
Poisson–Gamma Modeling of Inter-Relational Dependencies in Dynamic Knowledge Graphs

Nan Fang^{1,2}

Yijun Wang^{1,3}

Hao Liao²

Sikun Yang^{*1,4}

¹School of Computing and Information Technology, Great Bay University, Dongguan, Guangdong, China

²College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong, China

³Tsinghua University, Shenzhen, Guangdong, China

⁴Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems

Abstract

Dynamic knowledge graphs are ubiquitous in today’s AI applications, as we represent molecular structures, social relationships, and language information using these graph models. As knowledge graphs evolve over time and are often noisy and incomplete, modeling their temporal and relational dependencies becomes crucial for downstream tasks. To address these challenges, this paper proposes PGRE (Poisson–Gamma Relational Evolution), a probabilistic model for modeling inter-relational dependencies in dynamic knowledge graphs. PGRE represents multi-relational temporal links via a Poisson–Bernoulli formulation. It introduces Gamma-distributed latent variables to capture entity–factor associations and cross-relation dependencies mediated by shared latent communities. A Gamma Markov process further models the temporal evolution of these latent variables, enabling principled characterization of relational dynamics. Experiments on benchmark datasets show that PGRE achieves competitive performance in link prediction, particularly in sparse settings, while revealing meaningful relational evolution patterns in dynamic knowledge graphs.

1 INTRODUCTION

Knowledge graphs (KGs) [Ji et al., 2021] provide a fundamental representation of structured knowledge about entities and relations, and support a wide range of applications including question answering [Yasunaga et al., 2021, Chakraborty et al., 2021, Jia et al., 2021], recommender systems [Wang et al., 2019b, 2021, 2019a], and semantic search [Thingbaijam et al., 2024, Xiong et al., 2017, Ehrlinger and Wöß, 2016]. In practical settings, however,

knowledge is inherently dynamic: relations may emerge, evolve, or disappear over time. This motivates the study of dynamic knowledge graphs (DKGs) [Liang et al., 2024]. DKGs model multi-relational interactions as temporal event sequences to predict future or missing relations from historical observations.

Knowledge graph completion has been extensively studied over the past decade, with significant progress driven by representation learning methods based on neural architectures [Schlichtkrull et al., 2018, Trivedi et al., 2017, Rossi et al., 2020, Jin et al., 2019]. Despite their strong empirical performance, these approaches often rely on large amounts of training data and exhibit limited interpretability due to their black-box nature [Chen and Ying, 2023, Seo et al., 2024]. Such limitations become more pronounced in sparse or small-sample settings, which are common in real-world dynamic knowledge graphs [Huang et al., 2023, Zhou et al., 2022].

At the same time, relations in knowledge graphs are rarely independent. Different relation types may exhibit structured dependencies, and relational facts observed in the past can directly influence future interactions [Schlichtkrull et al., 2018, Trivedi et al., 2017]. As illustrated in Fig. 1, such dependencies often manifest as structured relation transitions over time, where the relation of an entity pair at time t depends on its state at the previous time step $t - 1$. Ignoring these inter-relational and temporal dependencies may result in an incomplete characterization of relational evolution. Bayesian approaches provide a principled alternative by explicitly modeling uncertainty and structured dependencies [Acharya et al., 2015, Yang and Koepl, 2018b,a, Schein et al., 2019, Yang and Koepl, 2020, Yang and Zha, 2023, 2024], offering a natural framework for capturing relational dynamics while maintaining interpretability.

In this paper, we propose PGRE (Poisson–Gamma Relational Evolution), a probabilistic model for dynamic knowledge graph completion. PGRE represents multi-relational temporal links through a Poisson–Gamma latent-variable

*Corresponding author: sikunyang@gbu.edu.cn

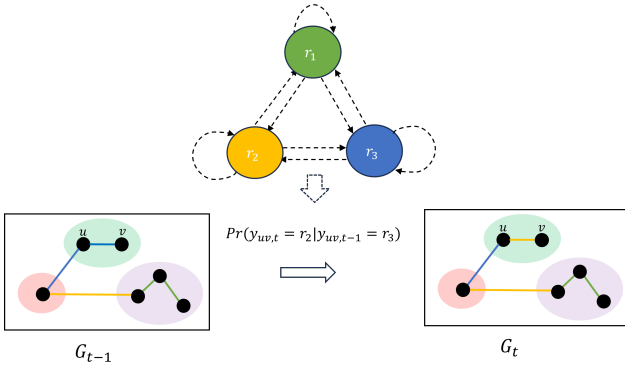


Figure 1: Illustration of a relation transition prior and its manifestation in dynamic knowledge graphs. The top panel shows an illustrative prior over relation transitions, where directed edges encode assumed transition probabilities between relation types. The bottom panel provides an example of how the relation between an entity pair may change from time $t - 1$ to t , depending on its previous relation (e.g., $\Pr(y_{uv,t} = r_2 \mid y_{uv,t-1} = r_3)$).

formulation and introduces a relation transition matrix within Gamma-Markov dynamics to explicitly capture inter-relational temporal dependencies. By jointly modeling entity–factor associations and relation-to-relation evolution in a unified generative framework, PGRE provides interpretable latent representations and supports tractable posterior inference for dynamic link prediction.

The main contributions of this paper are summarized as follows:

1. PGRE, a probabilistic model for dynamic knowledge graph completion, is proposed to jointly model entity–factor associations and inter-relational temporal dynamics.
2. Single-relation probabilistic modeling is extended to a multi-relational setting, enabling structured characterization of relational evolution in dynamic knowledge graphs.
3. An efficient Gibbs sampling algorithm based on negative-binomial–logarithmic data augmentation is developed for tractable posterior inference.
4. Extensive experiments on benchmark datasets are conducted, demonstrating that PGRE achieves competitive or superior link prediction performance, particularly in sparse and small-sample regimes, while revealing meaningful relational evolution patterns.

2 RELATED WORK

Dynamic knowledge graph completion has been extensively studied using neural representation learning methods. Early

approaches extend static embedding models by incorporating temporal information, including TTransE [García-Durán et al., 2018], TA-TransE and TA-DistMult [Leblay and Chekol, 2018], HyTE [Dasgupta et al., 2018], DacKGR [Lv et al., 2020], and DKGE [Wu et al., 2022], which introduce time-aware embeddings or temporal constraints to model relation evolution. Subsequent work combines graph neural networks with sequence modeling to capture both structural context and temporal dynamics. Representative models include DyRep [Trivedi et al., 2019], RE-NET [Jin et al., 2019], and RE-GCN [Li et al., 2021], which leverage attention mechanisms, recurrent units, or evolutionary aggregation to predict future relations. More recently, large language models have been leveraged for temporal knowledge graph forecasting to improve cross-scenario generalization [Bai et al., 2025, Tang et al., 2025]. While these neural approaches achieve strong empirical performance on large datasets, they typically rely on extensive training data, offer limited interpretability and face challenges in sparse or small-sample regimes.

Probabilistic modeling provides a principled alternative for dynamic network analysis by explicitly characterizing uncertainty and temporal evolution. Early work builds on the stochastic block model (SBM) [Holland et al., 1983, Nowicki and Snijders, 2001] and its nonparametric extensions such as IRM [Kemp et al., 2006] and GP-EPM [Zhou, 2015]. Dynamic extensions include the dynamic stochastic block model of Matias and Miele [2017], D-NGPPF [Acharya et al., 2015], which models smoothly evolving communities via gamma-Markov chains, and DPGM [Yang and Koepl, 2018b], which allows time-varying community memberships. A recent extension, G-HSEPM [Yu et al., 2025], introduces hierarchical transition structures to characterize latent community evolution in dynamic networks. However, these models typically focus on single undirected networks and do not explicitly address multi-relational dynamics. Another line of research studies temporal interaction data by reshaping dynamic knowledge graphs into event–time count matrices, leading to models such as PGDS [Schein et al., 2016], PRGDS [Schein et al., 2019], and NBRGDS [Huang et al., 2024]. Although effective for modeling temporal intensities, these approaches capture multi-relational structure indirectly through data transformation rather than directly modeling relation interactions. As a result, Bayesian modeling of dynamic multi-relational knowledge graphs remains relatively underexplored.

3 THE PROPOSED DYNAMIC KNOWLEDGE GRAPH MODEL

This section begins by defining the problem setting, followed by a detailed description of the proposed Poisson–Gamma model for modeling and tracking relational evolution in dynamic knowledge graphs.

3.1 NOTATION AND PROBLEM DEFINITION

A temporal knowledge graph (TKG) represents time-evolving relational facts as a set of time-stamped quadruples (s, r, o, t) , where s and o denote entities, r denotes a relation type, and t is a discrete timestamp. The collection of events at time t forms a snapshot G_t , and a TKG over T time steps is represented as a sequence $\mathcal{G} = \{G_1, \dots, G_T\}$.

For modeling convenience, we represent the TKG as a binary tensor $\mathcal{M} \in \{0, 1\}^{T \times R \times N \times N}$, where $\mathcal{M}[t][r][s][o] = 1$ indicates the presence of relation r from s to o at time t . Temporal knowledge graph completion then amounts to predicting missing entries of $\mathcal{M}$. In this work, we focus on the forecasting setting, where the objective is to infer future relations based on historical observations. Given historical snapshots up to time $t_q - 1$, the goal is to infer whether a relation (s, r, o) holds at a query time t_q .

3.2 POISSON-GAMMA RELATIONAL EVOLUTION

We first describe how to represent the structure of latent communities at a single time step, and then explain how the influence of these latent communities on relation types evolves over time. Throughout the paper, we use Gam, Pois, Dir, Bern, and NB to denote the Gamma, Poisson, Dirichlet, Bernoulli, and negative binomial distributions, respectively.

To characterize the community structure of knowledge graphs, we associate each entity with two nonnegative *entity-community factor* vectors, corresponding to its roles as a subject and as an object. These vectors capture role-specific soft affiliations with multiple latent communities, while their absolute scales are not interpreted in isolation. Instead, their relative patterns, together with the relation- and time-specific community weights, determine the strength of multi-relational interactions. Intuitively, entities may participate in multiple types of interactions simultaneously. For example, a researcher may collaborate with some peers while advising others. Such overlapping and role-dependent patterns can be represented through different subject- and object-side affiliations with multiple latent communities. Specifically, we assume that the dynamic knowledge graph contains K latent communities. The subject-side factor loading of entity i on community k and the object-side factor loading of entity j on community k are assigned Gamma priors as $\phi_{ik} \sim \text{Gam}(a_0, 1/c_i)$ and $\psi_{jk} \sim \text{Gam}(a_1, 1/c_j)$, respectively, where $\text{Gam}(\alpha, \theta)$ denotes a Gamma distribution with shape α and scale θ . The hyperparameters a_0 and a_1 control the prior shapes, while c_i and c_j regulate the magnitudes of the subject- and object-side community factor loadings, respectively.

We assume that latent communities evolve over time and are influenced by relation types. Therefore, we introduce

a time- and relation-dependent variable $\delta_k^{(t,r)}$ to represent the state of community k under relation r at time t . Intuitively, different relation types in a knowledge graph are not independent, and their dynamics can influence each other through shared latent community structures. Entities sharing similar semantic contexts often exhibit correlated relational behaviors over time. For example, the emergence of a ‘‘collaborates with’’ relation between two researchers may increase the likelihood of a subsequent ‘‘publishes with’’ relation within the same community. To capture such dependencies, we model the state of community k under relation r at time t as $\delta_k^{(t,r)} \sim \text{Gam}\left(\sum_{r'=1}^R \pi_{rr'} \delta_k^{(t-1,r')}, 1/\tau\right)$, which indicates that the weight of community k under relation r at time t may be influenced by the weights of the same community across all relations at the previous time step $t - 1$. The strength of this influence is determined by the coefficient vector $\Pi_r = [\pi_{r1}, \pi_{r2}, \dots, \pi_{rR}]^T$. In particular, we draw the initial community weight under relation r from a Gamma prior as $\delta_k^{(1,r)} \sim \text{Gam}(\nu_r/K, 1/\tau)$. We impose a Dirichlet prior on the transition kernel as $\pi_r \sim \text{Dir}(\nu_1\nu_r, \dots, \xi\nu_r, \dots, \nu_R\nu_r)$, and draw the hyperparameter ν_r from a Gamma prior as $\nu_r \sim \text{Gam}(\gamma_0/R, 1/\beta)$, where γ_0 is the concentration parameter and β is the hyperparameter.

It is worth noting that as the number of latent communities K approaches infinity, the hierarchical Gamma prior exhibits an inherent shrinkage property that drives the weights of redundant communities toward zero. This property enables the model to automatically infer an appropriate number of active communities from data, rather than relying on a manually fixed K , thereby improving flexibility and interpretability in modeling real-world dynamic knowledge graphs.

Given the subject- and object-side community factor loadings ϕ_i and ψ_j , together with the relation- and time-dependent community weights $\delta_k^{(t,r)}$, we model the probability of a link of type r from subject entity i to object entity j at time t using the Bernoulli–Poisson link function as

$$\mathcal{M}[t][r][i][j] \sim \text{Bern}\left(1 - \exp\left(-\sum_{k=1}^K \delta_k^{(t,r)} \phi_{ik} \psi_{jk}\right)\right). \quad (1)$$

Equivalently, this formulation can be expressed as

$$\mathcal{M}_{ij}^{(t,r)} = \mathbb{1}(x_{ij}^{(t,r)} \geq 1), \quad (2)$$

$$x_{ij}^{(t,r)} \sim \text{Pois}\left(\sum_{k=1}^K \delta_k^{(t,r)} \phi_{ik} \psi_{jk}\right), \quad (3)$$

where $\mathcal{M}_{ij}^{(t,r)}$, equivalent to $\mathcal{M}[t][r][i][j]$ in the tensor representation, indicates whether the event (i, r, j) occurs at time t . This Bernoulli–Poisson formulation effectively captures the discrete and sparse nature of temporal multi-relational data and enables tractable posterior inference.

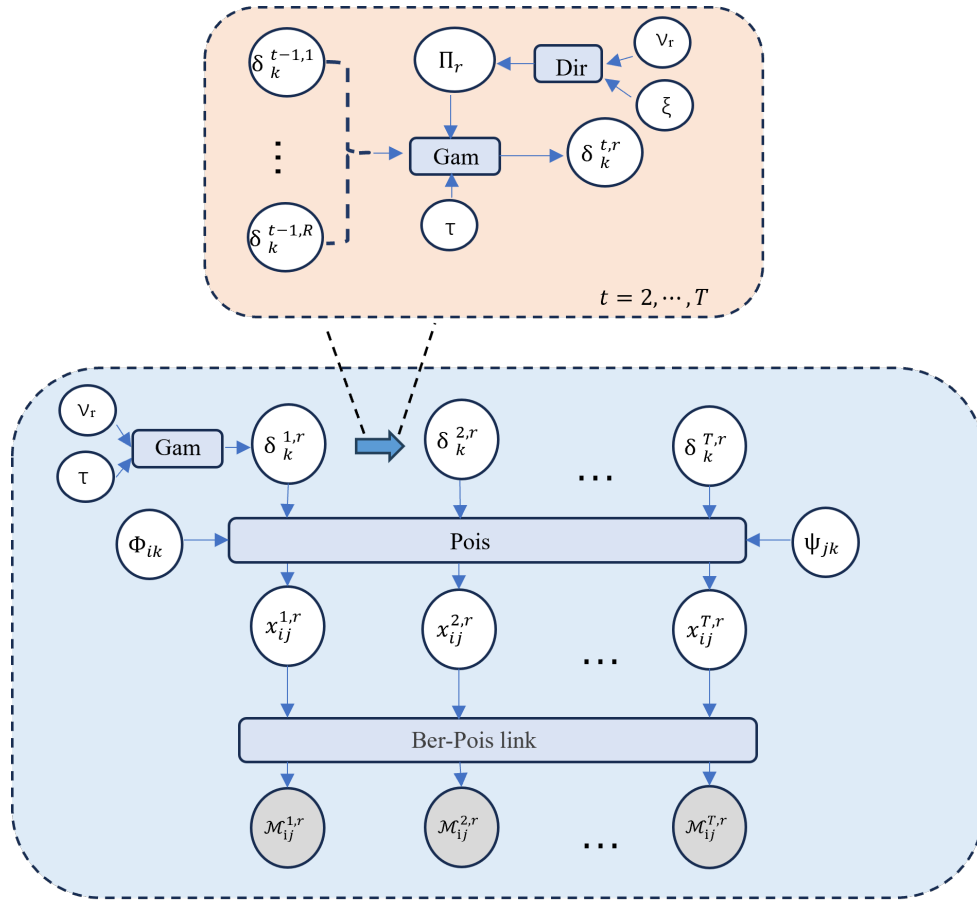


Figure 2: The overall structure of the proposed PGRE model. The upper part illustrates the hierarchical Gamma–Dirichlet process governing the temporal and relational evolution of community weights $\delta_k^{(t,r)}$, while the lower part shows the Poisson–Bernoulli generative process for multi-relational links over time.

The full generative Poisson–Gamma relational model is specified as follows.

$$\mathcal{M}_{ij}^{(t,r)} = \mathbb{1}(x_{ij}^{(t,r)} \geq 1), \quad (4)$$

$$x_{ij}^{(t,r)} \sim \text{Pois}\left(\sum_{k=1}^K \delta_k^{(t,r)} \phi_{ik} \psi_{jk}\right), \quad (5)$$

$$\phi_{ik} \sim \text{Gam}(a_0, 1/c_i), \quad \psi_{jk} \sim \text{Gam}(a_1, 1/c_j), \quad (6)$$

$$\delta_k^{(t,r)} \sim \text{Gam}\left(\sum_{r'=1}^R \pi_{rr'} \delta_k^{(t-1,r')}, 1/\tau\right), \quad t = 2, \dots, T, \quad (7)$$

$$\delta_k^{(1,r)} \sim \text{Gam}(\nu_r/K, 1/\tau), \quad (8)$$

$$\pi_r \sim \text{Dir}(\nu_1 \nu_r, \dots, \xi \nu_r, \dots, \nu_R \nu_r), \quad (9)$$

$$\nu_r \sim \text{Gam}(\gamma_0/R, 1/\beta). \quad (10)$$

Here, c_i , c_j , ξ , and β are assigned Gamma priors, while hyperparameters such as a_0 , a_1 , and γ_0 are themselves drawn from Gamma distributions. The overall structure of the proposed PGRE model is illustrated in Fig. 2.

4 INFERENCE

This section outlines the Gibbs sampling procedure for PGRE by presenting the posterior updates of key model parameters, with full derivations deferred to the supplementary material. The following standard results are recalled for completeness and are not claimed as new theoretical contributions [Zhou, 2015].

Theorem 1 Let $\lambda \sim \text{Gam}\left(r, \frac{p}{1-p}\right)$ and $y \mid \lambda \sim \text{Pois}(\lambda)$. Then the marginal distribution of y is $\text{NB}(r, p)$. Equivalently, a negative binomial random variable admits a gamma–Poisson mixture representation.

Theorem 2 The Poisson–logarithmic bivariate representation links the negative binomial distribution to the Poisson–CRT construction. If $y \sim \text{NB}(r, p)$ and $l \sim \text{CRT}(y, r)$, then $y = \sum_{s=1}^l u_s$ with $u_s \sim \text{Logarithmic}(p)$ and $l \sim \text{Pois}(-r \log(1-p))$.

Theorem 3 Let $y = \sum_{n=1}^N y_n$, where $y_n \sim \text{Pois}(\lambda_n)$ are independent Poisson random variables with rates λ_n . Then, according to the Poisson–multinomial equivalence, $(y_1, \dots, y_N) \mid y \sim \text{Mult}\left(y; \frac{\lambda_1}{\sum_n \lambda_n}, \dots, \frac{\lambda_N}{\sum_n \lambda_n}\right)$ and $y \sim \text{Pois}\left(\sum_{n=1}^N \lambda_n\right)$.

Sampling latent counts $x_{ijk}^{(t,r)}$. By Theorem 3, the latent counts are allocated across communities as

$$(x_{ijk}^{(t,r)} \mid -) \sim \text{Mult}\left(x_{ij}^{(t,r)}, \frac{\delta_k^{(t,r)} \phi_{ik} \psi_{jk}}{\sum_{k'=1}^K \delta_{k'}^{(t,r)} \phi_{ik'} \psi_{jk'}}\right). \quad (11)$$

Sampling community factor loadings ϕ_{ik} and ψ_{jk} . By Gamma–Poisson conjugacy, both subject- and object-side factor loadings admit closed-form Gamma updates:

$$(\phi_{ik} \mid -) \sim \text{Gam}\left(a_0 + \sum_{t,r,j} x_{ijk}^{(t,r)}, \left(c_i + \sum_{t,r,j} \delta_k^{(t,r)} \psi_{jk}\right)^{-1}\right), \quad (12)$$

$$(\psi_{jk} \mid -) \sim \text{Gam}\left(a_1 + \sum_{t,r,i} x_{ijk}^{(t,r)}, \left(c_j + \sum_{t,r,i} \delta_k^{(t,r)} \phi_{ik}\right)^{-1}\right). \quad (13)$$

Sampling community weights $\delta_k^{(t,r)}$. For the final time step $t = T$, applying the Gamma–Poisson conjugacy yields the following posterior update for $\delta_k^{(T,r)}$.

$$(\delta_k^{(T,r)} \mid -) \sim \text{Gam}\left(x_{..k}^{(T,r)} + \sum_{r'=1}^R \pi_{rr'} \delta_k^{(T-1,r')}, (\tau + s_k)^{-1}\right). \quad (14)$$

To incorporate forward information, we marginalize $\delta_k^{(t+1,r)}$ via Theorem 1 and introduce auxiliary variables

$$l_k^{(t,r)} \sim \text{CRT}\left(x_{..k}^{(t,r)} + l_k^{(t+1,r)}, \sum_{r'=1}^R \pi_{rr'} \delta_k^{(t-1,r')}\right), \quad (15)$$

which propagate forward information across time.

Combining backward counts and forward augmentation, the conditional posterior of $\delta_k^{(t,r)}$ is

$$(\delta_k^{(t,r)} \mid -) \sim \text{Gam}\left(x_{..k}^{(t,r)} + l_k^{(t+1,r)} + \sum_{r'=1}^R \pi_{rr'} \delta_k^{(t-1,r')}, \frac{1}{1/(\tau + s_k - \ln(1 - \rho_k^{(t+1)}))}\right), \quad (16)$$

The auxiliary parameter $\rho_k^{(t)}$ follows the recursion

$$\rho_k^{(t)} = \frac{s_k - \ln(1 - \rho_k^{(t+1)})}{\tau + s_k - \ln(1 - \rho_k^{(t+1)})}. \quad (17)$$

Sampling the transition matrix π_r . After marginalizing out δ , the auxiliary counts follow a multinomial distribution $(l_k^{(t,1r)}, \dots, l_k^{(t,Rr)}) \sim \text{Mult}(l_k^{(t,r)}, (\pi_{1r}, \dots, \pi_{Rr}))$. By the Dirichlet–multinomial conjugacy, the posterior distribution of π_r is given by

$$(\pi_r \mid -) \sim \text{Dir}\left(\nu_1 \nu_r + \sum_{t=2}^T \sum_{k=1}^K l_k^{(t,1r)}, \dots, \xi \nu_r + \sum_{t=2}^T \sum_{k=1}^K l_k^{(t,rr)}, \dots, \nu_R \nu_r + \sum_{t=2}^T \sum_{k=1}^K l_k^{(t,Rr)}\right). \quad (18)$$

Sampling ν_r and ξ . We marginalize over Π to obtain a Dirichlet–multinomial distribution

$$(l^{(\cdot,1r)}, \dots, l^{(\cdot,Rr)}) \sim \text{DirMult}\left(l^{(\cdot,r)}, (\nu_1 \nu_r, \dots, \xi \nu_r, \dots, \nu_R \nu_r)\right), \quad (19)$$

where $l^{(\cdot,r)} = \sum_{t=1}^T \sum_{k=1}^K \sum_{r'=1}^K l_k^{(t,r1r')}$.

By introducing a beta-distributed auxiliary variable q_r , the Dirichlet–multinomial distribution can be rewritten as a negative binomial distribution:

$$q_r \sim \text{Beta}\left(l^{(\cdot,r)}, \nu_r \left(\xi + \sum_{r' \neq r} \nu_{r'}\right)\right). \quad (20)$$

Then, $l^{(\cdot,rr)}$ and $l^{(\cdot,rr')}$ follow negative binomial distributions as $l^{(\cdot,rr)} \sim \text{NB}(\xi \nu_r, q_r)$ and $l^{(\cdot,rr')} \sim \text{NB}(\nu_r \nu_{r'}, q_r)$. To further facilitate posterior inference, we introduce auxiliary count variables h_{rr} and $h_{rr'}$ via the Chinese Restaurant Table (CRT) distribution:

$$\begin{aligned} h_{rr} &\sim \text{CRT}\left(l^{(\cdot,rr)}, \xi \nu_r\right), \\ h_{rr'} &\sim \text{CRT}\left(l^{(\cdot,rr')}, \nu_r \nu_{r'}\right). \end{aligned} \quad (21)$$

Using the gamma–Poisson conjugacy, ξ is then sampled as

$$(\xi \mid -) \sim \text{Gam}\left(b_0 + \sum_{r=1}^R h_{rr}, \frac{1}{e_0 - \sum_{r=1}^R \nu_r \ln(1 - q_r)}\right). \quad (22)$$

Next, we define

$$n_r = h_{rr} + \sum_{r_1 \neq r} h_{r_1 r} + \sum_{r_2 \neq r} h_{rr_2} + \sum_{k=1}^K l_k^{(1,r)},$$

Table 1: Performance comparison on three dynamic knowledge graph datasets, evaluated using AUC-PR and AUC-ROC. Best results for each metric are shown in **bold**.

Model	ICEWS18		GDELT		WIKI	
	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC	AUC-PR	AUC-ROC
PGDS	0.082 ± 0.004	0.863 ± 0.001	0.124 ± 0.014	0.836 ± 0.008	0.126 ± 0.006	0.854 ± 0.004
PRGDS	0.079 ± 0.003	0.865 ± 0.002	0.136 ± 0.016	0.843 ± 0.010	0.134 ± 0.006	0.889 ± 0.002
NBRGDS	0.071 ± 0.005	0.876 ± 0.004	0.102 ± 0.019	0.851 ± 0.009	0.437 ± 0.007	0.949 ± 0.003
Know-Evolve	0.072 ± 0.005	0.907 ± 0.002	0.105 ± 0.017	0.805 ± 0.012	0.554 ± 0.014	0.976 ± 0.007
DyRep	0.075 ± 0.006	0.914 ± 0.003	0.114 ± 0.012	0.824 ± 0.011	0.598 ± 0.015	0.984 ± 0.008
RENet	0.080 ± 0.005	0.916 ± 0.004	0.108 ± 0.019	0.837 ± 0.011	0.727 ± 0.005	0.998 ± 0.001
D-NGPPF	0.127 ± 0.004	0.930 ± 0.004	0.121 ± 0.012	0.880 ± 0.011	0.582 ± 0.003	0.987 ± 0.003
DPGM	0.105 ± 0.004	0.922 ± 0.003	0.108 ± 0.015	0.874 ± 0.008	0.498 ± 0.006	0.989 ± 0.002
G-HSEPM	0.120 ± 0.002	0.930 ± 0.002	0.119 ± 0.016	0.878 ± 0.005	0.556 ± 0.006	0.987 ± 0.004
PGRE	0.198 ± 0.004	0.936 ± 0.003	0.277 ± 0.015	0.902 ± 0.004	0.607 ± 0.003	0.998 ± 0.002

where $l_k^{(1,r)} \sim \text{Pois}(-\tau \nu_r \ln(1 - \rho_k^{(1)}))$. Using the Poisson additive property and gamma-Poisson conjugacy, we have

$$(\nu_r | -) \sim \text{Gam}\left(\frac{\gamma_0}{R} + n_r, 1/(\beta + t_r)\right), \quad (23)$$

where $t_r = -\ln(1 - q_r) \left(\xi + \sum_{r_1 \neq r} \nu_{r_1} \right) - \sum_{r_2 \neq r} \ln(1 - q_{r_2}) \nu_{r_2} - \sum_{k=1}^K \frac{1}{K} \ln(1 - \rho_k^{(1)})$.

5 EXPERIMENTS

This section evaluates PGRE through comparative performance analysis, ablation studies, and parameter visualizations to assess both effectiveness and interpretability.

5.1 RELATION PREDICTION

Datasets. We evaluated PGRE on three widely used temporal knowledge graph datasets: ICEWS18 [Boschee et al., 2015], GDELT [Leetaru and Schrod, 2013], and WIKI [Leblay and Chekol, 2018]. To construct moderate-scale subgraphs with sufficient relational-temporal support, we first ranked relation types according to their total numbers of observed events and retained the most frequent relations, resulting in 10 relations for ICEWS18, 8 for GDELT, and 15 for WIKI. Within each relation-filtered subgraph, entity activity was measured by the total number of events in which an entity appeared as either a subject or an object. We then retained the 500 most active entities and removed events whose subject or object was not included in the selected entity set. This preprocessing strategy keeps Gibbs inference computationally manageable while ensuring sufficient observations for estimating inter-relational temporal dependencies.

Baselines. We compared PGRE with representative methods from three categories: probabilistic models for dynamic networks (DPGM [Yang and Koepl, 2018b], D-NGPPF [Acharya et al., 2015], G-HSEPM [Yu et al., 2025]), probabilistic models for dynamic count data (PGDS [Schein et al., 2016], PRGDS [Schein et al., 2019], NBRGDS [Huang et al., 2024]), and neural models for temporal knowledge graphs (Know-Evolve [Trivedi et al., 2017], DyRep [Trivedi et al., 2019], RENet [Jin et al., 2019]).

Experimental Settings. Each dataset was divided into three temporal segments: train ($t = 1, \dots, T - 2$), valid ($t = T - 1$), and test ($t = T$), where the prediction query time is $t_q = T$. Given historical events up to time $t_q - 1$, models predict the probabilities of events occurring at time t_q . Performance is evaluated using AUC-ROC and AUC-PR. More experimental details are provided in the supplementary material.

Results on Relation Prediction. Table 1 summarizes the relation prediction performance on ICEWS18, GDELT, and WIKI. PGRE consistently achieves competitive results across all datasets, with particularly large improvements in AUC-PR on ICEWS18 and GDELT, both of which exhibit sparse and highly multi-relational structures. Even on WIKI, where relational dependencies are weaker, PGRE attains the second-best AUC-PR, indicating robust generalization across varying relational densities.

The superior AUC-PR performance of PGRE primarily stems from two modeling choices. First, PGRE captures latent transition dependencies among relations through the transition matrix $\Pi_{rr'}$, enabling the model to represent how different relations evolve and interact over time. Second, the explicit modeling of directed head-tail entity roles allows PGRE to capture asymmetric temporal dynamics that are ignored by conventional single-relation models.

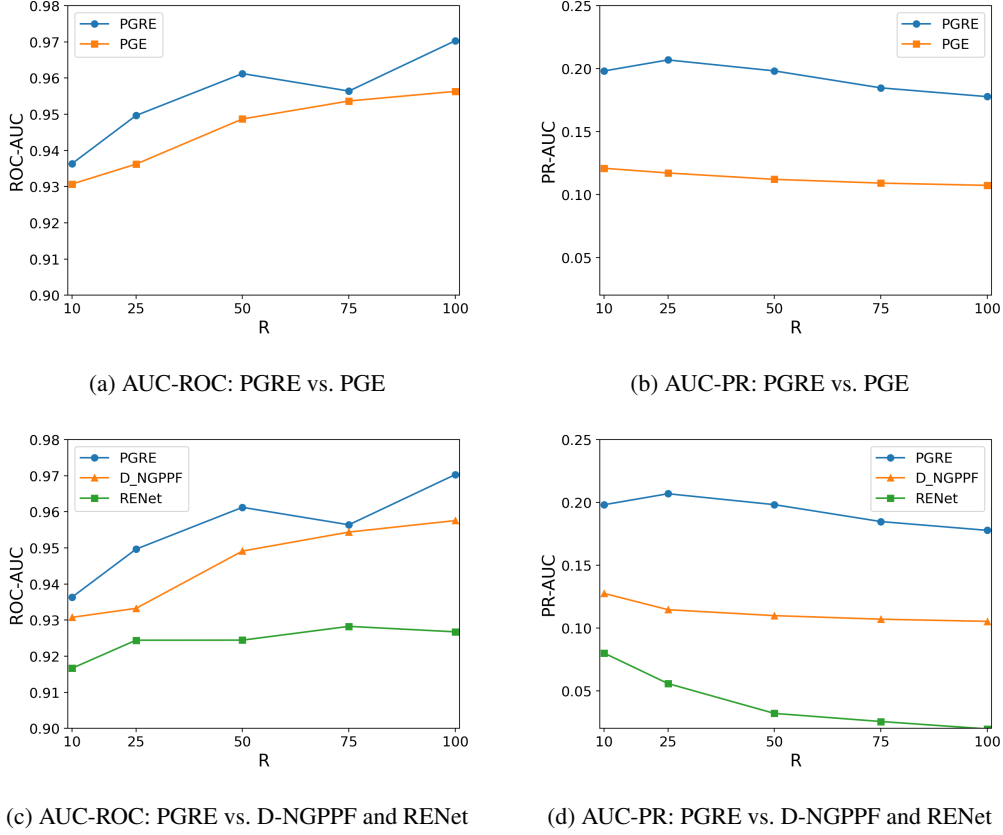


Figure 3: Comparison of AUC-ROC and AUC-PR under different relation counts $R \in \{10, 25, 50, 75, 100\}$. The first row compares PGRE with its variant PGE (without the relation transition matrix), while the second row compares PGRE with two strong baselines, D-NGPPF and RENet.

Table 2: Average time per iteration (seconds) on different datasets.

Model	ICEWS18	GDELT	WIKI
PGDS	73.28	135.52	51.91
PRGDS	75.07	117.54	54.85
NBRGDS	72.11	79.06	46.55
D-NGPPF	140.70	329.81	81.50
DPGM	172.78	459.88	88.86
G-HSEPM	204.01	339.26	145.98
PGRE	71.70	185.16	16.19

Compared with both count-based probabilistic methods and neural temporal models, PGRE is more robust under sparse multi-relational settings. While neural models often achieve higher AUC-ROC by focusing on frequent patterns, PGRE maintains substantially better AUC-PR by leveraging structured Bayesian priors that couple relations and entities, resulting in improved recall and interpretability.

Table 2 reports the average time per iteration on three benchmarks. PGRE achieves the lowest runtime on the sparsest

dataset (WIKI) and remains competitive on denser datasets such as ICEWS18 and GDELT. Overall, PGRE provides a favorable trade-off between computational efficiency and modeling expressiveness in multi-relational temporal settings.

5.2 ABLATION AND INTERPRETABILITY ANALYSIS

We conduct a mechanism-oriented relation-cardinality ablation on ICEWS18 to examine how relation coupling behaves as the number of retained relation types increases. D-NGPPF and RENet are selected as the strongest probabilistic and neural baselines, respectively, in the main comparison. Fig. 3 reports the performance under varying numbers of relations R . The AUC-PR of PGRE initially increases, reaches its maximum at $R = 25$, and then gradually decreases, while remaining consistently higher than those of the competing methods. Its AUC-ROC exhibits an overall upward trend, with a slight fluctuation at $R = 75$. In contrast, the AUC-PR values of the competing methods generally decline as R increases, suggesting that they benefit less from the additional cross-relation structure.

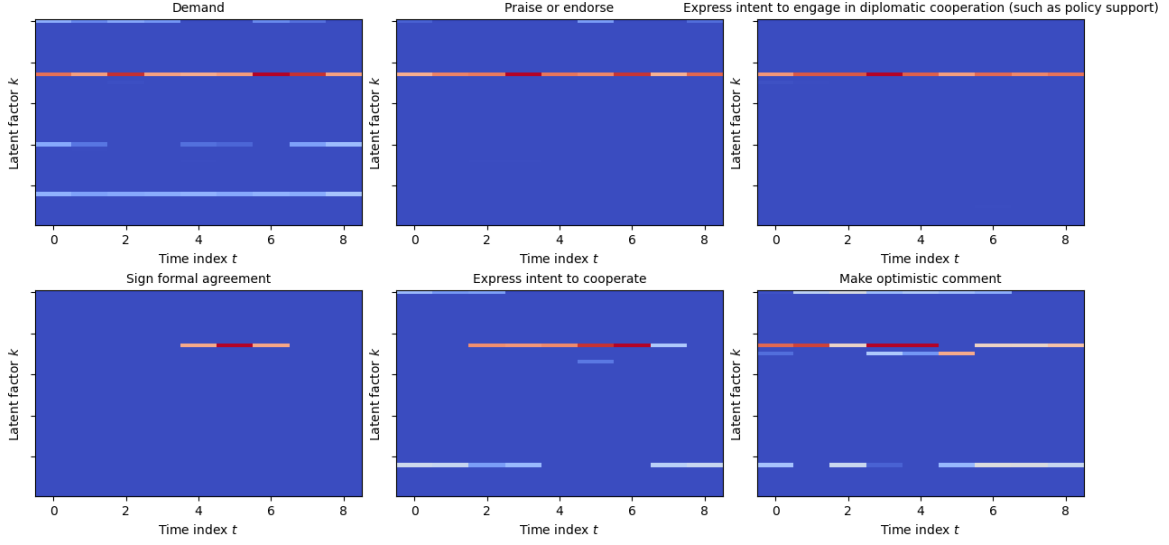


Figure 4: Temporal evolution of latent factors for six representative relations: *Demand*, *Praise or endorse*, *Express intent to engage in diplomatic cooperation*, *Sign formal agreement*, *Express intent to cooperate*, and *Make optimistic comment*, $\delta_k^{t,r} / (\sum_{k=1}^K \delta_k^{t,r})$. The y-axis represents latent factor indices (150–199) and the x-axis represents time steps t . Redder colors indicate higher activity levels.

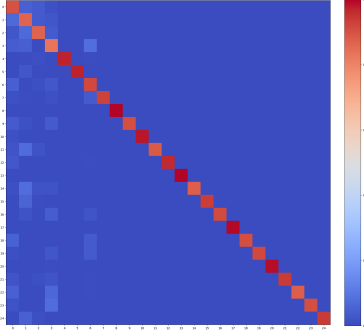


Figure 5: Transition probabilities among the 25 most active relations in the $R = 25$ relation-cardinality setting of ICEWS18.

The ablation variant PGE, which removes the relation transition matrix, exhibits a substantial reduction in AUC-PR across all settings, demonstrating the contribution of explicit relation coupling. Similarly, the AUC-PR values of D-NGPPF and RENet decrease as R grows, whereas their AUC-ROC improvements are less pronounced than those of PGRE. These results indicate that the proposed relation-transition mechanism helps PGRE maintain more robust predictive performance as relation diversity increases.

To further examine the learned structure, we visualize the temporal evolution of latent factors and the relation transition matrix $\Pi_{T,T'}$. Fig. 4 illustrates the temporal evolution of latent factors for representative positive interaction types. Certain latent factors remain consistently active across multiple relations, suggesting the presence of shared latent com-

munities that drive their temporal dynamics. Fig. 5 shows the transition probabilities in the $R = 25$ setting. The matrix displays a dominant diagonal pattern, reflecting strong temporal self-dependence of relations across adjacent time steps. In addition, several off-diagonal bands reveal structured transitions between semantically related relations, indicating dynamic coupling beyond self-persistence. These observations demonstrate that PGRE captures both interpretable latent evolution and structured inter-relational interactions, which jointly contribute to its robustness in sparse multi-relational settings.

6 CONCLUSION

We have proposed a probabilistic network model for capturing temporal evolution and relational interactions in dynamic knowledge graphs. In contrast to existing Bayesian network models, the proposed framework is directly applicable to dynamic knowledge graph completion. The model has introduced a structured relation transition kernel to characterize inter-relational dynamics within latent communities and has supported efficient closed-form posterior inference via Gibbs sampling. Experimental results on multiple temporal knowledge graph benchmarks have demonstrated competitive relation prediction performance, with substantial AUC-PR improvements on ICEWS18 and GDELT, together with interpretable relational and temporal structures revealed through ablation and visualization analyses.

Future work includes extending the model to settings with rapidly evolving event streams and learning time-varying relation transition dynamics. Another promising direction is

adapting the framework to heterogeneous temporal networks with irregular or partially overlapping node sets.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (NSFC) (Grant Nos. 62476047 and 62276171), the Peking University Mathematics Challenge Funding Program (Grant No. 2024SRMC10), the Shenzhen Science and Technology Program (Grant Nos. ZDCY20250901110940006 and JCYJ20240813141503005), the Dongguan Key Laboratory for AI and Dynamical Systems, the Dongguan Key Laboratory for Intelligence and Information Technology, the Dongguan Key Laboratory for Data Science and Intelligent Medicine, the Guangdong Research Team for Communication and Sensing Integrated with Intelligent Computing (Project No. 2024KCXTD047), and the Guangdong Multidisciplinary Innovation Research Group for New-Generation Intelligent Systems and Diagnostic-Therapeutic Applications (Grant No. 2025KCXTD031).

References

- Ayan Acharya, Avijit Saha, Mingyuan Zhou, Dean Teffer, and Joydeep Ghosh. Nonparametric dynamic network modeling. In *KDD Workshop on Mining and Learning from Time Series*, 2015.
- Long Bai, Zixuan Li, Xiaolong Jin, Jiafeng Guo, Xueqi Cheng, and Tat-Seng Chua. G2s: A general-to-specific learning framework for temporal knowledge graph forecasting with large language models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20927–20938, 2025.
- Elizabeth Boschee, Jennifer Lautenschlager, Sean O’Brien, Steve Shellman, James Starz, and Michael Ward. ICEWS Coded Event Data. <https://dataverse.harvard.edu/dataverse/icews>, 2015. Harvard Dataverse, accessed October 19, 2025.
- Nilesh Chakraborty, Denis Lukovnikov, Gaurav Maheshwari, Priyansh Trivedi, Jens Lehmann, and Asja Fischer. Introduction to neural network-based question answering over knowledge graphs. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(3):e1389, 2021.
- Jialin Chen and Rex Ying. Tempme: Towards the explainability of temporal graph neural networks via motif discovery. *Advances in Neural Information Processing Systems*, 36:29005–29028, 2023.
- Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha Talukdar. Hyte: Hyperplane-based temporally aware knowledge graph embedding. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 2001–2011, 2018.
- Lisa Ehrlinger and Wolfram Wöß. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCESS)*, 48(1-4):2, 2016.
- Alberto García-Durán, Sebastijan Dumančić, and Mathias Niepert. Learning sequence encoders for temporal knowledge graph completion. *arXiv preprint arXiv:1809.03202*, 2018.
- Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
- Rui Huang, Sikun Yang, and Heinz Koepl. Negative-binomial randomized gamma markov processes for heterogeneous overdispersed count time series. *arXiv preprint arXiv:2402.18995*, 2024.
- Shenyang Huang, Farimah Poursafaei, Jacob Danovitch, Matthias Fey, Weihua Hu, Emanuele Rossi, Jure Leskovec, Michael Bronstein, Guillaume Rabusseau, and Reihaneh Rabbany. Temporal graph benchmark for machine learning on temporal graphs. *Advances in Neural Information Processing Systems*, 36:2056–2073, 2023.
- Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S Yu. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2): 494–514, 2021.
- Zhen Jia, Soumajit Pramanik, Rishiraj Saha Roy, and Gerhard Weikum. Complex temporal question answering on knowledge graphs. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 792–802, 2021.
- Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. *arXiv preprint arXiv:1904.05530*, 2019.
- Charles Kemp, Joshua B Tenenbaum, Thomas L Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *AAAI*, volume 3, page 5, 2006.
- Julien Leblay and Melisachew Wudage Chekol. Deriving validity time in knowledge graph. In *Companion proceedings of the web conference 2018*, pages 1771–1776, 2018.
- Kalev Leetaru and Philip A. Schrod. Gdelt: Global data on events, location, and tone, 1979–2012. In *Proceedings of the International Studies Association Annual Convention*, pages 1–49. Citeseer, 2013.

- Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. Temporal knowledge graph reasoning based on evolutionary representation learning. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 408–417, 2021.
- Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, Fuchun Sun, and Kunlun He. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9456–9478, 2024.
- Xin Lv, Xu Han, Lei Hou, Juanzi Li, Zhiyuan Liu, Wei Zhang, Yichi Zhang, Hao Kong, and Suhui Wu. Dynamic anticipation and completion for multi-hop reasoning over sparse knowledge graph. *arXiv preprint arXiv:2010.01899*, 2020.
- Catherine Matias and Vincent Miele. Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1119–1141, 2017.
- Krzysztof Nowicki and Tom A B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association*, 96(455):1077–1087, 2001.
- Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*, 2020.
- Aaron Schein, Hanna Wallach, and Mingyuan Zhou. Poisson-gamma dynamical systems. *Advances in Neural Information Processing Systems*, 29, 2016.
- Aaron Schein, Scott Linderman, Mingyuan Zhou, David Blei, and Hanna Wallach. Poisson-randomized gamma dynamical systems. *Advances in Neural Information Processing Systems*, 32, 2019.
- Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Riianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer, 2018.
- Sangwoo Seo, Sungwon Kim, Jihyeong Jung, Yoonho Lee, and Chanyoung Park. Self-explainable temporal graph networks based on graph information bottleneck. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pages 2572–2583, 2024.
- Guo Tang, Zheng Chu, Wenxiang Zheng, Junjia Xiang, Yizhuo Li, Weihao Zhang, Ming Liu, and Bing Qin. Anre: Analogical replay for temporal knowledge graph forecasting. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4632–4650, 2025.
- Lenin Thingbaijam, Kowstubha Palle, P Venkata Prasad, Balasubbareddy Mallala, Shrishailappa Patil, et al. Incorporating knowledge graphs in semantic search. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–6. IEEE, 2024.
- Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *international conference on machine learning*, pages 3462–3471. PMLR, 2017.
- Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. Dyrep: Learning representations over dynamic graphs. In *International conference on learning representations*, 2019.
- Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 950–958, 2019a.
- Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5329–5336, 2019b.
- Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhenguang Liu, Xiangnan He, and Tat-Seng Chua. Learning intents behind interactions with knowledge graph for recommendation. In *Proceedings of the web conference 2021*, pages 878–887, 2021.
- Tianxing Wu, Arijit Khan, Melvin Yong, Guilin Qi, and Meng Wang. Efficiently embedding dynamic knowledge graphs. *Knowledge-based systems*, 250:109124, 2022.
- Chenyan Xiong, Russell Power, and Jamie Callan. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th international conference on world wide web*, pages 1271–1279, 2017.
- Sikun Yang and Heinz Koepl. Dependent relational gamma process models for longitudinal networks. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5551–5560. PMLR, 10–15 Jul 2018a.
- Sikun Yang and Heinz Koepl. A poisson gamma probabilistic model for latent node-group memberships in dynamic networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018b.

- Sikun Yang and Heinz Koepl. The hawkes edge partition model for continuous-time event-based temporal networks. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 460–469, 03–06 Aug 2020.
- Sikun Yang and Hongyuan Zha. Estimating latent population flows from aggregated data via inverting multi-marginal optimal transport. In *Proceedings of the 2023 SIAM International Conference on Data Mining, SDM*, pages 181–189, 2023.
- Sikun Yang and Hongyuan Zha. A variational autoencoder for neural temporal point processes with dynamic latent graphs. In *Thirty-Eighth AAAI Conference on Artificial Intelligence*, pages 16343–16351, 2024.
- Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. Qa-gnn: Reasoning with language models and knowledge graphs for question answering. *arXiv preprint arXiv:2104.06378*, 2021.
- Xincan Yu, Nan Fang, Hao Liao, and Sikun Yang. Tracking latent communities evolution with hierarchical edge partition models. In *2025 IEEE International Conference on Data Mining (ICDM)*, pages 1672–1681. IEEE, 2025.
- Hongkuan Zhou, Da Zheng, Israt Nisa, Vasileios Ioannidis, Xiang Song, and George Karypis. Tgl: A general framework for temporal gnn training on billion-scale graphs. *arXiv preprint arXiv:2203.14883*, 2022.
- Mingyuan Zhou. Infinite edge partition models for overlapping community detection and link prediction. In *Artificial intelligence and statistics*, pages 1135–1143. PMLR, 2015.

Poisson–Gamma Modeling of Inter-Relational Dependencies in Dynamic Knowledge Graphs (Supplementary Material)

Nan Fang^{1,2}

Yijun Wang^{1,3}

Hao Liao²

Sikun Yang^{*1,4}

¹School of Computing and Information Technology, Great Bay University, Dongguan, Guangdong, China

²College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong, China

³Tsinghua University, Shenzhen, Guangdong, China

⁴Guangdong Provincial Key Laboratory of Mathematical and Neural Dynamical Systems

7 SUPPLEMENTARY MATERIAL

7.1 BASELINE IMPLEMENTATIONS

We provide additional details on the baseline models used in the experiments.

Probabilistic Models for Dynamic Networks. DPGM, D-NGPPF, and G-HSEPM are designed for dynamic single-relation networks. Following standard practice, we train one independent model for each relation. For directed relations, each relational graph is converted into an undirected bipartite graph by separating entities into source and target roles, resulting in a $(2N \times 2N)$ adjacency matrix per relation.

Probabilistic Models for Dynamic Count Data. PGDS, PRGDS, and NBRGDS operate on temporal count tensors. We reshape the original $R \times T \times N \times N$ temporal knowledge graph into an event–time matrix, where each event corresponds to a triple (s, r, o) . Model inference is conducted following the settings described in the original papers.

Neural Models for Temporal Knowledge Graphs. Know-Evolve, DyRep, and RENet are neural temporal knowledge graph models that output real-valued scores for candidate events. We use the official implementations when available and follow the default hyperparameter configurations recommended by the authors. For models that do not directly produce probabilities, a sigmoid function is applied to convert scores into probabilistic predictions.

7.2 ADDITIONAL EXPERIMENTAL SETTINGS

All probabilistic generative models, including PGRE and count-based baselines, are trained using Gibbs sampling. For all datasets, the number of latent communities is fixed to $K = 200$. To ensure fair comparison, the remaining hyperparameters for each probabilistic model follow the recommendations in the original papers or official implementations. Model inference is performed with 2000 Gibbs sampling iterations, where the first 1000 iterations are discarded as burn-in and the remaining 1000 iterations are used for posterior estimation.

For neural network-based methods, we use the authors’ publicly released code and follow their recommended experimental settings. All neural models are trained in a supervised learning setting using the binary cross-entropy loss. Training is conducted for up to 1000 epochs with early stopping based on validation performance, using a patience of 20. The model checkpoint achieving the best validation performance is selected for testing. Unless otherwise specified, the learning rate and batch size are set to 0.0001 and 200, respectively.

To reduce the effect of randomness, each experiment was repeated five times with different random seeds, and the reported results correspond to average performance. All experiments are implemented in Python 3.9.12 with PyTorch 2.1.0 and

*Corresponding author: sikunyang@gbu.edu.cn

Table 3: Post-burn-in effective sample sizes for selected representative model parameters.

Parameter	ESS
π_{00}	49.97
π_{20}	336.70
π_{70}	246.59
$\tilde{\delta}_{61}^{(4,0)}$	23.63
$\tilde{\delta}_{24}^{(4,0)}$	12.99

Table 4: Sensitivity of PGRE to the temporal scale parameter τ . The results correspond to a single-run sensitivity analysis.

τ	AUC-ROC	AUC-PR
0.5	0.9324	0.2067
1.0	0.9327	0.1991
1.5	0.9329	0.2002
2.0	0.9316	0.2008

CUDA 12.1, and are conducted on a Dell Precision 7920 workstation running Ubuntu Linux (kernel 6.11.0-25-generic). The implementation code and experimental scripts are publicly available at <https://github.com/ffffkgh/PGRE>.

7.3 SAMPLING DIAGNOSTICS

We assessed the sampling behavior of the Gibbs sampler using effective sample sizes (ESSs) and trace plots for a subset of representative model parameters. The sampler was run for 2,000 iterations, with the first 1,000 iterations discarded as burn-in and the remaining 1,000 iterations retained for posterior estimation. Because the model contains a large number of latent variables, we report diagnostics for selected entries of the relation transition matrix and selected normalized temporal community weights on a representative dataset and temporal–relation slice.

For the relation transition parameters, the post-burn-in ESS values were 49.97 for π_{00} , 336.70 for π_{20} , and 246.59 for π_{70} . We additionally monitored two representative normalized temporal community weights,

$$\tilde{\delta}_k^{(t,r)} = \frac{\delta_k^{(t,r)}}{\sum_{k'=1}^K \delta_{k'}^{(t,r)}},$$

whose post-burn-in ESS values were 23.63 for $\tilde{\delta}_{61}^{(4,0)}$ and 12.99 for $\tilde{\delta}_{24}^{(4,0)}$. The corresponding results are summarized in Table 3.

Figs. 6 and 7 show the corresponding trace plots over all 2,000 iterations. The representative transition parameters exhibit an initial adaptation period followed by comparatively more stable trajectories in the retained sampling period. Their ESS values also indicate more effective mixing for π_{20} and π_{70} , while π_{00} exhibits stronger serial dependence.

The normalized temporal community-weight summaries display slower variation and stronger autocorrelation, consistent with their lower ESS values. Nevertheless, their trajectories become comparatively more stable after the initial adaptation period. Taken together, the trace plots and ESS values provide complementary evidence that the adopted sampling design captures a more stable posterior regime after burn-in. These diagnostics therefore support the use of 1,000 burn-in iterations followed by 1,000 retained iterations for posterior estimation, while also indicating that some latent community-weight summaries remain more strongly autocorrelated than the transition parameters.

7.4 SENSITIVITY ANALYSIS OF THE TEMPORAL SCALE PARAMETER

We examined the sensitivity of PGRE to the temporal scale parameter τ , which controls the scale of the Gamma–Markov evolution. We varied τ over $\{0.5, 1.0, 1.5, 2.0\}$ while keeping all other experimental settings unchanged. The results are reported in Table 4.

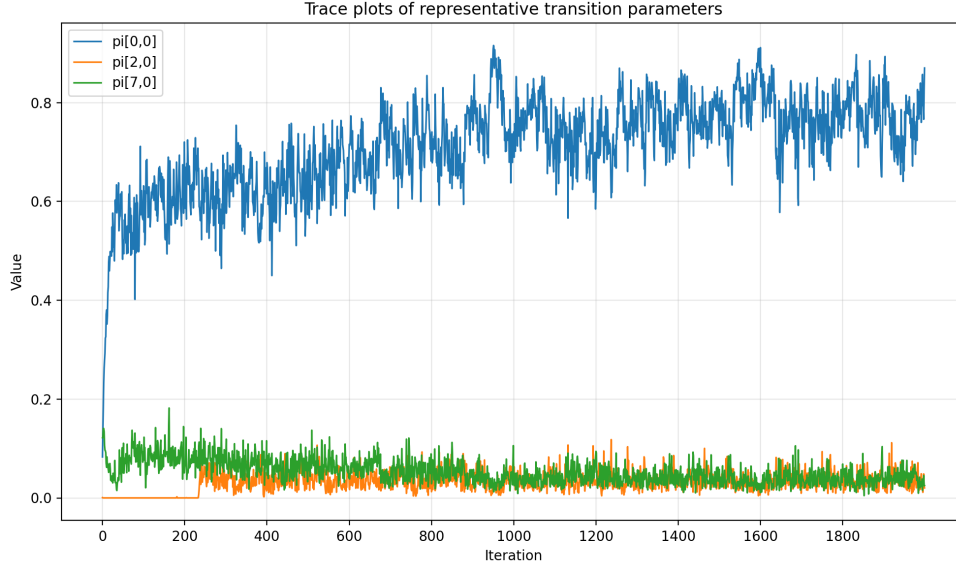


Figure 6: Trace plots of selected representative relation transition parameters over 2,000 Gibbs sampling iterations. The first 1,000 iterations are treated as burn-in, and the remaining 1,000 iterations are retained for posterior estimation. The trajectories exhibit an initial adaptation period followed by comparatively more stable post-burn-in behavior.

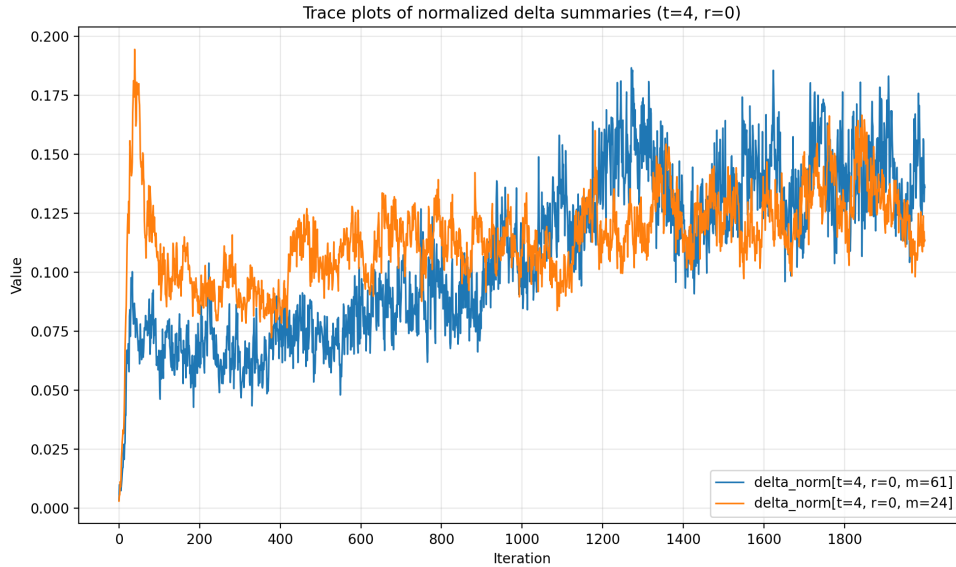


Figure 7: Trace plots of selected normalized temporal community weights at $(t, r) = (4, 0)$ over 2,000 Gibbs sampling iterations. Compared with the transition parameters, these summaries exhibit stronger serial dependence and slower variation, consistent with their lower post-burn-in ESS values.

The performance remains stable across the tested values of τ . Specifically, the AUC-ROC values vary only from 0.9316 to 0.9329, while the AUC-PR values range from 0.1991 to 0.2067. Although $\tau = 0.5$ achieves the highest AUC-PR and $\tau = 1.5$ gives the highest AUC-ROC, the differences are modest. These results indicate that PGRE is reasonably robust to moderate variations in the temporal scale parameter and does not rely on a narrowly tuned value of τ . We therefore use $\tau = 1.0$ as the default setting in the main experiments.

7.5 MARKOV CHAIN MONTE CARLO INFERENCE

Here we present the Gibbs sampling for the PGRE model.

Sampling latent counts $x_{ij}^{(t,r)}$. We sample the latent count associated with each observed event $\mathcal{M}_{ij}^{(t,r)}$ as

$$(x_{ij}^{(t,r)} | -) \sim \mathcal{M}_{ij}^{(t,r)} \text{Po}_+(\sum_{k=1}^K \delta_k^{(t,r)} \phi_{ik} \psi_{jk}). \quad (24)$$

where $\text{Pois}_+(\cdot)$ denotes a zero-truncated Poisson distribution. Since $x_{ij}^{(t,r)} = \sum_{k=1}^K x_{ijk}^{(t,r)}$, by the additive property of the Poisson distribution we have $x_{ijk}^{(t,r)} \sim \text{Pois}(\delta_k^{(t,r)} \phi_{ik} \psi_{jk})$. According to the Poisson–multinomial equivalence (Theorem 3), the latent count $x_{ijk}^{(t,r)}$ can be sampled as

$$(x_{ijk}^{(t,r)} | -) \sim \text{Mult}(x_{ij}^{(t,r)}, (\frac{\delta_k^{(t,r)} \phi_{ik} \psi_{jk}}{\sum_{k=1}^K \delta_k^{(t,r)} \phi_{ik} \psi_{jk}})). \quad (25)$$

Sampling community factor loadings ϕ_{ik} and ψ_{jk} . By the Gamma–Poisson conjugacy, the posterior distributions of the subject- and object-side memberships are given by

$$(\phi_{ik} | -) \sim \text{Gam} \left(a_0 + \sum_{t=1}^T \sum_{r=1}^R \sum_{\substack{j=1 \\ j \neq i}}^N x_{ijk}^{(t,r)}, \right. \\ \left. 1 / \left(c_i + \sum_{t=1}^T \sum_{r=1}^R \sum_{\substack{j=1 \\ j \neq i}}^N \delta_k^{(t,r)} \psi_{jk} \right) \right), \quad (26)$$

$$(\psi_{jk} | -) \sim \text{Gam} \left(a_1 + \sum_{t=1}^T \sum_{r=1}^R \sum_{\substack{i=1 \\ i \neq j}}^N x_{ijk}^{(t,r)}, \right. \\ \left. 1 / \left(c_j + \sum_{t=1}^T \sum_{r=1}^R \sum_{\substack{i=1 \\ i \neq j}}^N \delta_k^{(t,r)} \phi_{ik} \right) \right). \quad (27)$$

Sampling entity-specific scaling parameters c_i and c_j . We place Gamma priors over the entity-specific scaling parameters as $c_i \sim \text{Gam}(f_0, 1/g_0)$ and $c_j \sim \text{Gam}(z_0, 1/k_0)$. Using the Gamma–Gamma conjugacy, the posterior distributions can be derived in closed form as:

$$(c_i | -) \sim \text{Gam}(f_0 + K a_0, 1/(g_0 + \sum_{k=1}^K \phi_{ik})), \quad (28)$$

$$(c_j | -) \sim \text{Gam}(z_0 + K a_1, 1/(k_0 + \sum_{k=1}^K \psi_{jk})), \quad (29)$$

where K denotes the number of latent communities.

Sampling community weights $\delta_k^{(t,r)}$. The community weights $\delta_k^{(t,r)}$ evolve under a Markovian structure, so both backward and forward information must be incorporated into the posterior updates.

For the last time step $t = T$, we define the aggregated latent count as $x_{..k}^{(T,r)} = \sum_{i=1}^N \sum_{j \neq i}^N x_{ijk}^{(T,r)}$, which follows a Poisson distribution $x_{..k}^{(T,r)} \sim \text{Pois}(\delta_k^{(T,r)} s_k)$, where $s_k = \sum_{i=1}^N \sum_{j \neq i}^N \phi_{ik} \psi_{jk}$. By applying the gamma–Poisson conjugacy, the posterior of $\delta_k^{(T,r)}$ is given by

$$(\delta_k^{(T,r)} | -) \sim \text{Gam} \left(x_{..k}^{(T,r)} + \sum_{r'=1}^R \pi_{rr'} \delta_k^{(T-1,r')}, 1/(\tau + s_k) \right). \quad (30)$$

For $t = T - 1$, to incorporate forward information, we marginalize out $\delta_k^{(T,r)}$ using Theorem 1, which leads to

$$x_{..k}^{(T,r)} \sim \text{NB} \left(\sum_{r'=1}^R \pi_{rr'} \delta_k^{(T-1,r')}, s_k/(\tau + s_k) \right). \quad (31)$$

According to Theorem 2, the negative binomial distribution can be augmented with an auxiliary variable as

$$l_k^{(T,r)} \sim \text{CRT} \left(x_{..k}^{(T,r)}, \sum_{r'=1}^R \pi_{rr'} \delta_k^{(T-1,r')} \right). \quad (32)$$

The joint distribution of $x_{..k}^{(T,r)}$ and $l_k^{(T,r)}$ can then be written as

$$x_{..k}^{(T,r)} \sim \text{SumLog}(l_k^{(T,r)}, \rho_k^{(T,r)}), \quad (33)$$

$$l_k^{(T,r)} \sim \text{Pois} \left(- \sum_{r'=1}^R \pi_{rr'} \delta_k^{(T-1,r')} \ln(1 - \rho_k^{(T)}) \right), \quad (34)$$

where $\rho_k^{(T)} = \frac{s_k}{\tau + s_k}$. Since $l_k^{(T,r)} = l_k^{(T,r \cdot)} = \sum_{r_2=1}^R l_k^{(T,rr_2)}$, the distribution of $l_k^{(T,rr')}$ is

$$l_k^{(T,rr')} \sim \text{Pois} \left(- \pi_{rr'} \delta_k^{(T-1,r')} \ln(1 - \rho_k^{(T)}) \right). \quad (35)$$

Using the additive property of the Poisson distribution and $\sum_{r_1=1}^R \pi_{rr_1} = 1$, we have

$$l_k^{(T,r \cdot)} \sim \text{Pois} \left(- \delta_k^{(T-1,r)} \ln(1 - \rho_k^{(T)}) \right). \quad (36)$$

Given $x_{..k}^{(T-1,r)} \sim \text{Pois}(\delta_k^{(T-1,r)} s_k)$, the Poisson additive property leads to

$$x_{..k}^{(T-1,r)} + l_k^{(T,r \cdot)} \sim \text{Pois} \left(\delta_k^{(T-1,r)} (s_k - \ln(1 - \rho_k^{(T)})) \right). \quad (37)$$

This combines backward and forward information at $t = T - 1$. With the gamma prior on $\delta_k^{(T-1,r)}$, its conditional posterior is

$$(\delta_k^{(T-1,r)} | -) \sim \text{Gam} \left(x_{..k}^{(T-1,r)} + l_k^{(T,r \cdot)} + \sum_{r'}^R \pi_{rr'} \delta_k^{(T-2,r')}, 1/(\tau + s_k - \ln(1 - \rho_k^{(T)})) \right). \quad (38)$$

For intermediate steps $t = T - 2, \dots, 2$, we introduce

$$l_k^{(t,r)} \sim \text{CRT} \left(x_{\cdot k}^{(t,r)} + l_k^{(t+1,r)}, \sum_{r'=1}^R \pi_{rr'} \delta_k^{(t-1,r')} \right), \quad (39)$$

and apply the same augmentation–sampling procedure:

$$\begin{aligned} (\delta_k^{(t,r)} | -) &\sim \text{Gam} \left(x_{\cdot k}^{(t,r)} + l_k^{(t+1,r)} + \sum_{r'}^R \pi_{rr'} \delta_k^{(t-1,r')}, \right. \\ &\quad \left. 1/(\tau + s_k - \ln(1 - \rho_k^{(t+1)})) \right), \end{aligned} \quad (40)$$

where

$$\rho_k^{(t+1)} = \frac{s_k - \ln(1 - \rho_k^{(t+2)})}{\tau + s_k - \ln(1 - \rho_k^{(t+2)})}. \quad (41)$$

Finally, for $t = 1$, we augment

$$l_k^{(1,r)} \sim \text{CRT} \left(x_{\cdot k}^{(1,r)} + l_k^{(2,r)}, \nu_r/K \right), \quad (42)$$

and sample

$$\begin{aligned} (\delta_k^{(1,r)} | -) &\sim \text{Gam} \left(x_{\cdot k}^{(1,r)} + l_k^{(2,r)} + \nu_r/K, \right. \\ &\quad \left. 1/(\tau + s_k - \ln(1 - \rho_k^{(2)})) \right). \end{aligned} \quad (43)$$

Sampling the transition matrix π_r . After marginalizing out δ , the auxiliary counts follow a multinomial distribution $(l_k^{(t,1r)}, \dots, l_k^{(t,Rr)}) \sim \text{Mult} \left(l_k^{(t,\cdot r)}, (\pi_{1r}, \dots, \pi_{Rr}) \right)$. By the Dirichlet–multinomial conjugacy, the posterior distribution of π_r is given by

$$\begin{aligned} (\pi_r | -) &\sim \text{Dir} \left(\nu_1 \nu_r + \sum_{t=2}^T \sum_{k=1}^K l_k^{(t,1r)}, \dots, \right. \\ &\quad \xi \nu_r + \sum_{t=2}^T \sum_{k=1}^K l_k^{(t,r r)}, \dots, \\ &\quad \left. \nu_R \nu_r + \sum_{t=2}^T \sum_{k=1}^K l_k^{(t,Rr)} \right). \end{aligned} \quad (44)$$

Sampling ν_r and ξ . We marginalize over Π to obtain a Dirichlet–multinomial distribution

$$\begin{aligned} (l_{\cdot}^{(\cdot,1r)}, \dots, l_{\cdot}^{(\cdot,Rr)}) \\ \sim \text{DirMult} \left(l_{\cdot}^{(\cdot,\cdot r)}, (\nu_1 \nu_r, \dots, \xi \nu_r, \dots, \nu_R \nu_r) \right), \end{aligned} \quad (45)$$

where $l_{\cdot}^{(\cdot,\cdot r)} = \sum_{t=1}^T \sum_{k=1}^K \sum_{r_1=1}^K l_k^{(t,r_1 r)}$.

By introducing a beta-distributed auxiliary variable q_r , the Dirichlet–multinomial distribution can be rewritten as a negative binomial distribution:

$$q_r \sim \text{Beta} \left(l_{\cdot}^{(\cdot,\cdot r)}, \nu_r \left(\xi + \sum_{r' \neq r} \nu_{r'} \right) \right). \quad (46)$$

Then, $l^{(\cdot, rr)}$ and $l^{(\cdot, rr')}$ follow negative binomial distributions as $l^{(\cdot, rr)} \sim \text{NB}(\xi\nu_r, q_r)$ and $l^{(\cdot, rr')} \sim \text{NB}(\nu_r\nu_{r'}, q_r)$. To further facilitate posterior inference, we introduce auxiliary count variables h_{rr} and $h_{rr'}$ via the Chinese Restaurant Table (CRT) distribution:

$$\begin{aligned} h_{rr} &\sim \text{CRT}\left(l^{(\cdot, rr)}, \xi\nu_r\right), \\ h_{rr'} &\sim \text{CRT}\left(l^{(\cdot, rr')}, \nu_r\nu_{r'}\right). \end{aligned} \quad (47)$$

Using the gamma–Poisson conjugacy, ξ is then sampled as

$$(\xi|-) \sim \text{Gam}\left(b_0 + \sum_{r=1}^R h_{rr}, \frac{1}{e_0 - \sum_{r=1}^R \nu_r \ln(1 - q_r)}\right). \quad (48)$$

Next, we define

$$n_r = h_{rr} + \sum_{r_1 \neq r} h_{r_1 r} + \sum_{r_2 \neq r} h_{r r_2} + \sum_{k=1}^K l_k^{(1, r \cdot)},$$

where $l_k^{(1, r \cdot)} \sim \text{Pois}(\tau\nu_r \ln(1 - \rho_k^{(1)}))$. Using the Poisson additive property and gamma–Poisson conjugacy, we have

$$(\nu_r|-) \sim \text{Gam}\left(\frac{\gamma_0}{R} + n_r, 1/(\beta + t_r)\right), \quad (49)$$

where $t_r = -\ln(1 - q_r) \left(\xi + \sum_{r_1 \neq r} \nu_{r_1} \right) - \sum_{r_2 \neq r} \ln(1 - q_{r_2}) \nu_{r_2} - \sum_{k=1}^K \frac{1}{K} \ln(1 - \rho_k^{(1)})$.

Sampling β . Finally, by gamma–gamma conjugacy, the posterior of β is derived as

$$(\beta|-) \sim \text{Gam}(d_0 + \gamma_0, 1/(h_0 + \sum_{r=1}^R \nu_r)). \quad (50)$$

Algorithm 1 summarises the full sampling procedure of PGRE.

7.6 FULL PROBABILISTIC GRAPHICAL MODEL

Fig. 8 presents the full probabilistic graphical model of PGRE, which illustrates the complete generative process including entity-level latent factors, relation-level temporal evolution, and observed interactions.

Algorithm 1 Gibbs sampling algorithm for PGRE

Require: Dynamic relational data $G^{(1)}, \dots, G^{(T)}$.

```
1: Initialise the number of communities  $K$  and other parameters;
2: repeat
3:   Sample  $x_{ij}^{(t,r)}$  (Eq. 24) and update  $x_{ijk}^{(t,r)}$  (Eq. 25)
4:   Sample  $\phi_{ik}$  (Eq. 26) and  $\psi_{jk}$  (Eq. 27)
5:   update  $c_i$  (Eq. 28) and  $c_j$  (Eq. 29)
6:   for  $t = T, \dots, 1$  do
7:     if  $t = T$  then
8:       Sample  $l_k^{(T,r)}$  (Eq. 32)
9:     else if  $t = T - 1, \dots, 2$  then
10:      Sample  $l_k^{(t,r)}$  (Eq. 39)
11:    else if  $t = 1$  then
12:      Sample  $l_k^{(1,r)}$  (Eq. 42)
13:    end if
14:    Update  $\rho_k^{(t)}$  (Eq. 41)
15:  end for
16:  for  $t = 1, \dots, T$  do
17:    if  $t = 1$  then
18:      Sample  $\delta_k^{(1,r)}$  (Eq. 43)
19:    else if  $t = 2, \dots, T - 1$  then
20:      Sample  $\delta_k^{(t,r)}$  (Eq. 40)
21:    else if  $t = T$  then
22:      Sample  $\delta_k^{(T,r)}$  (Eq. 30)
23:    end if
24:  end for
25:  Sample  $\pi_r$  (Eq. 44), update  $q_r$  (Eq. 46) and  $h_{rr}$  (Eq. 47)
26:  Sample  $\xi$  (Eq. 48) and  $\nu_r$  (Eq. 49)
27:  Sample  $\beta$  (Eq. 50)
28: until convergence
Ensure: Posterior mean of  $\phi_{ik}$ ,  $\psi_{jk}$  and  $\delta_k^{(t,r)}$ 
```

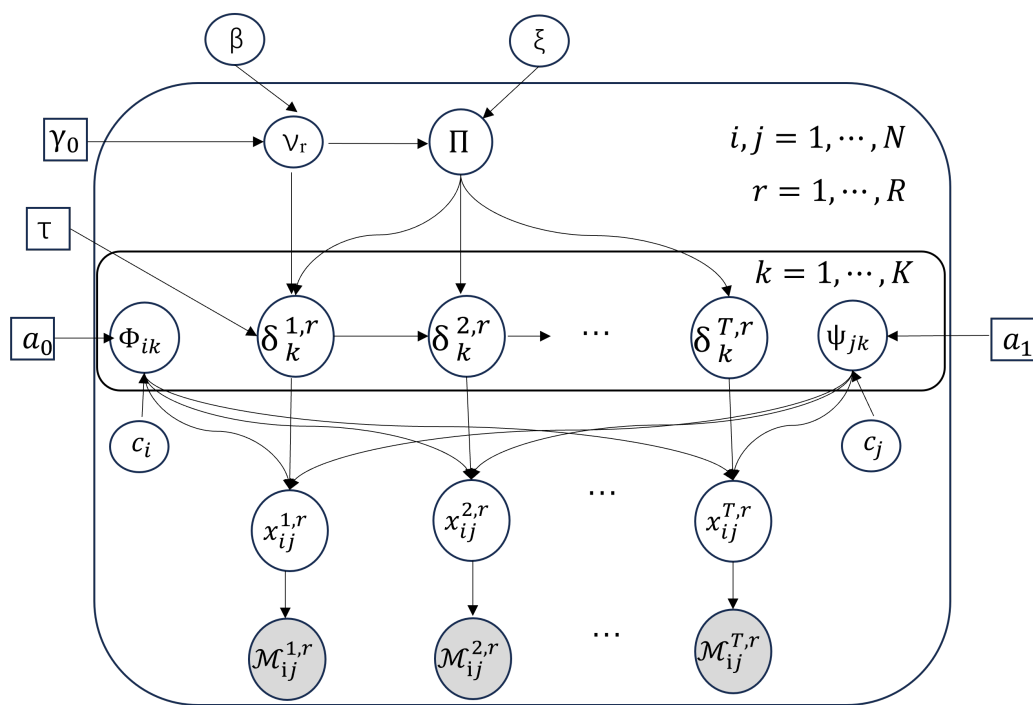


Figure 8: Full probabilistic graphical model of PGRE.