Dependent Relational Gamma Process Models for Longitudinal Networks

Sikun Yang [1] **Heinz Koeppl** [1]

Abstract

A probabilistic framework based on the covariate-dependent relational gamma process is developed to analyze relational data arising from longitudinal networks. The proposed framework characterizes networked nodes by *nonnegative* node-group memberships, which allow each node to belong to multiple latent groups simultaneously, and encodes edge probabilities between each pair of nodes using a Bernoulli Poisson link to the embedded latent space. Within the latent space, our framework models the *birth* and *death* dynamics of individual groups via a thinning function. Our framework also captures the evolution of individual node-group memberships over time using gamma Markov processes. Exploiting the recent advances in data augmentation and marginalization techniques, a simple and efficient Gibbs sampler is proposed for posterior computation. Experimental results on a simulation study and three real-world temporal network data sets demonstrate the model's capability, competitive performance and scalability compared to state-of-the-art methods.

1. Introduction

The study of relational data arising from various networks including social, biological and physical networks is becoming increasingly important due to the emergence of massive relational data collected from these domains. Many efforts have been dedicated to develop statistical models in terms of community detection and missing link prediction for analyzing relational data arising from static networks, where either a single snapshot of the network of interest or an aggregated network over time is presented; see (Goldenberg et al., 2010) for a review of the literature. However, network data, such as friendships or interactions in a social network, is often dynamic since the relations among the entities within the network may appear or disappear over time (Mucha et al., 2010). Accordingly, the latent groups composed of those temporally connected entities also *form* and *decay* over time. Hence, appropriate models are needed to enable a better understanding of the formation and evolution of dynamic networks (Phan & Airoldi, 2015).

A probabilistic framework is proposed to model such dynamic networks by assuming the network of interest is composed of a set of latent groups. Each node of the observed network is hence associated with a time-dependent memberships vector that governs its involvement in multiple groups and interactions with other nodes. The node-group memberships are assumed to be gamma distributed, thus, naturally *nonnegative* real-valued. Moreover, to capture time-evolving interactions between groups of nodes, we model the *birth* and *death* dynamics of individual groups explicitly via a dependent relational gamma process (dRGaP). The ideal number of latent groups can be adaptively learned from data via the shrinkage mechanism of the dRGaP.

Explicitly modelling group birth/death dynamics can be useful in many applications. For instance, latent groups in a network of military disputes between countries could mean alliances such as NATO coordinating collective defence to attacks by external forces (Schein et al., 2016a). These groups can be born and die afterwards. For example, the Warsaw Pact was established during the Cold War and dissolved in later years. We demonstrate that our model can discover interpretable latent structure on a real network of military interstate disputes (Ghosn et al., 2004) that agrees with our knowledge of international relations (Section 5). Furthermore, it is reasonable to model the time-evolving memberships of each individual node to interpret its joining and withdrawing behavior to these groups. Hence, we capture the dynamics of individual node-group memberships evolving over time via gamma Markov processes.

In contrast to dynamic network modelling using logistic or probit mapping functions (Foulds et al., 2011; Heaukulani et al., 2013; Durante et al., 2014a), we leverage the Bernoulli-Poisson link (BPL) function (Dunson & Herring, 2005; Zhou, 2015) to generate edges from the latent space representation, which makes the computational cost of our

[1]Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt, Germany. Correspondence to: Sikun Yang <sikun.yang@bcs.tu-darmstadt.de>, Heinz Koeppl <heinz.koeppl@bcs.tu-darmstadt.de>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

model to scale linearly with the number of edges, rather than quadratically with the number of nodes. In addition, the Bernoulli-Poisson link is also a more appropriate model for *imbalanced* binary data (Zhou, 2017; Hu et al., 2015), which makes the proposed model appealing for analyzing real-world relational data that are usually extremely sparse. To perform inference, we present an efficient Gibbs sampling algorithm exploiting the Pólya-gamma data augmentation technique (Polson et al., 2013) and the data augmentation and marginalization technique for discrete data (Zhou et al., 2015).

The paper is organized as follows. In Section 2, we shortly review the gamma process and the thinned completely random measure framework. In Section 3, we present our generative model. In Section 4, we discuss some related work. Experimental results are provided in Section 5. The complete Gibbs sampling algorithm and additional experimental results are presented in the supplementary material.

2. Covariate-Dependent Random Measures

Our dynamic network model is based on the thinned completely random measures (tCRMs) framework, originally proposed in (Foti et al., 2013) for the construction of covariate-dependent topic models and latent feature models. We generalize this construction for longitudinal network modelling. More specifically, a set of latent groups that constitute the underlying structure of the observed dynamic network is generated. Via the tCRMs framework, the generated groups are allowed to *form* and *decay* over time. To facilitate understanding, we shortly review the gamma process and the thinned CRMs.

2.1. Gamma Process

The gamma process (GaP) is a *completely random measure* (CRM) (Kingman, 1967) defined on the product space $\Theta \times \mathbb{R}_{>0}$ as $G \sim \mathrm{GaP}(G_0, c)$, where c is a *scale* parameter, and G_0 is a finite and continuous *base measure* over a complete separable metric space Θ, such that $G(S_k) \sim \mathrm{Gamma}(G_0(S_k), c)$ are independent gamma random variables for disjoint subsets $\{S_k\}_{k=1}^{\infty}$ of Θ. The positive Lévy measure of the gamma process can be expressed as $\nu(\mathrm{d}r) = cr^{-1}e^{-cr}\mathrm{d}r$. As a completely random measure, the gamma process can be regarded as a Poisson process on $\Theta \times \mathbb{R}_{>0}$ with mean measure $\nu(\mathrm{d}\theta, \mathrm{d}r)$. A sample from this Poisson process consists of countably infinite atoms because $\int \int_{\Theta \times \mathbb{R}_{>0}} \nu(\mathrm{d}\theta, \mathrm{d}r) = \infty$. Thus, a sample from the gamma process can be expressed as $G = \sum_{k=1}^{\infty} r_k \delta_{\theta_k} \sim \mathrm{GaP}(G_0, c)$. More detailed information about the gamma process can be found in (Wolpert et al., 1998; 2011).

2.2. The Thinned CRMs Framework

Let $\Pi = \{(x_k, \theta_k, r_k)\}_{k=1}^{\infty}$ be generated by a Poisson process on the augmented product space $\mathcal{X} \times \Theta \times \mathbb{R}_{>0}$ with mean measure $\nu(\mathrm{d}x, \mathrm{d}\theta, \mathrm{d}r)$. Let $G = \sum_{k=1}^{\infty} r_k \delta_{(x_k, \theta_k)}$ be a CRM on $\mathcal{X} \times \Theta \times \mathbb{R}_{>0}$, and let $\mathcal{T}$ denote the time set as the coveriate. Our goal is to construct a family of random measures $\{G^{(t)}\}_{t \in \mathcal{T}}$ dependent on covariate values $t \in \mathcal{T}$. To achieve this, we generate a set of binary random variables $b_k^{(t)}$ for each point $(x_k, r_k, \theta_k) \in \Pi$ such that $p(b_k^{(t)} = 1) = P_{x_k}(t)$, where $P_x : \mathcal{T} \to [0, 1]$ denotes the thinning function which determines the probability that atom k in the global measure G appears in the local measure $G^{(t)}$ at covariate value t. Then, the set of covariate-dependent CRMs $\{G^{(t)}\}_{t \in \mathcal{T}}$ can be specified as

$$G^{(t)} = \sum_{k=1}^{\infty} b_k^{(t)} r_k \delta_{\theta_k}, \qquad t \in \mathcal{T}.$$

The new CRMs are well-defined by the mapping theorem for the Poisson processes (Kingman, 1993), that is proved in (Foti et al., 2013). As a concrete example, we exploit a thinned gamma process (tGaP) to model the global atoms and their *activity/inactivity* at multiple time points originally developed for dynamic topic models. Let $\nu(\mathrm{d}x, \mathrm{d}\theta, \mathrm{d}r) = H(\mathrm{d}x)G_0(\mathrm{d}\theta)\nu_0(\mathrm{d}r)$, where $\nu_0(\mathrm{d}r) = cr^{-1}e^{-cr}\mathrm{d}r$ is the Lévy measure of the gamma process. We transform a Gaussian basis kernel pointwise using a logistic function as the thinning function:

$$P_{x_k}(t) = \sigma\left\{\omega_{0k} + \sum_{l=1}^{T} \omega_{lk} \exp[-\phi_k (t-l)^2]\right\},$$

where $\sigma(x) = 1/(1 + \exp(-x))$ denotes the logistic function. We fix the centres of these kernels to the T discrete time points in covariate space $\mathcal{T}$. We characterize each location $x_k \in \mathcal{X}$ by a set of $T + 1$ kernel weights $\omega_{lk} \in \mathbb{R}$, and a (shared) kernel width ϕ_k uniformly drawn from a fixed dictionary $\{\phi_1^*, \ldots, \phi_D^*\}$ of size D. To encourage sparsity of the kernel weights, we place a normal-inverse gamma prior over ω_{lk}, i.e., $\omega_{lk} \sim \mathcal{NIG}(\omega_{lk}; 0, 1, 1)$. Hence, the base measure $H(dx)$ can be expressed as $H(dx) = \mathcal{NIG}(\omega_{lk}; 0, 1, 1)\mathrm{Cat}(\phi_k; \{\phi_1^*, \ldots, \phi_D^*\})$. The generative procedure can be expressed as

$$G = \sum_{k=1}^{\infty} r_k \delta_{(x_k, \theta_k)} \sim \mathrm{CRM}(\nu(\mathrm{d}x, \mathrm{d}\theta, \mathrm{d}r)), \qquad (1)$$

$$\omega_{lk} \sim \mathcal{NIG}(0, 1, 1), \qquad \phi_k \sim \mathrm{Cat}(\phi_1^*, \ldots, \phi_D^*),$$

$$P_{x_k}(t) = \sigma\left\{\omega_{0k} + \sum_{l=1}^{T} \omega_{lk} \exp[-\phi_k (t-l)^2]\right\},$$

$$b_k^{(t)} \sim \mathrm{Bernoulli}\Big[P_{x_k}(t)\Big], \qquad G^{(t)} = \sum_{k=1}^{\infty} b_k^{(t)} r_k \delta_{\theta_k}.$$

3. Model Formulation

We represent a dynamic network by a sequence of adjacency matrices $A^{(t)}$ for each time $t = 1, 2, \ldots, T$. For the sake of clarity, we limit our focus to unweighted (binary) undirected (symmetric) dynamic networks without self-links although the proposed model can straightforwardly be generalized to nonnegative real-weighted networks via the Poisson randomized gamma distribution (Zhou et al., 2016). We denote each snapshot of a dynamic network by $A^{(t)} \in \{0,1\}^{N\times N}$ with N being the number of nodes. The binary symmetric matrix $A^{(t)}$ has entries $A^{(t)}_{ij} = 1$ if an edge between nodes i and j is present at time t, and $A^{(t)}_{ij} = 0$ otherwise. Let $\Pi^{(t)}_{ij}$ be the link probability between nodes i and j. Our model specifies

$$A^{(t)}_{ij} \mid \Pi^{(t)}_{ij} \sim \text{Bernoulli}(\Pi^{(t)}_{ij}), \qquad t \in \mathcal{T}$$

independently for each $i = 2, \ldots, N$ and $j = 1, \ldots, i-1$, with

$$\mathbf{E}\left[A^{(t)}_{ij} \mid \Pi^{(t)}_{ij}\right] = 1 - \exp\left(-\sum_{k=1}^{K}\sum_{k'=1}^{K} z^{(t)}_{ik}\lambda^{(t)}_{kk'}z^{(t)}_{jk'}\right),$$

where $z^{(t)}_{ik} \in \mathbb{R}_{>0}$ characterizes the membership of node i to group k at time $t \in \mathcal{T}$. In contrast to the latent feature relational models (Foulds et al., 2011; Heaukulani et al., 2013; Kim et al., 2013) that assume *binary* node-group relationships, our nonnegative memberships capture how strongly each node associates with multiple groups. The group interaction weight $\lambda^{(t)}_{kk'}$ modulates the probability that there exists a link between a node affiliated to group k and a second node affiliated to group k' at time t. Our framework exploits the dynamics of the underlying relations between nodes on two levels: (1) The latent groups can be *active* and *inactive*; and (2) the memberships of each node to groups evolve over time. We now proceed to describe each component of our framework in the following sections.

3.1. Model of Active Groups

Many previous works (Kim et al., 2013; Xu, 2015) have shown that explicitly modelling the dynamics of latent groups using a distance-dependent Indian buffet process (dd-IBP) (Gershman et al., 2015) or a linear dynamical system discovers interpretable latent structures and achieves good predictive performance. Here, we build the group interaction weight $\lambda_{kk'}$ on the relational gamma process construction (Zhou, 2015). For implementation convenience, we use a truncated version of the infinite capacity model by fixing the maximum number of groups to K. That is, we first generate a group weight r_k independently for each group k as $r_k \sim \text{Gamma}(\gamma_0/K, c)$, where γ_0 denotes the concentration parameter and c denotes the rate parameters. Then, the inter-group interaction weight $\lambda_{kk'}$ and intra-group interaction weight λ_{kk} can be generated as

$$\lambda_{kk'} \sim \begin{cases} \text{Gamma}(\xi r_k, \beta), & \text{if } k = k' \\ \text{Gamma}(r_k r_{k'}, \beta), & \text{otherwise} \end{cases} \tag{2}$$

where $\xi \sim \text{Gamma}(1,1)$ and $\beta \in \mathbb{R}_{>0}$.

For dynamic relational data, we exploit the thinned CRMs framework to capture the birth/death dynamics of latent groups assuming that the status of group k can be either *active* or *inactive* at time t. More specifically, we use a Bernoulli random variable $b^{(t)}_k = 1$ to indicate the presence of group k at time t, and $b^{(t)}_k = 0$ otherwise. Accordingly, the interaction weight between group k and k' is active at time t only if the two groups are both active at that time, i.e., $\lambda^{(t)}_{kk'} = \lambda_{kk'} b^{(t)}_k b^{(t)}_{k'}$.

Given the group interaction weight matrix Λ defined in Eq. (2), we generate the time-dependent group interaction weights $\lambda^{(t)}_{kk'}$ using the thinning function introduced in Section 2.2 with the prior specification:

$$\omega_{lk} \sim \mathcal{NIG}(0,1,1), \;\; \phi_k \sim \text{Cat}(\phi^*_1, \ldots, \phi^*_D),$$
$$P_{x_k}(t) = \sigma\left\{\omega_{0k} + \sum_{l=1}^{T} \omega_{lk} \exp[-\phi_k (t-l)^2]\right\},$$

where we fix the centres of the covariate-dependent kernel functions to the T discrete time points of the considered dynamic network. The probability of activity/inactivity of group k at time t can be determined by the thinning function. A smooth thinning function can encourage the snapshots of a dynamic network at nearby covariate values t to share a similar set of groups.

3.2. Dynamics of Node-Group Memberships

To capture the temporal dynamics of node-group memberships, we represent the individual memberships as independently evolving gamma Markov chains. More specifically, the individual membership $z^{(t)}_{ik}$ is assumed to be gamma distributed with shape parameter $z^{(t-1)}_{ik}$, where $z^{(t-1)}_{ik}$ corresponds to the membership of the same node at the previous time, and rate parameter τ as

$$z^{(t)}_{ik} \sim \text{Gamma}(z^{(t-1)}_{ik}, \tau), \qquad t \in \mathcal{T} \tag{3}$$
$$z^{(1)}_{ik} \sim \text{Gamma}(\theta_{ik}, \tau),$$

where we assume each membership starts off in a dummy state (θ_{ik} for $i = 1, \ldots, N$, and $k = 1, \ldots, K$). We place a gamma prior over the dummy node membership θ_{ik}, i.e. $\theta_{ik} \sim \text{Gamma}(1,1)$. Given the group interaction weights $\{\lambda^{(t)}_{kk'}\}$ and node-group memberships $\{z^{(t)}_{ik}\}$ generated by our model, we generate an edge $A^{(t)}_{ij}$ between two nodes i

and j at time t via the Bernoulli-Poisson link (BPL) function as

$$m_{ij}^{(t)} \sim \text{Poisson}\left(\sum_{k=1}^{K}\sum_{k'=1}^{K} z_{ik}^{(t)}\lambda_{kk'}^{(t)}z_{jk'}^{(t)}\right), \qquad (4)$$
$$A_{ij}^{(t)} = \mathbf{1}(m_{ij}^{(t)} \geq 1), \qquad t \in \mathcal{T}$$

where m_{ij} is a latent Poisson count that measures how often nodes i and j interact in the latent space representation. Marginalizing out $m_{ij}^{(t)}$ from Eq. (4) yields

$$A_{ij}^{(t)} \sim \text{Bernoulli}\left[1 - \exp\left(-\sum_{k=1}^{K}\sum_{k'=1}^{K} z_{ik}^{(t)}\lambda_{kk'}^{(t)}z_{jk'}^{(t)}\right)\right].$$

The conditional distribution of the latent count $m_{ij}^{(t)}$ can be expressed as

$$(m_{ij}^{(t)} \mid A_{ij}^{(t)}, -) \sim A_{ij}^{(t)}\text{Poisson}_{+}\left(\sum_{k=1}^{K}\sum_{k'=1}^{K} z_{ik}^{(t)}\lambda_{kk'}^{(t)}z_{jk'}^{(t)}\right),$$

where $x \sim \text{Poisson}_{+}(\sigma)$ is the zero-truncated Poisson distribution with support only on the positive integers, and "–" denotes all other variables. We note that if $A_{ij}^{(t)} = 0$, then $m_{ij}^{(t)} = 0$ almost surely (a.s.), and if $A_{ij}^{(t)} = 1$, then $m_{ij}^{(t)} \sim \text{Poisson}_{+}(\sum_{k=1}^{K}\sum_{k'=1}^{K} z_{ik}^{(t)}\lambda_{kk'}^{(t)}z_{jk'}^{(t)})$. Thus, the latent count $m_{ij}^{(t)}$ only needs to be sampled for $A_{ij}^{(t)} = 1$ with a rejection sampler (Zhou, 2015). This property makes the proposed model appealing for modelling large sparse dynamic networks because the computational cost of our model scales linearly with the number of edges.

3.3. Bayesian Nonparametric Interpretation

As $K \to \infty$, the group weights and their corresponding dummy node memberships vector constitute a draw from a gamma process as $G = \sum_{k=1}^{\infty} r_k\delta_{\boldsymbol{\theta}_k}$, where $\boldsymbol{\theta}_k = (\theta_{1k}, \ldots, \theta_{Nk}) \in \Theta$ is an atom sampled from a N-dimensional base measure $G_0(\mathrm{d}\boldsymbol{\theta}_k)/G_0(\Theta) = \prod_{i=1}^{N}\text{Gamma}(1,1)$. Accordingly, the intra- and inter-group interaction weights and their corresponding pair of node memberships vector constitute a draw $\Lambda \mid G = \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\lambda_{kk'}\delta_{(\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k'})}$ from a relational gamma process (Zhou, 2015). Via the thinned CRMs framework, $\Lambda^{(t)} \mid \Lambda = \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} b_k^{(t)}b_{k'}^{(t)}\lambda_{kk'}\delta_{(\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k'})}$ can be viewed as a draw from a covariate-dependent relational gamma process. For brevity, the definitions of these processes are presented in the supplementary material.

3.4. The Full Generative Model

The full generative model (truncated) for the observed dynamic network data $\{A^{(t)}\}_{t\in\mathcal{T}}$ along with the latent variables, parameters, and hyperparameters, is given by

$$\begin{aligned}
r_k &\sim \text{Gamma}(\gamma_0/K, c),\ \theta_{ik} \sim \text{Gamma}(1,1), \qquad (5)\\
\xi &\sim \text{Gamma}(1,1),\\
\lambda_{kk'} &\sim \begin{cases}\text{Gamma}(\xi r_k, \beta), & \text{if } k = k'\\ \text{Gamma}(r_k r_{k'}, \beta), & \text{otherwise}\end{cases}\\
\omega_{lk} &\sim \mathcal{NIG}(0,1,1), \qquad \phi_k \sim \text{Cat}(\phi_1^*, \ldots, \phi_D^*),\\
b_k^{(t)} &\sim \text{Bernoulli}\left(\sigma\Big\{\omega_{0k} + \sum_{l=1}^{T}\omega_{lk}\exp[-\phi_k(t-l)^2]\Big\}\right),\\
z_{ik}^{(t)} &\sim \text{Gamma}(z_{ik}^{(t-1)}, \tau), \qquad z_{ik}^{(1)} \sim \text{Gamma}(\theta_{ik}, \tau),\\
\lambda_{kk'}^{(t)} &= b_k^{(t)}\lambda_{kk'}b_{k'}^{(t)}, m_{ij}^{(t)} \sim \text{Poisson}\left(\sum_{k,k'=1}^{K} z_{ik}^{(t)}\lambda_{kk'}^{(t)}z_{jk'}^{(t)}\right),\\
A_{ij}^{(t)} &= \mathbf{1}(m_{ij}^{(t)} \geq 1).
\end{aligned}$$

3.5. Inference via Gibbs Sampling

Let $A^{(1:t)}$ denotes the sequence $A^{(1)}, \ldots, A^{(t)}$ and similarly for $Z^{(1:t)}$ and $\Lambda^{(1:t)}$. The model parameters that need to be sampled include: latent node-group memberships $\{z_{ik}^{(t)}\}$, $\{\theta_{ik}\}$, individual group weights $\{r_k\}$, scale parameter ξ, groups interaction weights $\{\lambda_{kk'}\}$, kernel weights $\{\omega_{lk}\}$, kernel widths $\{\phi_k\}$, thinning variables $\{b_k^{(t)}\}$, and latent counts $\{m_{ij}^{(t)}\}$. Exploiting the Pólya-gamma data augmentation technique (Polson et al., 2013) and the data augmentation and marginalization technique (Zhou et al., 2015), a simple and efficient Gibbs sampling algorithm is developed to perform the model inference. The details of our inference algorithm are presented in the supplementary material.

4. Related Work

Prior works on dynamic networks modelling include the exponential random graph model (ERGM) (Guo et al., 2007), matrix and tensor factorization based methods (Dunlavy et al., 2011) and statistical models (Sarkar et al., 2007; 2014; Ishiguro et al., 2010; Durante et al., 2014b; Schein et al., 2016a; Palla et al., 2016). Statistical dynamic network models received considerable attention because these models have favourable interpretability by providing uncertainty estimates for the uncovered latent representations (Hoff et al., 2001). Dynamic extensions of the mixed membership stochastic blockmodel (MMSB) (Airoldi et al., 2008) have been developed (Fu et al., 2009; Xing et al., 2010; Ho et al., 2011) using linear state space models to capture the evolution of *real-valued* node-group memberships. Recently, an extended Kalman filter (EKF) based algorithm (Xu et al., 2014) was proposed to infer dynamic stochastic blockmodels (SBM) with competitive performance. Dynamic extensions of the latent feature relational model (LFRM) (Miller et al., 2009) using an infinite factorial hidden Markov pro-

cess to capture the evolution of *binary* node-group memberships include the dynamic relational infinite feature model (DRIFT) (Foulds et al., 2011), the latent feature propagation model (LFP) (Heaukulani et al., 2013), and the dynamic multi-group membership graph model (DMMG) (Kim et al., 2013).

Our proposed model is a form of Poisson factorization model (Zhou et al., 2015; Gopalan et al., 2015), and can be considered as the dynamic extension of the hierarchical gamma process edge partition model (HGPEPM) (Zhou, 2015). The dependent CRM framework (Foti et al., 2013) has been exploited for dynamic topic models and dependent latent feature models previously. To the best of our knowledge, this is the first attempt to model activity and inactivity of latent groups using a thinned CRMs framework in longitudinal network modelling. Our Markov chain construction, used to capture the evolution of node-group memberships, is inspired by the data augmentation technique (Zhou et al., 2015) that has been exploited for dynamic matrix factorization (Acharya et al., 2015a; Schein et al., 2016b) and deep gamma belief networks (Zhou et al., 2016). We note that the dynamic gamma process Poisson factorization (D-GPPF) (Acharya et al., 2015b) has been proposed using gamma Markov chains to model the evolution of latent groups while the D-GPPF assumes node memberships are static over time. Yang & Koeppl (2018) directly generalized the relational gamma process model to dynamic networks using gamma Markov chains to model node-group evolving behavior. For longitudinal networks, it is more reasonable to explicitly model the birth and death dynamics of latent groups by switching off redundant groups to avoid overfitting the data and to strengthen the interpretability of the latent network structure.

5. Experiments

We demonstrate that the proposed **D**ependent **R**elational **G**amma **P**rocess **M**odel (DRGPM) infers more interpretable latent structure compared with the related methods using a synthetic example. Quantitive evaluations of our model, compared with state-of-the-art methods as discussed in Section 4, are performed in terms of missing link prediction and future network forecasting on three real-world data sets. The first baseline is the dynamic relational infinite feature model (DRIFT) for which we used the code provided by Foulds et al. (2011).[1] Additionally, we implemented the D-GPPF, where we set the hyperparameters and initialized the model parameters with the values provided in (Acharya et al., 2015b). The third baseline is the dynamic stochastic blockmodel (DSBM) based on an extended Kalman filter (EKF) augmented with a local search, for which we use the released code[2] with the default settings. We also compare DRGPM with the hierarchical gamma process edge partition model (HGPEPM) (Zhou, 2015)[3] and the dynamic Poisson gamma model (DPGM) (Yang & Koeppl, 2018) that only models the evolving node memberships. Moreover, we demonstrate that our model can discover highly interpretable latent structure on a military interstate disputes dataset. In the experiments, we set the hyperparameters for our model as $\gamma_0 = 1, \beta = 1, c = 1, \tau = 1$. Unless otherwise stated, we use $K = N/2$ for initilization, where N is the number of nodes. A sensitivity analysis revealed that we obtain similar results when instead setting $\gamma_0 = 0.1$ or $\gamma_0 = 10$. All the experiments were run on a standard desktop with 2.7 GHz CPU and 24 GB RAM.

5.1. Simulation Study

Following the procedure suggested by Durante et al. (2016), we generate synthetic data to evaluate our proposed model in estimating the formation and evolution of the latent network sturcture. We consider a dynamic network with $N = 50$ nodes monitored for $T = 70$ equally spaced time snapshots. To generate a time-varying network, we first generate five regimes defining the true edge probabilities, as shown in Figure 2. We then simulate the network edges $A_{ij}^{(t)} \mid \Pi_{ij}^{(t)} \sim \text{Bernoulli}(\Pi_{ij}^{(t)})$ with each of the five regimes according to Figure 1. To demonstrate that DRGPM can infer interpretable latent structure while avoiding to overfit the data, we compare DRGPM with D-GPPF and DPGM. We initialize all methods setting $K = 30$.

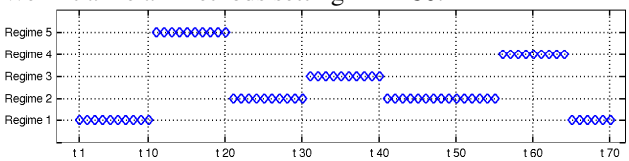

Figure 1. The graph showing which regime – i.e. true edge probabilities – for each snapshot is used to simulate the data.

In Figure 3, we depict the inferred link probabilities by DRGPM and D-GPPF in columns (b) and (d), respectively. We also depict the time-evolving node-group connections by computing the node-group association weights $z_{ik}^{(t)} \lambda_{kk}^{(t)}$ for DRGPM, and $r_k^{(t)} z_{ik}$ for D-GPPF in columns (c) and (e), respectively. We note that D-GPPF needs to generate many redundant groups to capture the time-evolving behavior of each node because of its unfavourable assumption that node memberships z_{ik} are static while group weights $r_k^{(t)}$ are time-dependent. Without this restriction, DRGPM characterizes the evolving node-group associations by explicitly modelling time-dependent node-group memberships, and hence generates an appropriate number of groups. In particular, using the thinned CRM framework, DRGPM can effectively activate newly-formed groups and switch off

[1] http://jfoulds.informationsystems.umbc.edu/code/DRIFT.tar.gz.

[2] https://tinyurl.com/ydf29he9.

[3] https://github.com/mingyuanzhou/EPM.

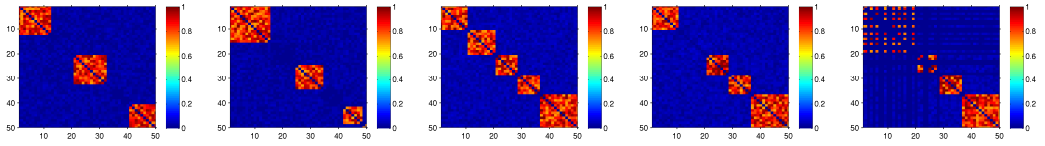

Figure 2. True edge probabilities for the five regimes in the simulation study.

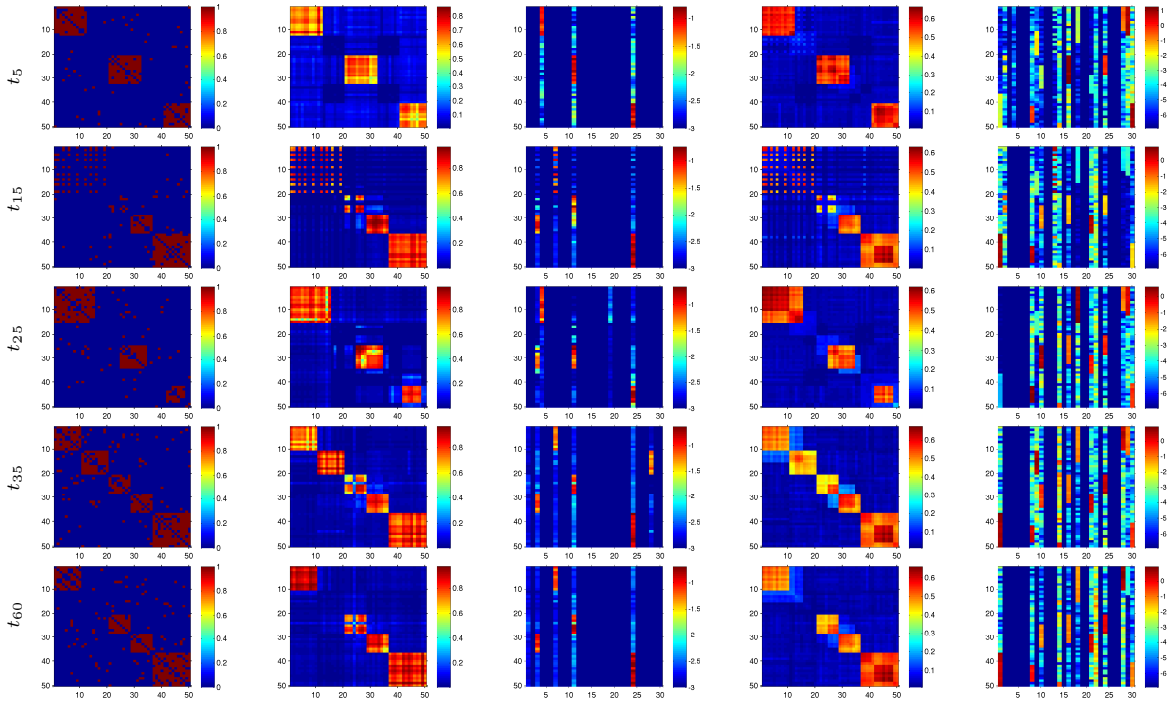

Figure 3. We select five snapshots of the simulated network as shown in column (a). The link probabilities inferred by DRGPM and D-GPPF are shown in columns (b) and (d), respectively. The association weights of each node (row variable) to the groups (column variable), as shown in columns (c) and (e), can be calculated as $z_{ik}^{(t)}\lambda_{kk}^{(t)}$ for DRGPM and $r_k^{(t)}z_{ik}$ for D-GPPF, respectively. The pixel values are displayed on $\log_{10}$ scale.

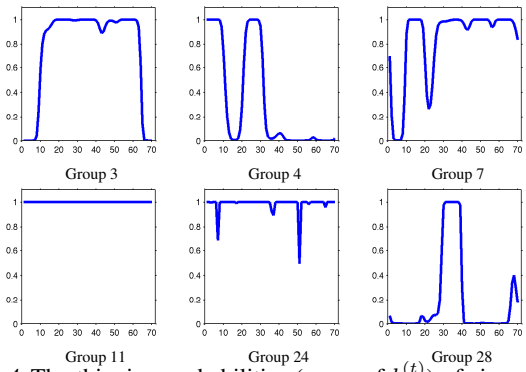

Figure 4. The thinning probabilities (mean of $b_k^{(t)}$) of six active groups inferred by DRGPM.

redundant groups over time, which strengthens the model interpretability for longitudinal network analysis. In Figure 4, we depict the thinning probabilities (mean of $b_k^{(t)}$) over time for six inferred active groups by DRGPM. We notice that DRGPM infers three groups $(4, 11, 24)$ at $t = 0$, turns off Group 4 and turns on Groups 3 and 7 at $t = 10$. Group 28 is only active from $t = 30$ to 40. The comparison of DRGPM to DPGM is presented in the supplementary material.

5.2. Quantitative Results

For the quantitative evaluation, we consider the following data sets: **(1) Face-to-face dynamic contacts network (FFDC):** This dataset (Mastrandrea et al., 2015) records timestamped face-to-face contacts among 180 students for 7 school days. We generated a dynamic network considering each school day as a snapshot, and created an edge between each pair of students at time t if they have at least one contact recorded at that given time. **(2) DBLP:** The DBLP co-authorship network data (Asur et al., 2009) contains the co-authorship information among 958 authors over ten years (1997-2006) in 28 conferences which spans three related research topics-database, data mining, and artificial

Table 1. Quantitive evalution. We highlight the performance of the best scoring model in bold.

Missing Link Prediction						
	FFDC		DBLP		Enron	
Model	AUROC	PR	AUROC	PR	AUROC	PR
HGPEPM	0.917 ± 0.006	0.354 ± 0.018	$\mathbf{0.979} \pm 0.004$	$\mathbf{0.791} \pm 0.014$	0.972 ± 0.001	0.443 ± 0.016
DSBM	0.878 ± 0.011	0.251 ± 0.017	0.913 ± 0.006	0.256 ± 0.009	0.916 ± 0.007	0.225 ± 0.023
D-GPPF	0.908 ± 0.005	0.313 ± 0.019	0.914 ± 0.005	0.308 ± 0.018	0.977 ± 0.002	0.499 ± 0.022
DRIFT	$\mathbf{0.933} \pm 0.006$	$\mathbf{0.416} \pm 0.020$	0.970 ± 0.019	0.491 ± 0.025	NA	NA
DPGM	0.921 ± 0.004	0.359 ± 0.020	0.960 ± 0.002	0.423 ± 0.032	0.979 ± 0.002	0.565 ± 0.014
DRGPM	0.924 ± 0.005	0.357 ± 0.018	0.963 ± 0.003	0.425 ± 0.023	$\mathbf{0.983} \pm 0.002$	$\mathbf{0.597} \pm 0.017$
Future Network Forecasting						
HGPEPM	0.733 ± 0.025	0.164 ± 0.022	0.714 ± 0.035	0.106 ± 0.027	0.828 ± 0.073	0.246 ± 0.140
DSBM	0.825 ± 0.085	0.181 ± 0.039	0.704 ± 0.030	0.091 ± 0.009	0.853 ± 0.059	0.325 ± 0.116
D-GPPF	0.842 ± 0.028	0.203 ± 0.046	0.734 ± 0.080	0.109 ± 0.046	0.878 ± 0.057	$\mathbf{0.360} \pm 0.121$
DRIFT	0.848 ± 0.056	$\mathbf{0.224} \pm 0.025$	0.745 ± 0.060	0.121 ± 0.054	NA	NA
DPGM	0.846 ± 0.017	0.221 ± 0.036	0.744 ± 0.053	$\mathbf{0.123} \pm 0.064$	$\mathbf{0.883} \pm 0.051$	$\mathbf{0.361} \pm 0.131$
DRGPM	$\mathbf{0.852} \pm 0.033$	$\mathbf{0.226} \pm 0.040$	$\mathbf{0.753} \pm 0.057$	$\mathbf{0.127} \pm 0.053$	$\mathbf{0.886} \pm 0.067$	$\mathbf{0.363} \pm 0.130$

intelligence. We focus on a subset of 324 most connected authors over all time period. **(3) Enron:** The Enron data[4] contains 517,431 emails among 151 users over 38 months (from May 1999 to June 2002). We generated a dynamic network aggregating the data into monthly snapshots, and created an edge between each pair of users at time t if they have at least one email recorded at that given time. The summary statistics are detailed in Table 2.

Table 2. Details of the data sets used in our experiments.

Dataset	FFDC	DBLP	Enron
# Nodes	180	324	151
# Slices	7	10	38
# Edges	8,332	11,154	11,414

Task 1: Predicting missing links First, we perform missing link prediction on the real-world data sets, and show the proposed model's predictive performance compared to the baseline models. We randomly hold out 20% of the network entries (either links or non-links) for each snapshot as test data, and use the remaining 80% to predict the held-out entries. DRIFT was infeasible to run on the Enron dataset in a reasonable amount of time given our computational resource. For DSBM, we either set K to the true number of classes provided by the data set or initialize it by examining the singular values of the first snapshot (Xu et al., 2014). We apply HGPEPM to each snapshot of dynamic networks independently. For all probabilistic methods, we use 2000 burn-in iterations, and collect 1000 samples from the model posterior distribution. We estimate the posterior mean of the edge probability for each held-out edge in the test data by averaging over the collected Gibbs samples. We then use these edge probabilities to evaluate the predictive performance of each model by calculating the area under the curve of the receiver operating characteristic (AUROC) and of the precision-recall (PR). In Table 1, we report the average evaluation metrics for each model over 10 runs. Overall, we found that DRIFT performs slightly better than DRGPM, although DRGPM has a significant advantage in terms of computational cost due to the Bernoulli-Poisson link (see Section 5.3). HGPEPM performs better than the dynamic models on the DBLP dataset because co-authorship links change dramatically from one year to the next one, and hence, the static model is better at fitting each snapshot independently. For the longitudinal Enron email network that is recorded monthly, DRGPM performs better than the baseline methods.

Task 2: Forecasting future networks Next, we consider the task of forecasting an unseen network snapshot $A^{(t)}$ given observed snapshots $A^{(1:t-1)}$. Following previous works (Foulds et al., 2011; Heaukulani et al., 2013; Kim et al., 2013), we train the models on the first $(t-1)$ snapshots of the considered network, and then estimate the predictive distribution of the unseen snapshot $A^{(t)}$ by running MCMC sampling one time step into the future. We apply HGPEPM to the most recent snapshot $A^{(t-1)}$, and then to perform prediction on the unseen snapshot $A^{(t)}$. For DRGPM, we set $\Lambda^{(t)} = \Lambda^{(t-1)}$, assuming the snapshots at nearby time points share a similar set of groups. We generated 10 samples of $Z^{(t)}$ for each of the 1000 samples collected for $Z^{(t-1)}$. For DSBM, we use the method detailed in (Xu et al., 2014) to perform future network forecasting. Table 1 shows the averaged performance for each model over different network snapshots from 3 to T. Overall, DRGPM shows competitive performance on all three datasets. This confirms that DRGPM can flexibly characterize temporally local links via time-evolving node memberships and switch off redundant groups to avoid overfitting the data.

5.3. Running Time

The probilistic models achieve higher accuracy although these methods require more computation time to collect MCMC samples. DSBM is much faster than the probabilistc models because its inference is performed using the

[4] https://www.cs.cmu.edu/~enron/.

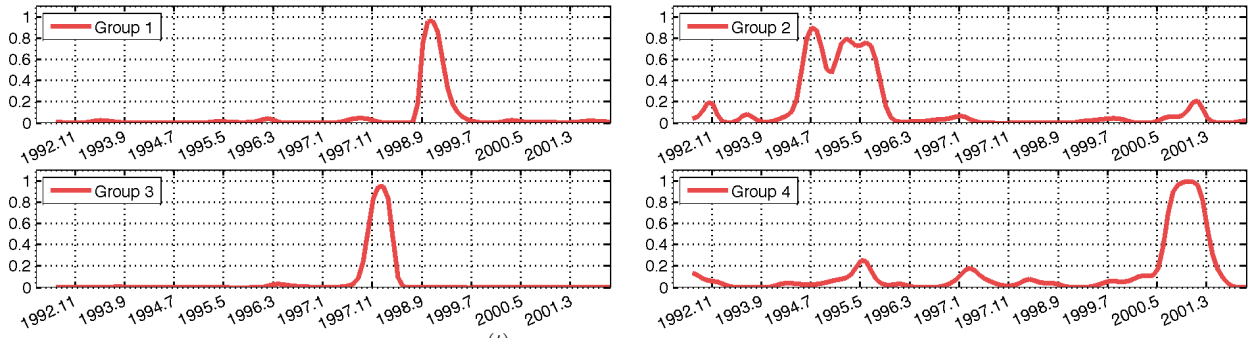

Figure 5. The activity (mean of $b_k^{(t)}$) of the four selected groups inferred from the MID network.

extended Kalman filter. Table 3 compares the per-iteration computation time of the sampling-based models (all models are implemented in Matlab). The computational cost of DRIFT scales in $\mathcal{O}(\bar{K}^2N^2T)$, where $\bar{K}$ is the expected number of groups. The Bernoulli-Poisson link based models (D-GPPF, DPGM, DRGPM) are much faster than the logistic link based method (DRIFT) because the former models scale linearly with the number of non-zero entries in network data. For DRGPM, sampling $\{m_{ij}^{(t)}\}_{i,j,t}$ and $\{m_{ikk'j}^{(t)}\}_{i,j,k,k',t}$ takes $\mathcal{O}(N^e\bar{K}^2)$ with N^e being the number of non-zero entries. Sampling $\{z_{ik}^{(t)}\}_{i,k,t}$ takes $\mathcal{O}(N\bar{K}T)$ and sampling $\{\lambda_{kk'}^{(t)}\}_{k,k',t}$ takes $\mathcal{O}(\bar{K}^2T)$. Overall, the computational complexity of DRGPM is $\mathcal{O}(N^e\bar{K}^2 + N\bar{K}T + \bar{K}^2T)$. The computational complexity of D-GPPF and DPGM is $\mathcal{O}(N^e\bar{K} + N\bar{K} + \bar{K}T)$ and $\mathcal{O}(N^e\bar{K}^2 + N\bar{K}T + \bar{K}^2)$, respectively. DRGPM is slightly faster than DPGM because DRGPM can effectively turn off redundant groups and hence achieves a lower computational cost.

Table 3. Comparison of computation time (seconds per iteration).

	FFDC	DBLP	Enron
DRIFT	164.342	382.119	-
D-GPPF	0.145	0.242	0.292
DPGM	0.748	1.676	1.705
DRGPM	0.623	1.217	1.234

5.4. Case Study: Military Interstate Disputes Dataset

We investigate the military interstate disputes (MID) dataset that contains disputes events between 138 countries from 1992 to 2001 (Ghosn et al., 2004) to explore the latent structure discovered by DRGPM. A dynamic network was generated by aggregating the data into monthly snapshots and a link was created between each pair of two countries if either country has disputes with the other one at that given time. We applied DRGPM to this dynamic network initializing $K = 30$ groups. Most of the identified groups correspond to some regional relations or conflicts. In Figure 5, we depict four interesting groups inferred by DRGPM and show the group activity by plotting the mean of the thinning function $b_k^{(t)}$. We normalized the node memberships to $[0, 1]$ by dividing them by the sum of memberships within the same group. In Table 4, we report the top 20 nodes associated to each of four groups with positive memberships. For instance, we found that Group 1 corresponds to the second Congo war (1998-2000). The first six nodes of the group are indeed the belligerents of this war. Group 2 corresponds to the Bosnian War (1992-1995), and its associated nodes are Yugoslavia and some NATO members that are indeed the belligerents of this war. Groups 3 and 4 are related to the regional disputes between some African countries. Additional graphs of the inferred groups, and tables showing the associated nodes are presented in the supplementary material.

Table 4. The top 20 nodes associated with each of the four selected groups as shown in Fig. 5 from the MID network. The highest node memberships to the corresponding selected groups throughout the whole period are reported for each node in parentheses.

Group	Country
1	Namibia (0.22), Chad (0.21), Zimbabwe (0.21), Angola (0.20), Dem. Rep. Congo (0.10), Sudan (0.05), Zambia (0.01)
2	Yugoslavia (0.60), Greece (0.13), Italy (0.04), UK (0.04), France (0.04), Belgium (0.03), Albania (0.03), Turkey (0.03), USA (0.02), Spain (0.02), Netherlands (0.01), Germany (0.01)
3	Nigeria (0.45), Ghana (0.31), Guinea (0.24)
4	Liberia (0.98), Sierra Leone (0.02)

6. Conclusion

We proposed a probabilistic framework for longitudinal network modelling based on the covariate-dependent relational gamma process. Our framework can characterize the group birth/death dynamics using the thinned CRM, which enables us to investigate the evolution of the inferred latent structure. The inferred latent dynamic structure can be useful for various qualitative analyses in practical applications. We experimentally demonstrated the competitive predictive performance and scalability of our framework on three real-world datasets. Our generative model can be easily extended in many interesting ways. For instance, it can be extended to *dynamic multilayer networks* (Durante et al., 2017) with relational data arising from multiple-related contexts via multi-level hierarchical gamma processes (Zhou et al., 2015). Additionally, network side information can be flexibly leveraged via a regression model (Rai et al., 2015).

Acknowledgements

The authors thank the anonymous reviewers and Adrian Šošić for their useful comments and suggestions. This research is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement 668858.

References

Acharya, Ayan et al. Nonparametric Bayesian factor analysis for dynamic count matrices. In *AISTATS*, pp. 1–9, 2015a.

Acharya, Ayan et al. Nonparametric dynamic network modeling. In *KDD Workshop on MLT*, pp. 104–113, 2015b.

Airoldi, Edoardo M. et al. Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.*, 9:1981–2014, June 2008.

Asur, Sitaram et al. An event-based framework for characterizing the evolutionary behavior of interaction graphs. *ACM Trans. Knowl. Discov. Data*, 3(4):16:1–16:36, 2009.

Dunlavy, Daniel M. et al. Temporal link prediction using matrix and tensor factorizations. *ACM Trans. Knowl. Discov. Data*, 5(2):10:1–10:27, 2011.

Dunson, David B. and Herring, Amy H. Bayesian latent variable models for mixed discrete outcomes. *Biostatistics*, 6 (1):11–25, 2005.

Durante, Daniele et al. Bayesian logistic Gaussian process models for dynamic networks. In *AISTATS*, pp. 194–201, 2014a.

Durante, Daniele et al. Nonparametric Bayes dynamic modelling of relational data. *Biometrika*, 101(4):883–898, 2014b.

Durante, Daniele et al. Locally adaptive dynamic networks. *Ann. Appl. Stat.*, 10(4):2203–2232, 2016.

Durante, Daniele et al. Bayesian learning of dynamic multilayer networks. *JMLR*, 18(43):1–29, 2017.

Ferguson, Thomas S. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1(2):209–230, 1973.

Foti, Nicholas et al. A unifying representation for a class of dependent random measures. In *AISTATS*, pp. 20–28, 2013.

Foulds, James R. et al. A dynamic relational infinite feature model for longitudinal social networks. In *AISTATS*, pp. 287–295, 2011.

Fu, Wenjie et al. Dynamic mixed membership blockmodel for evolving networks. In *ICML*, pp. 329–336, 2009.

Gershman, Samuel et al. Distance dependent infinite latent feature models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(2):334–345, 2015.

Ghosn, Faten et al. The mid3 data set, 1993-2001: Procedures, coding rules, and description. *Conflict Management and Peace Science*, 21(2):133–154, 2004.

Goldenberg, Anna et al. A survey of statistical network models. *Found. Trends Mach. Learn.*, 2(2):129–233, February 2010.

Gopalan, Prem et al. Scalable recommendation with hierarchical poisson factorization. In *UAI*, pp. 326–335, 2015.

Guo, Fan et al. Recovering temporally rewiring networks: a model-based approach. In *ICML*, pp. 321–328, 2007.

Heaukulani, Creighton et al. Dynamic probabilistic models for latent feature propagation in social networks. In *ICML*, pp. 275–283, 2013.

Ho, Qirong et al. Evolving cluster mixed-membership blockmodel for time-evolving networks. In *AISTATS*, pp. 342–350, 2011.

Hoff, Peter D. et al. Latent space approaches to social network analysis. *JASA*, 97:1090–1098, 2001.

Hu, Changwei et al. Zero-truncated Poisson tensor factorization for massive binary tensors. In *UAI*, pp. 375–384, 2015.

Ishiguro, Katsuhiko et al. Dynamic infinite relational model for time-varying relational data analysis. In *NIPS*, pp. 919–927, 2010.

Kim, Myunghwan et al. Nonparametric multi-group membership model for dynamic networks. In *NIPS*, pp. 1385–1393, 2013.

Kingman, J. F. C. Completely random measures. *Pacific J. Math.*, 21(1):59–78, 1967.

Kingman, J. F. C. *Poisson processes*. OUP, New York, 1993.

Mastrandrea, Rossana et al. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE*, 10(9):1–26, 2015.

Miller, Kurt et al. Nonparametric latent feature models for link prediction. In *NIPS*, pp. 1276–1284, 2009.

Mucha, Peter J et al. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

Palla, K. et al. Bayesian nonparametrics for sparse dynamic networks. *ArXiv e-prints*, 2016.

Phan, Tuan Q. and Airoldi, Edoardo M. A natural experiment of social network formation and dynamics. *PNAS*, 112(21):6595–6600, 2015.

Polson, Nicholas G. et al. Bayesian inference for logistic models using Pólya–Gamma latent variables. *JASA*, 108 (504):1339–1349, 2013.

Rai, Piyush et al. Large scale Bayesian multi-label learning via topic-based label embeddings. In *NIPS*, pp. 3222–3230, 2015.

Sarkar, Purnamrita et al. A latent space approach to dynamic embedding of co-occurrence data. In *AISTATS*, pp. 420–427, 2007.

Sarkar, Purnamrita et al. Nonparametric link prediction in large scale dynamic networks. *Electron. J. Statist.*, 8(2): 2022–2065, 2014.

Schein, Aaron et al. Bayesian Poisson Tucker decomposition for learning the structure of international relations. In *ICML*, pp. 2810–2819, 2016a.

Schein, Aaron et al. Poisson-gamma dynamical systems. In *NIPS*, pp. 5006–5014, 2016b.

Wolpert, Robert L. et al. Poisson/gamma random field models for spatial statistics. *Biometrika*, 85(2):251–267, 1998.

Wolpert, Robert L. et al. Stochastic expansions using continuous dictionaries: Lévy adaptive regression kernels. *Ann. Statist.*, 39(4):1916–1962, 08 2011.

Xing, Eric P. et al. A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.*, 4(2):535–566, 06 2010.

Xu, Kevin. Stochastic block transition models for dynamic networks. In *AISTATS*, pp. 1079–1087, 2015.

Xu, Kevin S. et al. Dynamic stochastic blockmodels for time-evolving social networks. *J. Sel. Topics Signal Processing*, 8(4):552–562, 2014.

Yang, Sikun and Koeppl, Heinz. A Poisson gamma probabilistic model for latent node-group memberships in dynamic networks. In *AAAI*, 2018.

Zhou, Mingyuan. Infinite edge partition models for overlapping community detection and link prediction. In *AISTATS*, pp. 1135–1143, 2015.

Zhou, Mingyuan. Discussion on "Sparse graphs using exchangeable random measures" by F. Caron and E. B. Fox. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1359–1360, 2017.

Zhou, Mingyuan et al. Negative binomial process count and mixture modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(2):307–320, 2015.

Zhou, Mingyuan et al. Augmentable gamma belief networks. *Journal of Machine Learning Research*, 17:1–44, 2016.