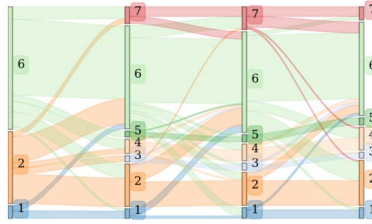
Figure 11: Alluvial diagram for the Game of Thrones character interaction networks.
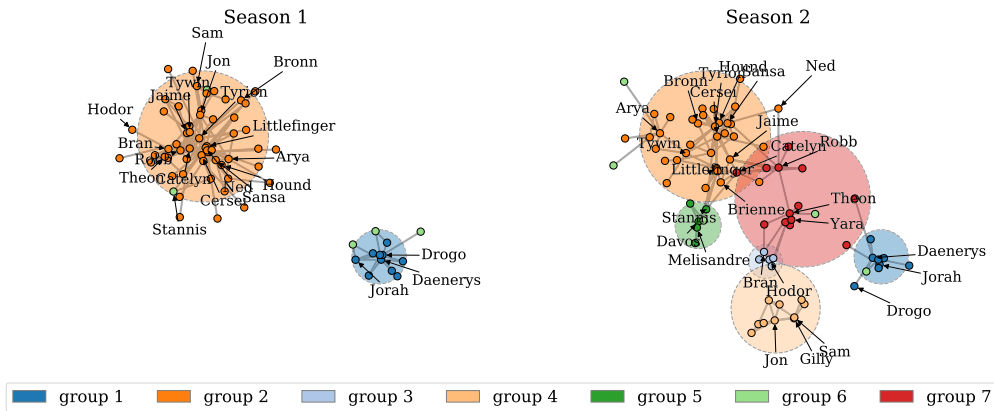

Figure 12: Latent space of character interactions in Game of Thrones for seasons 1 and 2. We removed all unconnected nodes and group 6's two-standard deviation ellipse.

contained in group 7. This group is composed of both Robb Stark's contingent as well as those related to Theon Greyjoy, who currently rules Winterfell the ancestral home of the Starks. Group 4 pertains to the story north of the wall. It most notably contains Jon Snow and Sam Tarly. Lastly, group 5 revolves around the story of Stannis Baratheon and his quest to regain the Iron throne, which was introduced in season two. In conclusion, our model recovers narratively coherent groups within the Game of Thrones character interaction networks and identifies their dynamics within the storyline.

7   Discussion

In this article, we proposed the hierarchical Dirichlet process latent position cluster model (HDP-LPCM) for dynamic networks. This is the first, to our knowledge, latent position model that can detect evolving community structures. In addition, we demonstrated that the HDP-LPCM still performs well when there is a static group structure. To accomplish our modeling goals, we used Bayesian nonparametric priors to provide simultaneous inference over the number of communities at each time point in the network

and the dynamics of each actor's latent position. Furthermore, our MCMC inference procedure has computational advantages over existing approaches. In particular, we avoided the BIC approximations of Handcock et al. (2007) and the computationally expensive process of estimating a large number of models.

In this work, we focused solely on community detection for binary undirected networks; however, our method can easily be extended to other network types. For example, both directed and weighted networks can be accommodated through the minor modifications of the likelihood presented in Sewell and Chen (2016). Due to the HDP-LPCM's conditional dependence structure, one only needs to modify the Metropolis-Hastings steps (i.e., steps 2 and 3) in Algorithm 1 to use the new network likelihood.

Common to most latent distance models, our proposed MCMC estimation method is time-intensive for large networks due to the quadratic scaling of the procedure. Specifically, the proposed sampler produced 3.3 and 4 samples per second on a 2015 MacBook Pro with a 2.2 GHz Intel Core i7 processor for the networks in Section 6.1 and Section 6.2, respectively. To alleviate this issue, one could use the case-control likelihood of Raftery et al. (2012). This method approximates the full likelihood by sub-sampling the unconnected edges, which allows for inference that scales linearly with the number of nodes. To further decrease the model's runtime, one could create a variational Bayes or stochastic variational Bayes (Hoffman et al., 2013) algorithm. However, the lack of conjugacy in the distance model means such an algorithm is non-trivial to implement. For this reason, we leave these extensions to future work.

Although our model can infer the number of communities, it does not provide inference for the latent space's dimension, p. In the numerical studies, we fixed $p = 2$, which allowed for visualization and was adequate for the networks considered in this work. Furthermore, we believe such visualizations allow for human-in-the-loop model scrutiny, which is a primary strength of the latent position model over other approaches such as latent feature models (Miller et al., 2009). Of course, selecting an appropriate dimension of the latent space is crucial if one is focused on tasks such as link prediction. A straightforward approach is to use a model selection criteria such as BIC or the Watanabe-Akaike information criterion (WAIC) (Watanabe, 2010). Another possibility is to put a prior over the dimensionality such as in Durante and Dunson (2014). Regardless, the selection of p remains an interesting and open question for future research.

Finally, while the HDP-LPCM's assumption of static group locations matches certain real-world networks, one can devise situations more naturally described by moving group centers. For example, increased polarization between two groups resulting from the groups moving further apart over time. We chose to keep the groups fixed for two reasons: the simplification still captures the merging, splitting, creation, and deletion phenomena of interest, and it removes the non-identifiability issue of distinguishing between the creation of a group and the movement of a group to a new location. Devising a model that efficiently infers dynamic group centers is an area of future research interest.

Despite these possible extensions, we believe the HDP-LPCM has a broad range of applicability. As demonstrated on the real-world networks in this paper, applications include the detection of changing alliances structures as well as inference regarding

narrative plotlines. We believe our methodology can be built upon to provide new tools for understanding the sequential evolution of communities in dynamic networks.

Supplementary Material

Supplemental Material for "A Bayesian Nonparametric Latent Space Approach to Modeling Evolving Communities in Dynamic Networks" (DOI: 10.1214/21-BA1300SUPPA; .pdf).

The `dynetlsm` Python package (DOI: 10.1214/21-BA1300SUPPB; .zip). Contains code to sample from the HDP-LPCM. This package (Loyal and Chen, 2022b) also includes all datasets and scripts used to produce the results presented in this work . This package is also available at https://github.com/joshloyal/dynetlsm.

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). "Mixed Membership Stochastic Blockmodels." *Journal of Machine Learning Research*, 9: 1981–2014.   52

Beveridge, A. and Chemers, M. (2018). "The Game of 'The Game of Thrones': Networked Concordances and Fractal Dramaturgy." In Brembilla, P. and Pascalis, I. A. D. (eds.), *Reading Contemporary Serial Television Universes: A Narrative Ecosystem Framework*, chapter 12. Routledge.   51, 70

Binder, D. A. (1978). "Bayesian Cluster Analysis." *Biometrika*, 65(1): 31–38. MR0501592. doi: https://doi.org/10.1093/biomet/65.1.31.   60

Durante, D. and Dunson, D. B. (2014). "Nonparametric Bayes Dynamic Modelling of Relational Data." *Biometrika*, 101(4): 883–898. MR3286923. doi: https://doi.org/10.1093/biomet/asu040.   73

Fan, X., Cao, L., and Xu, R. Y. D. (2015). "Dynamic Infinite Mixed-Membership Stochastic Blockmodel." *IEEE Transactions on Neural Networks and Learning Systems*, 26(9): 2072–2085. MR3453141. doi: https://doi.org/10.1109/TNNLS.2014.2369374.   52

Fox, E., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011a). "Bayesian Nonparametric Inference of Switching Dynamic Linear Models." *IEEE Transactions on Signal Processing*, 59(4): 1569–1585.   58

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011b). "A Sticky HDP-HMM with Applications to Speaker Diarization." *Annals of Applied Statistics*, 5(2A): 1020–1056. MR2840185. doi: https://doi.org/10.1214/10-AOAS395.   57, 59

Fu, W., Song, L., and Xing, E. P. (2009). "Dynamic Mixed Membership Blockmodel for Evolving Networks." In *Proceedings of the 26th Annual International Conference of Machine Learning*, 329–336. ACM.   52

Gibler, D. M. (2009a). "The Changing Nature of Military Alliances." In Gibler, D. M. (ed.), *International Military Alliances, 1648–2008*, chapter 6, 553–556. Washington: CQ Press. 70

Gibler, D. M. (2009b). *International Military Alliances, 1648–2008*, volume 2. Washington, D.C.: CQ Press. 51, 69

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). "Model-Based Clustering of Social Networks." *Journal of the Royal Statistical Society – Series A*, 170(2): 301–354. MR2364300. doi: https://doi.org/10.1111/j.1467-985X.2007.00471.x. 50, 55, 58, 73

Ho, Q., Song, L., and Xing, E. P. (2011). "Evolving Cluster Mixed-Membership Blockmodel for Time-Evolving Networks." In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, 342–350. PMLR. 52

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association*, 97(460): 1090–1098. MR1951262. doi: https://doi.org/10.1198/016214502388618906. 50, 53, 62

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013). "Stochastic Variational Inference." *Journal of Machine Learning Research*, 14: 1303–1347. MR3081926. 73

Ishiguro, K., Iwata, T., Ueda, N., and Tenenbaum, J. B. (2010). "Dynamic Infinite Relational Model for Time-varying Relational Data Analysis." In Lafferty, J. D., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A. (eds.), *Advances in Neural Information Processing Systems 23*, 919–927. Curran Associates, Inc. 51

Ishwaran, H. and Zarepour, M. (2000). "Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models." *Biometrika*, 87(2): 371–390. MR1782485. doi: https://doi.org/10.1093/biomet/87.2.371. 59

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N. (2006). "Learning Systems of Concepts with an Infinite Relational Model." In *Proceedings of the 21st National Conference on Artificial Intelligence*, 381–388. 51

Kim, M. and Leskovec, J. (2013). "Nonparametric Multi-group Membership Model for Dynamic Networks." In *Advances in Neural Information Processing Systems 26*, 1385–1393. Curran Associates, Inc. 52

Krivitsky, P. N., Handcock, M. S., and Raftery, A. E. (2009). "Representing Degree Distributions, Clustering, and Homophily in Social Networks with Latent Cluster Random Effects Models." *Social Networks*, 31(3): 204–213. 54

Lau, J. W. and Green, P. J. (2012). "Bayesian Model-Based Clustering Procedures." *Journal of Computational and Graphical Statistics*, 16(3): 526–558. MR2351079. doi: https://doi.org/10.1198/106186007X238855. 60

Loyal, J. D. and Chen, Y. (2022a). "Supplementary Material for "A Bayesian Nonparametric Latent Space Approach to Modeling Evolving Communities in Dynamic

Networks".". *Bayesian Analysis*. doi: https://doi.org/10.1214/21-BA1300SUPPA. 56

Loyal, J. D. and Chen, Y. (2022b). "Supplementary Material for "A Bayesian Non-parametric Latent Space Approach to Modeling Evolving Communities in Dynamic Networks"." *Bayesian Analysis*. doi: https://doi.org/10.1214/21-BA1300SUPPB. 74

Matias, C. and Miele, V. (2017). "Statistical Clustering of Temporal Networks through a Dynamic Stochastic Block Model." *Journal of the Royal Statistical Society – Series B*, 79(4): 1119–1141. MR3689311. doi: https://doi.org/10.1111/rssb.12200. 63

Meilă, M. (2007). "Comparing Clusterings – an Information Based Distance." *Journal of Multivariate Analysis*, 98(5): 873–895. MR2325412. doi: https://doi.org/10.1016/j.jmva.2006.11.013. 60

Miller, K., Jordan, M. I., and Griffiths, T. L. (2009). "Nonparametric Latent Feature Models for Link Prediction." In *Advances in Neural Information Processing Systems*, 1276–1284. MR3022515. 73

Paez, M. S., Amini, A. A., and Lin, L. (2019). "Hierarchical Stochastic Block Model for Community Detection in Multiplex Networks." *arXiv preprint arXiv:1904.05330*. 52

Papastamoulis, P. and Iliopoulos, G. (2010). "An Artificial Allocations Based Solution to the Label Switching Problem in Bayesian Analysis of Mixtures of Distributions." *Journal of Computational and Graphical Statistics*, 19(2): 313–331. MR2758306. doi: https://doi.org/10.1198/jcgs.2010.09008. 63

Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012). "Fast Inference for the Latent Space Network Model Using a Case-Control Approximate Likelihood." *Journal of Computational and Graphical Statistics*, 21(4): 901–919. MR3005803. doi: https://doi.org/10.1080/10618600.2012.679240. 73

Robert, C. P., Ryden, T., and Titterington, D. M. (2000). "Bayesian Inference in Hidden Markov Models through the Reversible Jump Markov Chain Monte Carlo Method." *Journal of the Royal Statistical Society – Series B*, 62(1): 57–75. MR1747395. doi: https://doi.org/10.1111/1467-9868.00219. 57

Sampson, S. (1969). "Crisis in a Cloister." Ph.D. thesis, Cornell University. 69

Sarkar, P. and Moore, A. W. (2005). "Dynamic Social Network Analysis using Latent Space Models." *SIGKDD Explorations*, 7(2): 31–40. 50, 54

Sethuraman, J. (1994). "A Constructive Definition of Dirichlet Priors." *Statistica Sinica*, 4(2): 639–650. MR1309433. 56

Sewell, D. K. and Chen, Y. (2015). "Latent Space Models for Dynamic Networks." *Journal of the American Statistical Association*, 110(512): 1646–1657. MR3449061. doi: https://doi.org/10.1080/01621459.2014.988214. 50, 54, 64

Sewell, D. K. and Chen, Y. (2016). "Latent Space Models for Dynamic Networks with

Weighted Edges." *Social Networks*, 44: 105–116. MR3449061. doi: https://doi.org/10.1080/01621459.2014.988214.   73

Sewell, D. K. and Chen, Y. (2017). "Latent Space Approaches to Community Detection in Dynamic Networks." *Bayesian Analysis*, 12(2): 351–377. MR3620737. doi: https://doi.org/10.1214/16-BA1000.   50, 51, 54, 58, 64

Sewell, D. K., Chen, Y., Bernhard, W. T., and Sulkin, T. E. (2016). "Model-Based Longitudinal Clustering with Varying Cluster Assignments." *Statistica Sinica*, 26(1): 205–233. MR3467114.   55

Small, M. and Singer, D. J. (1969). "Formal Alliances, 1816–1966: An Extension of the Basic Data." *Journal of Peace Research*, 6(3): 257–282.   70

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). "Hierarchical Dirichlet Processes." *Journal of the American Statistical Association*, 101(476): 1566–1581. MR2279480. doi: https://doi.org/10.1198/016214506000000302.   57

Wade, S. and Ghahramani, Z. (2018). "Bayesian Cluster Analysis: Point Estimation and Credible Balls (with Discussion)." *Bayesian Analysis*, 13(2): 559–626. MR3807860. doi: https://doi.org/10.1214/17-BA1073.   60, 62

Watanabe, S. (2010). "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory." *Journal of Machine Learning Research*, 11: 867–897. MR3049492.   73

Acknowledgments