Negative-Binomial Randomized Gamma Dynamical Systems for Heterogeneous Overdispersed Count Time Sequences

Rui Huang\textsuperscript{1,2,3,4}, Sikun Yang\textsuperscript{1,2,3,4*}, Heinz Koeppl\textsuperscript{5}

\textsuperscript{1}School of Computing and Information Technology, Great Bay University, 523000 Dongguan, China
\textsuperscript{2}Great Bay Institute for Advanced Study, Great Bay University
\textsuperscript{3}Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems, Great Bay University
\textsuperscript{4}Dongguan Key Laboratory for Data Science and Intelligent Medicine, Great Bay University
\textsuperscript{5}Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt

\texttt{huangrui@pku.edu.cn, sikunyang@gbu.edu.cn, heinz.koeppl@bcs.tu-darmstadt.de}

Abstract

Modeling count-valued time sequences has been receiving growing interests because count time sequences naturally arise in physical and social domains. Poisson gamma dynamical systems (PGDSs) are newly-developed methods, which can well capture the expressive latent transition structure and bursty dynamics behind count sequences. In particular, PGDSs demonstrate superior performance in terms of data imputation and prediction, compared with canonical linear dynamical system (LDS) based methods. Despite these advantages, PGDSs cannot capture the heterogeneous overdispersed behaviours of the underlying dynamic processes. To mitigate this defect, we propose a negative-binomial-randomized gamma Markov process, which not only significantly improves the predictive performance of the proposed dynamical system, but also facilitates the fast convergence of the inference algorithm. Moreover, we develop methods to estimate both factor-structured and graph-structured transition dynamics, which enable us to infer more explainable latent structure, compared with PGDSs. Finally, we demonstrate the explainable latent structure learned by the proposed method, and show its superior performance in imputing missing data and forecasting future observations, compared with the related models.

1 Introduction

Count time sequences, naturally arise in many domains such as text mining [Blei and Lafferty, 2006; Wang et al., 2008; Rudolph and Blei, 2018; Acharya et al., 2018; Dieng et al., 2019], cell genomic analysis [Levitin et al., 2019; Tong et al., 2020; Jones et al., 2023], population movement forecasting [Sheldon et al., 2013; Stuart and Wolfram, 2020; Roy and Dunson, 2020], and etc. Modeling count sequences has been drawing increasing research attention because these real-world count data usually exhibit bursty and overdispersed behaviours, which cannot be well-captured by canonical linear dynamical systems (LDS) [Ghahramani and Roweis, 1998]. In addition, some previous works use extended rank likelihood functions [Han et al., 2014] which link the count observations to latent continuous dynamics to model count time sequences. Nonetheless, the extended rank likelihood functions cannot faithfully capture bursty dynamics underlying real-world count sequences. Meanwhile, the extended rank likelihood functions often require an approximate inference scheme, and thus scale poorly with high-dimensional count sequences, such as single-cell RNA sequencing data [Chandra et al., 2023]. Notably, some recent works [Acharya et al., 2015; Schein et al., 2016a; Schein et al., 2016b; Schein et al., 2019] model sequential count observations using gamma Poisson family distributions. More specifically, [Acharya et al., 2015] develops a gamma Markov process to capture continuous dynamics underlying count-valued sequences. In particular, the number of latent factors behind high-dimensional count data, can be appropriately determined by the gamma process prior, in a Bayesian non-parametric manner. Following the success of [Acharya et al., 2015], Schein et al. [2016a] study a Poisson gamma dynamical system, in which a transition kernel is designed to capture how the latent dimensions interact with each other to model complicated observed dynamics. Another appealing aspect of the Poisson gamma dynamic model is that the posterior simulation can be performed using a tractable-yet-efficient Gibbs sampling algorithm via Poisson-Logarithm data augmentation strategy [Zhou and Carin, 2012; Zhou and Carin, 2015]. Hence, the Poisson gamma dynamic models [Acharya et al., 2015; Schein et al., 2016a; Schein et al., 2019] are in particular well-fitted to impute missing entries, to predict future unseen observations and to estimate uncertainties.

Despite these advantages, these models still cannot well capture the heterogeneous overdispersion effects of the latent dynamic processes behind count observations. For instance, international event data, usually consists of multiple latent dynamic processes, which often change rapidly with the different magnitudes [King, 2001; Stewart, 2014]. To capture such heterogeneous overdispersed behaviours, we de-
velop a negative-binomial-randomized gamma Markov chain structure, which not only greatly enhances the model flexibility, but also facilitates the fast convergence of the derived Gibbs sampling algorithms. Moreover, the transition dynamics behind real-world high-dimensional count data, are often sparse, and exhibit a certain amount of graph structure. Hence, we propose to learn the graph-structured transition dynamics using relational gamma processes [Zhou, 2015]. To the best of our knowledge, this is the first attempt to learn the latent graph-structured transition dynamics under the Poisson gamma dynamical system.

The main contributions of the paper include: 1) A negative-binomial-randomized gamma Markov process (NBRGMP) is proposed to estimate the heterogeneous overdispersion effects of the latent dimensions underlying sequential count observations; 2) Relational gamma processes are thoroughly studied to learn both factor-structured and graph-structured transition dynamics, which renders the estimated latent structure more explainable, compared with transition structure inferred using non-informative priors; 3) Although the proposed NBRGMP and its factor-structured and graph-structured extensions are intractable, simple-yet-efficient Gibbs sampling algorithms are developed via Negative-binomial data augmentation strategies to perform inference; 4) Extensive experiments are conducted to illustrate the explainable transition structure learned by the proposed model. We demonstrate the superior performance of the proposed method in missing data imputation and future snapshot forecasting, with several related works.

2 Preliminary

Suppose we have a sequentially-observed count data over time interval $[0, T]$ specified by $N = (n_1, \ldots, n_{V})^T$ of $V$ dimensions, where $n_v = (n_v^{(0)}, \ldots, n_v^{(T)})^T$ with $n_v^{(t)}$ denoting the $v$-th observation at time $t$. The Poisson gamma dynamical system [Schein et al., 2016a] models the count $n_v^{(t)}$ as

$$n_v^{(t)} \sim \text{Pois} (\delta^{(t)} \sum_{k=1}^{K} \phi_{vk} \theta_k^{(t)}),\quad (1)$$

where $\delta^{(t)}$ captures the strength of latent component $k$ at time $t$, and $\phi_{vk}$ represents the involvement degree of dimension $v$ in latent component $k$. To model the underlying dynamics, the PGDS assumes that the latent components evolve over time according to a gamma Markov chain structure as

$$\theta_k^{(t)} \sim \text{Gam} (\tau_0 \sum_{k=2}^{K} \pi_{kk_2} \theta^{(t-1)}_{k_2}, \tau_0),\quad (2)$$

where the latent components $\theta^{(t-1)} = (\theta_1^{(t-1)}, \ldots, \theta_K^{(t-1)})^T$ evolve over time through the transition matrix $\Pi$. The $\theta_k^{(t-1)}$ captures how strongly the $k$-th latent component activates at time $t-1$, and $\pi_{kk_2}$ models how strongly the $k_2$-th component affects the $k$-th component at time $t$. Eq. 2 naturally defines a gamma Markov chain structure. The expectation and variance of the gamma Markov chain can be calculated respectively as $E(\theta^{(t)} | \theta^{(t-1)}, \Pi) = \Pi \theta^{(t-1)}$ and $\text{Var}(\theta^{(t)} | \theta^{(t-1)}, \Pi) = (\Pi \theta^{(t-1)}) \tau_0^{-1}$, where $\tau_0$ controls the variance of $\theta^{(t)}$.

Schein et al. [2019] further develop a Poisson-randomized gamma Markov chain (PRGMC) structure specified by

$$\theta_k^{(t)} \sim \text{Gam} (\epsilon_0^{(0)} + h_k^{(t)}, \tau), \quad h_k^{(t)} \sim \text{Pois} (\tau \sum_{k_2} \pi_{kk_2} \theta_k^{(t-1)}),$$

By marginalizing out the Poisson latent states $h_k^{(t)}$, we have a continuous-valued dynamical system given by

$$\theta_k^{(t)} \sim \text{RG1} (\epsilon_0^{(0)}, \tau \sum_{k_2} \pi_{kk_2} \theta_k^{(t-1)}),$$

where RG1 is the randomized gamma distribution of the first type. The marginal expectation and variance of the PRGMC is $E(\theta^{(t)} | \theta^{(t-1)}, \Pi) = \Pi \theta^{(t-1)} + \epsilon_0^{(0)} \tau^{-1}$ and $\text{Var}(\theta^{(t)} | \theta^{(t-1)}, \Pi) = 2 \Pi \theta^{(t-1)} \tau^{-1} + \epsilon_0^{(0)} \tau^{-2}$, respectively.

3 The Proposed Model

In this section we will introduce the novel negative-binomial-randomized gamma Markov chain structure to capture the heterogeneous overdispersion effects of the latent dimensions behind count data. Then we shall describe how to learn explainable latent transition structure with relational gamma processes. The proposed negative-binomial-randomized gamma dynamical system is defined by

$$n_v^{(t)} \sim \text{Pois} (\delta^{(t)} \sum_{k=1}^{K} \lambda_k \phi_{vk} \theta_k^{(t)}),\quad (3)$$

where $\delta^{(t)}$ is a nonnegative multiplicative term capturing time-dependent bursty dynamics. We place a gamma prior on $\delta^{(t)}$ as $\delta^{(t)} \sim \text{Gam} (\epsilon_0, \epsilon_0)$, and let $\delta^{(t)} = \delta$ if the generative process (Eq. 3) is stationary over time. Here $\phi_k = (\phi_{1k}, \phi_{2k}, \ldots, \phi_{Vk})^T$ denotes the loading coefficient of $k$-th latent component, and $\lambda_k$ denotes the weight of $k$-th latent component. To ensure model identifiability, we require $\sum_v \phi_{vk} = 1$ and thus have a Dirichlet prior over $\phi_k$ given by $\phi_k \sim \text{Dir} (\epsilon_0, \ldots, \epsilon_0)$. More specifically, we draw $\lambda_k$ from a hierarchical prior as $\lambda_k \sim \text{Gam} (\frac{\lambda}{K} + g_k, \beta)$, in which $g_k \sim \text{Pois} (\frac{\lambda}{K})$. We specify gamma priors over $\gamma$ and $\beta$ as $\gamma \sim \text{Gam} (\epsilon_0, \epsilon_0)$, $\beta \sim \text{Gam} (\epsilon_0, \epsilon_0)$. Note that as $K \rightarrow \infty$, the summation of the weight expectation remains finite and fixed, i.e., $\sum_k \mathbb{E} [\lambda_k] = \beta^{-1} (\epsilon_0 + \gamma)$. Hence, this hierarchical prior enables us to effectively estimate a finite number of latent factors that are representative to capture the temporal dynamics.

3.1 Negative-Binomial Randomized Gamma Markov Processes

To capture the heterogeneous overdispersed behaviors of the latent dimensions behind count sequences, we introduce a negative-binomial randomized gamma Markov process (NBRGMP) specified by

$$\theta_k^{(t)} \sim \text{Gam} (\epsilon_0^{(0)} + \psi_k^{(t)}, \tau), \quad h_k^{(t)} \sim \text{NB} (\tau \sum_{k_2} \pi_{kk_2} \theta_k^{(t-1)} - 1, \psi),\quad (4)$$

where we set $\theta_0^{(0)} = \lambda_k$, and $\psi_k^{(t)}$ is gamma distributed with shape parameter $\epsilon_0^{(0)} + h_k^{(t)}$, where $\epsilon_0^{(0)} \geq 0$, and the rate
because of its large expectation and variance. When the flexibility of a limited range of overdispersion effects. If 
while, the variance of the NBRGMP still allows us to capture to be smaller compared with the expectation of the PRGMC, 
the expectation of the NBRGMP will be three times of the variance of the PRGMC (as discussed in Sec. 2), while the variance of the 
the expectation and variance converge to zeros as \( \psi \) goes to inft-
ity. Fig. 1 plots the realizations of the NBRGMP by varying the parameter \( \psi \). Note that the negative-binomial distributed latent state \( h_k^{(t)} \) can be equivalently drawn from a gamma-Poisson mixture as

\[
h_k^{(t)} \sim \text{Pois}(\hat{h}_k^{(t)}), \quad \hat{h}_k^{(t)} \sim \text{Gam}(\tau \sum_{k=1}^{K} \pi_{kk} \theta_k^{(t-1)}, \psi). \]

Fig. 2 shows the graphical representation of the developed NBRGMP. When \( \tau \sum_{k=1}^{K} \pi_{kk} \theta_k^{(t-1)} \to \infty \), the \( h_k^{(t)} \) is approximately characterized by \( \text{Pois}(\frac{1}{2} \sum_{k=1}^{K} \pi_{kk} \theta_k^{(t-1)}) \). Hence, by marginalizing the Poisson distributed latent states \( h_k^{(t)} \) from Eq. 4, the negative-binomial randomized gamma dynamical system can be equivalently represented by randomized gamma distribution of the first type as

\[
\theta_k^{(t)} \sim \text{RG1}(\epsilon_0^{(\theta)}, \frac{\tau}{\psi} \sum_{k=1}^{K} \pi_{kk} \theta_k^{(t-1)}, \tau). \]

### 3.2 Factor-structured Transition Dynamics

We first propose to learn the latent factor structure behind transition dynamics. To that end, we specify a hierarchical Dirichlet prior over \( \pi_k \) as \( \pi_k \sim \text{Dir}(a_{1k}, \ldots, a_{Kk}) \), where \( a_k = (a_{1k}, \ldots, a_{Kk}) \) is the hyper-parameter. Our goal here is to capture the correlation structure between the latent dimensions of the transition kernel. Thus, we model the hyper-parameter \( A = [a_{k1} a_{k2}]^{K \times K} \) using a Poisson factor model as

\[
a_{k1} a_{k2} \sim \text{Pois}(\sum_{c=1}^{C} m_{k1c} r_c m_{k2c}), \quad m_{k1c} \sim \text{Gam}(a_{k1}, \hat{b}_k), \quad r_c \sim \text{Gam}(\frac{r_0}{C}, c_0),
\]

respectively. Here, \( C \) is the maximum number of latent factors. As \( C \to \infty \), the weights of the latent factors \( r_c \) tend
and the factor loading \( \{ \mathbf{m}_c \}^C \) can be considered as a draw
\[ G = \sum_{r=1}^{\infty} r \mathbf{C}_r \] from a gamma process \( \text{GaP}(G_0, c_0) \),
where \( G_0 \) denotes the base measure over the metric space \( \Omega \),
and \( c_0 \) the concentration parameter [Ferguson, 1973].

### 3.3 Graph-Structured Transition Dynamics

For high-dimensional count sequences, the underlying transition
dynamics are often sparse and exhibit a certain amount
of graph structure. Hence, we further study to learn the latent
transition kernel behind count time series, using relational gamma process prior. In particular, we sample
the transition parameter \( \pi_k \) from a hierarchical Dirichlet
prior, as \( \pi_k \sim \text{Dir}(a_{1:k}, \ldots, a_{K:k}) \). To introduce a sparse
graph-structured transition kernel, we model the matrix of the
hyper-parameter \( \mathbf{A} = [a_{k_1:k_2}]_{k_1,k_2} \) as \( \mathbf{A} = \mathbf{D} \odot \mathbf{Z} \),
where \( \mathbf{D} = [d_{k_1:k_2}]_{k_1,k_2} \) denotes the matrix of the nonnegative
hyper-parameters, and \( \mathbf{Z} = [z_{k_1:k_2}]_{k_1,k_2} \) is a binary mask.
More specifically, we consider the dimensions of the transition
kernel as vertices, and the non-zero transition behaviours
as graph edges. Naturally, we can capture the sparse structure
of the transition kernel \( \Pi \) using a graph. As shown in Fig. 3,
for each pair of two vertices \( i \) and \( j \), \( z_{ij} = 1 \) means that the
transition probability from \( i \)-th component to \( j \)-th component
is non-zero, and vice versa. In particular, we model the binary
mask \( \mathbf{Z} \) using relational gamma processes as
\[ z_{k_1:k_2} \sim \text{Ber}(1 - \exp(\sum_{c=1}^C m_{k_1:c} r_c m_{k_2:c})) \]
where \( r_c \) can be considered as the weight of latent community
\( c \), and \( m_{k:c} \) measures how strongly \( k \)-th vertex (the dimension
of the transition kernel) relate to \( c \)-th latent community,
as illustrated in Fig. 3. Note that the binary mask \( \mathbf{Z} \) can be
equivalently drawn via the Bernoulli-Poisson link function as
\[ z_{k_1:k_2} \sim \delta(w_{k_1:k_2} \geq 1), \quad w_{k_1:k_2} \sim \text{Pois}(\sum_{c=1}^C m_{k_1:c} r_c m_{k_2:c}) \]

To ensure the model explainability, we restrict \( r_c \) and \( m_{k:c} \)
to be nonnegative, and thus place gamma priors over these two
parameters as \( m_{k:c} \sim \text{Gam}(\delta_{k_1}, \delta_{k_2}) \), \( r_c \sim \text{Gam}(\hat{r}_c, c_0) \),
respectively. As we discussed in Sec. 3.2, this hierarchical
gamma prior can be considered as a draw \( G = \sum_{r=1}^{\infty} r \mathbf{C}_r \)
from a gamma process \( \text{GaP}(G_0, c_0) \). In particular, we call this
Bayesian non-parametric prior the relational gamma Markov
process, as a graph-structured can be naturally induced. The
non-negative hyper-parameters \( \mathbf{D} = [d_{k_1:k_2}]_{k_1,k_2} \) are drawn from
a gamma distribution as \( d_{k_1:k_2} \sim \text{Gam}(\epsilon_0, c_0) \).

The proposed gamma dynamical systems are not fully
conjugate. Nonetheless, tractable-yet-efficient Gibbs sampling
algorithms are developed to perform posterior simulation
via negative-binomial data augmentation strategies [Zhou,
2016a]. The full derivation of the inference procedure is pre-
sented in the supplementary material.

### 4 Related Work

Modeling sequentially observed count sequences has been
receiving growing interests in recent years. Here we dis-
cuss several types of methods closely related to our stud-
es. [Acharya et al., 2015] first studies the gamma Markov
process on sequentially count sequences, in which the latent
states evolve independently over time. [Schein et al., 2016a]
tries to capture the excitations among the latent gamma
Markov processes using a transition structure. [Schein
et al., 2019] investigates a Poisson-randomized gamma Markov
process which can capture a certain amount of bursty dynam-
ic, and thus demonstrates advantages over gamma Markov
processes. [Virtanen and Girolami, 2020] studies a second
type of gamma Markov chain structure via the scale pa-
rameter of the latent gamma states, which demonstrate bet-
ter stationary property over the gamma Markov chain
proposed by [Acharya et al., 2015]. Filstroff et al., 2021
recently provides a thorough survey on the studies of the devel-
oped gamma Markov processes, and evaluates these models
through standard tasks including data smoothing and fore-
casting. [Han et al., 2014] first tries to capture sequential
count observations using linear dynamical systems, via the
extend rank likelihood function. [Linderman et al., 2017]
proposes to learn switching behaviors of sequential data
using recurrent linear dynamical systems (rLDS). [Nassar et al.,
2018] further develops a tree-structured extension of rLDS,
with multi scale resolution. [Chen et al., 2020] extends the
Poisson gamma dynamical systems to learn non-stationary
transition dynamics behind count time series. Some efforts
are also dedicated to developing Bayesian deep models to
capture count sequences. [Gan et al., 2015] develops a tem-
poral sigmoid belief network for count time series.

### 5 Experiments

We evaluate the proposed relational gamma process dynam-
ical systems, and compare it with closely-related methods,
using both synthetic and real-world count data.
Real-world data. We conducted the experiments with the following real-world datasets: (1) Integrated Crisis Early Warning System (ICEWS) dataset contains the count number of 6,000 pairwise interactions between 233 countries over 365 days. By screening out 4,800 dimensions where the sample sparsity exceeds 99%, we used a subset of ICEWS data which contains $V = 1,200$ dimensions, and $T = 365$ time steps; (2) Last.fm contains the listening information of 7,071 music artists over 51 months, where we have $T = 51$ time steps, and $V = 7,071$ dimensions; (3) Earthquake Reports Database (EQDB): records more than 120,000 earthquake reports over 15,000 earthquakes whose epicenters in the United States and nearby U.S. territories from 1928 to 1985. We created a count matrix where each column represents a month and each row represents a state. The EQDB used in the experiments, contains $T = 696$ time steps, and $V = 64$ dimensions. (4) COVID-19 contains the daily death toll in the fifty states and Washington DC of the United States, from March 2020 to March 2021. We have $T = 365$ time steps, and $V = 51$ dimensions.

Baselines. In the experiments we compared the predictive performance of the proposed models with (1) the gamma process dynamic Poisson factor analysis (GaP-DPFA) [Acharya et al., 2015], in which the gamma Markov chain involves independently over time; (2) the Poisson-gamma dynamical system (PGDS) [Schein et al., 2016a], in which a transition kernel is used to capture the excitations among latent gamma Markov chains; (3) the Poisson-randomized gamma dynamical system (PRGDS) [Schein et al., 2019] where the Poisson-randomized gamma Markov chain structure can capture a certain amount of bursty dynamics.

We denote the proposed negative-binomial-randomized gamma dynamical system as NBRGDS. The proposed NBRGDS with factor-structured prior imposed over the transition kernel, is denoted as FS-NBRGDS. The proposed NBRGDS with graph-structured prior placed over the transition structure, is denoted by GS-NBRGDS.

To evaluate the performance of the compared models in capturing heterogeneous overdispersed behaviours of latent dynamic processes behind count sequences, we considered a subset of ICEWS data that consists of heterogeneous overdispersed counts. More specifically, we sorted the observed dimensions according to their variance/expectation ratio, and selected the first $L$ dimensions in descending order, i.e., those dimensions with larger variance/expectation ratio.

We present the result for the setting $L = 300$. We treated the 80 percent of the data as the training set, and the remaining 20 percent as the test set. Then, we trained all the compared models with the training set, and evaluated the model performance using the test set. In the experiments, we used mean absolute error (MAE) and mean relative error (MRE) to evaluate the model performance in fitting count sequences:

$$\text{MAE} = \frac{1}{VT} \sum_{v=1}^{V} \sum_{t=1}^{T} |n_v(t) - \hat{n}_v(t)|,$$

$$\text{MRE} = \frac{1}{VT} \sum_{v=1}^{V} \sum_{t=1}^{T} \frac{|n_v(t) - \hat{n}_v(t)|}{1 + n_v(t)},$$

where $n_v(t)$ and $\hat{n}_v(t)$ denotes the ground true value and estimated value of dimension $v$ at time $t$, respectively. They differ because MRE considers the relative magnitude of the errors in relation to the actual values, taking into account the scale of the data, while MAE simply measures the absolute magnitude of the errors without considering the data scale. Fig. 4 shows the results of the compared models averaged over ten random training-testing repeats.

As shown in Fig. 4, NBRGDS has started to converge to its steady states after almost $10^3$ iterations, while both PGDS and PRGDS start to converge until $10^5$ iterations. Fig. 4 (b) and (c) compares the mean absolute errors and mean relative errors of the compared models, respectively. Overall, NBRGDS achieves the lowest MAE and MRE. PRGDS performs better than PGDS as PRGDS can capture a certain amount of overdispersion effects via its Poisson-randomized chain structure. We also note that NBRGDS with a time-varying scaling factor $\delta(t)$, performs better than stationary NBRGDS because this scaling factor $\delta(t)$ can also capture bursty dynamics. Nonetheless, stationary NBRGDS still out-
performs both the PRGDS and PGDS with time-varying $\delta$ magnitude (the ratio of variance to expectation) of the synthetic count sequences.

Figure 5: The proposed NBRGDS consistently achieves lower mean error (MRE), lower values are better.

Table 1: Results for the data smoothing ("S") and future data forecasting ("F") tasks. For both mean absolute error (MAE) and mean relative error (MRE), lower values are better.

Table:  
<table>
<thead>
<tr>
<th>Data Set</th>
<th>Model</th>
<th>MAE</th>
<th>MRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICEWS</td>
<td>GaP-DPFA</td>
<td>1.29±0.01</td>
<td>1.04±0.00</td>
</tr>
<tr>
<td></td>
<td>PGDS</td>
<td>1.06±0.01</td>
<td>1.04±0.00</td>
</tr>
<tr>
<td></td>
<td>PRGDS</td>
<td>1.04±0.00</td>
<td>1.04±0.00</td>
</tr>
<tr>
<td></td>
<td>NBRGDS</td>
<td>1.04±0.00</td>
<td>1.04±0.00</td>
</tr>
<tr>
<td></td>
<td>FS-NBRGDS</td>
<td>1.04±0.00</td>
<td>1.04±0.00</td>
</tr>
<tr>
<td></td>
<td>GS-NBRGDS</td>
<td>1.04±0.00</td>
<td>1.04±0.00</td>
</tr>
<tr>
<td>Last.fm</td>
<td>GaP-DPFA</td>
<td>1.71±0.03</td>
<td>1.39±0.01</td>
</tr>
<tr>
<td></td>
<td>PGDS</td>
<td>1.38±0.01</td>
<td>1.37±0.01</td>
</tr>
<tr>
<td></td>
<td>PRGDS</td>
<td>1.36±0.01</td>
<td>1.36±0.01</td>
</tr>
<tr>
<td></td>
<td>NBRGDS</td>
<td>1.36±0.01</td>
<td>1.36±0.01</td>
</tr>
<tr>
<td></td>
<td>FS-NBRGDS</td>
<td>1.36±0.01</td>
<td>1.36±0.01</td>
</tr>
<tr>
<td></td>
<td>GS-NBRGDS</td>
<td>1.36±0.01</td>
<td>1.36±0.01</td>
</tr>
</tbody>
</table>

Figure 5: The proposed NBRGDS consistently achieves lower mean absolute and relative errors, when we vary the overdispersion magnitude (the ratio of variance to expectation) of the synthetic count sequences, compared with the other closely-related models.

performs both the PRGDS and PGDS with time-varying $\delta(t)$. We conjecture that this improved prediction accuracy is because the time-varying scaling factor $\delta(t)$ fail to capture the underlying overdispersed behaviours, although it still can model a certain amount of bursty dynamics in observed dimensions. This observation further demonstrates the strong ability of the NBRGDS in capturing heterogeneous overdispersion effects of the latent dimensions behind count sequences.

Synthetic data. To further evaluate the performance of the compared models in capturing overdispersion effects, we also considered generating synthetic data with heterogeneous overdispersed dynamics. To that end, we considered to simulate synthetic data using zero-inflated negative-binomial (ZINB) models given by

$$f_{\text{ZINB}}(n \mid p_0, r, p) = p_0 f_0(n) + (1 - p_0)f_{\text{NB}}(n \mid r, p),$$

where $f_{\text{ZINB}}$ and $f_{\text{NB}}$ represents the probability mass function (PMF) of the zero-inflated negative-binomial distribution and negative-binomial distribution, respectively. Here, $f_0(n)$ is an indicator function that takes 1 when $n = 0$, otherwise 0. The parameter $p_0 \in [0, 1]$ controls the ratio of zero counts, while $r$ and $p$ are the two parameters of the negative-binomial distribution. Hence, we can effectively control the sparsity and overdispersion magnitude of the dimensions by tuning the values of $p_0$ and $p$, respectively.

More specifically, we generated five groups of synthetic data, in which each group contains $V = 10$ dimensions and $T = 365$ time steps, using the following configurations: (1) $p_0 = 0.9$, $r = 5$, $p = 0.9$, $V/E = 1.6$; (2) $p_0 = 0.9$, $r = 5$, $p = 0.8$, $V/E = 2.3$; (3) $p_0 = 0.9$, $r = 5$, $p = 0.7$, $V/E = 3.3$; (4) $p_0 = 0.9$, $r = 5$, $p = 0.6$, $V/E = 4.7$; (5) $p_0 = 0.9$, $r = 5$, $p = 0.5$, $V/E = 6.5$, where $V$ and $E$ represents variance and expectation of each group data, respectively. Then $V/E$ denotes the ratio of variance to expectation, and thus measures overdispersion effects. Fig. 5 plots the model performance of the compared methods by varying the ratio of variance to expectation. NBRGDS models including its factor-structured and graph-structured versions, consistently outperforms the other methods. In particular, NBRGDS achieves a significant improvement compared with PRGDS, although PRGDS still can capture a certain amount of overdispersion effects. Additional experiments on synthetic data under different configurations are available in the supplementary material.

Data Smoothing and Forecasting. To quantitatively evaluate the predictive performance of the compared methods, we considered two standard tasks: data smoothing and future data forecasting. The data smoothing is to predict $y_{v(t)}$ given the remaining observations $y_{v(1:T) \setminus y_{v(t)}}$, while the future data prediction is to predict future observations at next $S$ time steps $y_{v(t+1):(T+S)}$ given the history up to $T$, $y_{v(1:T)}$. Here we considered to predict next two time steps ($S = 2$). We used the default settings of GaP-DPFA, and PRGDS as

Assume an random variable $x \sim \text{ZINB}(p_0, r, p)$. The expectation and variance of $x$ are $E[x] = r(1 - p_0)(1 - p)/p$, and $\text{Var}[x] = (1 - p_0)r(1 - p)/p^2 + p_0(1 - p_0)r^2(1 - p)^2/p^2$, respectively. Thus, the ratio of variance to expectation of $x$ is $(1 + r p_0(1 - p))/p$.  

Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)  
4179
Figure 6: The latent graph structure inferred by the proposed method on ICEWS data

provided in the corresponding releases [Acharya et al., 2015; Schein et al., 2019]. For NBRGDS, FS-NBRGDS and GS-NBRGDS, we choose $K = 100$ when $V \geq 1000$, while the dimensions of EQDB and COVID-19 datasets are smaller than 100, thus we choose $K = 25$. We set $C = K$ for FS-NBRGDS and GS-NBRGDS. We ran 5000 iterations of the Gibbs sampler, which have started to converge after 1000 iterations. We discarded the first 3000 samples which were treated as burn-in time and collected a posterior sample every tenth sample thereafter. Tab. 1 shows the results of the compared methods in these two tasks. Overall, NBRGDS outperforms the GaP-DPFA, PGDS and PRGDS on almost all the datasets. In particular, we found that FS-NBRGDS and GS-NBRGDS show superior performance in future data forecasting. We conjecture this improved prediction accuracy is due to that FS-NBRGDS and GS-NBRGDS can effectively leverage the structure information underlying dynamic count data, and thus yields better predictive accuracy. We provide more comparative results on data smoothing and forecasting over different models in appendix Sec.D.

6 Conclusion

Novel negative-binomial-randomized gamma dynamical systems, have been proposed to capture the heterogeneous overdispersed behaviors of latent dynamics behind count time sequences. The new framework demonstrates more explainable latent structure, by learning the factor structure and sparse graph structure of the transition kernels, compared with transition kernel by non-informative priors. Although the prior specification of the proposed framework lacks conjugacy, tractable-yet-efficient sampling algorithms are developed to perform posterior inference. In the future, we plan to capture time-varying graph-structured transition dynamics, which will enable to better understand non-stationary count sequences. We are also considering to enhance the modeling capacities of gamma belief networks [Zhou et al., 2016; Zhou, 2018] and convex polytopes [Zhou, 2016b] using the negative-binomial-randomized gamma Markov processes.
Acknowledgments
This work was partly supported by the Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems.

References


