Community-Aware Variational Autoencoder for Continuous Dynamic Networks

Junwei Cheng[1], Chaobo He[1*], Pengxing Feng[2], Weixiong Liu[1], Kunlin Han[3,4], Yong Tang[1*]

[1]School of Computer Science, South China Normal University
[2]Department of Electrical Engineering, City University of Hong Kong
[3]CMT US Holdings LLC
[4]Computer Science Department, University of Southern California
{jung, hechaobo, weixiongliu, ytang}@m.scnu.edu.cn,
pengxfeng2-c@my.cityu.edu.hk, kunlinha@usc.edu

Abstract

Variational autoencoder performs well in community detection on static networks, but it is difficult to directly extend to continuous dynamic networks. The main reason is that traditional methods mainly rely on adjacency structures to complete the inference and generation processes. However, continuous dynamic networks cannot be described by this structure because the inherent timeliness and causality information of the network would be lost. To address this issue, we propose a novel variational autoencoder, CT-VAE, for community detection in continuous dynamic networks, along with its scalable variant, CT-CAVAE. By conceptualizing node interactions as event streams and adopting the Hawkes process to capture temporal dynamics and causality, and incorporating them into the inference process, CT-VAE can effectively extend the traditional inference approach to continuous dynamic networks. Additionally, in the generation phase, CT-VAE combines pseudo-labeling and compact constraint strategies to facilitate the reconstruction process of non-adjacent structures. For the scalable variant, CT-CAVAE, end-to-end community detection is achieved by cleverly combining Gaussian mixture distribution. Extensive experimental results demonstrate that the proposed CT-VAE and CT-CAVAE achieve more favorable performance compared with the state-of-the-art baselines.

Introduction

Variational autoencoder is a crucial methodology within the generative paradigm, primarily comprising three components: inference, generation and the evidence lower bound (ELBO). The inference component is responsible for learning latent representations from data, while the generation component reconstructs new data based on these learned representations. ELBO functions as an optimization objective, ensuring that the learned distribution of representations closely aligns with the actual yet unknown distribution of data. Given that variational autoencoder provides a natural optimization objective and has advantages in handling complex data (Blei, Kucukelbir, and McAuliffe 2017), its application to graph machine learning (Liang et al. 2023; Bhattacharya, Liu, and Maiti 2024; Liang et al. 2024), and in particular to the community detection, seems inherently appropriate.

Community detection, aka. graph clustering, aims to divide nodes into optimal cluster structures with dense intra-edge and sparse inter-edge connections. For nodes with attributes, such as those in text-attributed networks (Cheng et al. 2024), nodes within the same community tend to have similar attribute information, and vice versa. To uncover the community structure in static networks, the inference component of variational autoencoder methods typically learns the latent representations of nodes through the message-passing mechanism of graph convolutional networks (GCN), while the generation component focuses on reconstructing the network structure (Kipf and Welling 2016). Subsequent methods build upon this approach and attempt to introduce improvements to various components (Mrabah, Bouguessa, and Ksantini 2024; Li, Li, and Lian 2023; Choong, Liu, and Murata 2018).

While these methods have demonstrated promising performance, they predominantly focus on the static networks and do not explore the continuous dynamic networks. Continuous dynamic networks primarily differ from static networks in the timeliness and causality of edges between nodes (Khoshraftar and An 2024; Su et al. 2024). As illustrated in Figure 1(A), Tom interacts with Jim, Mia and Ella at different times, with a lower frequency of interaction with Ella. In the context of continuous dynamic networks, based on the graph motif (Paranjape, Benson, and Leskovec 2017), Jim and Mia are likely to interact in the future due to their common and close friend, Tom. Conversely, Jim and Ella are unlikely to interact, as Ella appears to be more of a transient visitor. However, in static networks, the absence of edge timeliness information promotes the formation of a connection between Jim and Ella.

Although continuous dynamic networks provide a more realistic representation of scenarios, research on community detection within such networks remains limited. One challenge is that existing relevant datasets often lack node labels, making it difficult to quantify the performance of proposed methods. Additionally, current variational autoencoder-based methods are difficult to extend directly to continuous patterns. Specifically,

- Non-adjacent structures complicate the inference process

*Corresponding authors.

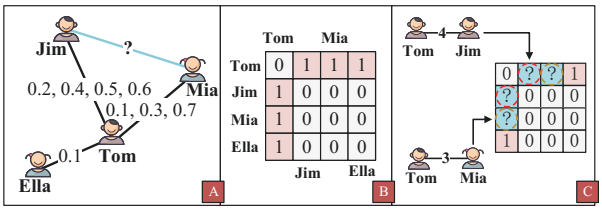

Figure 1: A toy network with user interactions and its representation across different network patterns.

in learning latent representations of nodes. Static networks, such as text-attributed networks, multiplex networks (Cheng et al. 2023) and discrete dynamic networks (Li et al. 2023), can be represented by adjacency matrices. As shown in Figure 1(B), an element is set to 1 if there is an interaction between two nodes, and 0 otherwise. However, the timeliness and causality inherent in continuous dynamic networks are challenging to capture using adjacency matrices. As illustrated in Figure 1(A and C), an adjacency matrix can only represent the interaction between Jim and Tom, but it fails to reflect their interactions across four different time periods. Additionally, as the saying goes, *"A butterfly flapping its wings in the Amazon rainforest can cause a tornado in Texas,"* historical events can influence future outcomes—a characteristic unique to continuous dynamic networks that adjacency matrices cannot adequately represent. Against this backdrop, existing approaches of using GCN for the inference process becomes ineffective, as they cannot perform message passing on non-adjacent structures.

- Generation component is unable to directly reconstruct non-adjacent structures. During the generation phase, existing methods (Li, Li, and Lian 2023; Salha-Galvan et al. 2022) typically assume that the connection probability between two nodes follows a Bernoulli distribution, and reconstruct the original network structure. However, as discussed above, the unique data structure of continuous dynamic networks makes it challenging to directly apply traditional approaches to the generation process.

Notably, the availability of labeled datasets summarized and proposed by (Liu et al. 2024b) has significantly advanced research on community detection in continuous dynamic networks. Based on these datasets, in this paper, we propose a novel method called CT-VAE to address the aforementioned problems. Specifically, we conceptualize the process of nodes interacting at different times in a continuous dynamic network as event streams, represented in the form of triples rather than the adjacency matrix. To better capture the complex relationships in non-adjacent structures, we approach the first problem by combining variational autoencoder with the Hawkes process (Zuo et al. 2018). The underlying mechanism is that the self-exciting nature of the Hawkes process captures node features and causal relationships between nodes from both current and historical events. By incorporating the learned features into the inference process, ELBO can naturally impose goal-oriented constraints. In accordance with the principles of community homogeneity, nodes within the same community should demonstrate close associations and similar attributes. Inspired by this, we introduce pseudo labels to determine the community affiliation of nodes in order to address another problem. For nodes that belong to the same community and share similar attributes, we construct edges between them. Conversely, we assume the absence of edges for nodes that do not meet these criteria.

Since the optimization objective of CT-VAE is not oriented towards community detection, K-means is required for identifying communities. To overcome this limitation, we propose a scalable variant of CT-VAE, named CT-CAVAE, which enables end-to-end community detection. Specifically, inspired by (Choong, Liu, and Murata 2018), we impose a community constraint on the learned node representations that follows a Gaussian mixture distribution. For nodes within the same community, their representations should exhibit similar probability distributions under the same Gaussian component. Furthermore, we update the ELBO to ensure that the introduced constraints effectively enhance the ability of node representations to reflect community characteristics. Figure 2 outlines the overall process of our proposed methods, CT-VAE and CT-CAVAE. The optimization objective unique to CT-CAVAE is highlighted within a red dashed box.

The contributions of this paper are:

- We have implemented an effective extension of variational autoencoder community detection for continuous dynamic networks, named CT-VAE, which addresses the challenges in designing the generation and inference components caused by the non-adjacent structure of these networks.
- We propose a scalable version, CT-CAVAE, which not only achieves end-to-end community detection by introducing a Gaussian mixture distribution but also enhances the performance of CT-VAE through community alignment.
- Extensive experiments on six real-world datasets verify that the proposed CT-VAE outperforms state-of-the-art baselines and demonstrate the promising performance of CT-CAVAE in specific data scenarios.

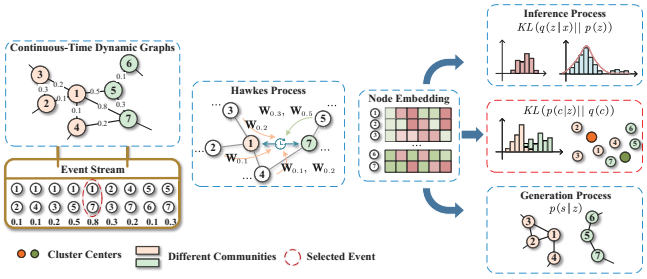

Figure 2: Framework of the proposed CT-VAE and CT-CAVAE.

Methodology

Problem Formulation

A continuous dynamic network can be denoted as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{T}\}$, where $\mathcal{V}$ and $\mathcal{E}$ are sets of nodes and edges, respectively. The interactions between nodes u and v in different time periods can be stored as triples $\mathcal{T}_{u,v} := \{(u, v, t_1), (u, v, t_2), \cdots, (u, v, t_T)\}$, where T denotes the maximum time step. Given the current time t, the neighbors nodes of u (i.e., the events that have occurred) are described by the pair $N_t(u) := \{(v, t_1), (s, t_2), \cdots, (v, t)\}$.

Given a network $\mathcal{G}$, the objective of community detection is to partition the nodes into k disjoint communities $\mathcal{C} = \{c_1, c_2, \cdots, c_k\}$ in an unsupervised paradigm.

Overview

In general, we aim for CT-VAE and CT-CAVAE to be simple yet effective methods for continuous dynamic networks by adhering to the core principles of variational autoencoder methods used in static networks. However, due to the non-adjacent structures of continuous dynamic networks, we need to emphasize two key challenges that need to be solved:

- Typical approach of using GCN to encode the posterior distribution $q(z \mid x)$ is not applicable due to the non-adjacent property of continuous dynamic networks. Nevertheless, we introduce the Hawkes process to address this gap, the learned representation cannot be modeled by a Gaussian distribution. Therefore, the optimization objective of ELBO needs to shift from the parameters of a Gaussian distribution to directly optimizing the node representation z with an unknown distribution.
- The generation component aims to reconstruct the network structure via learned representations. Although (Liu et al. 2024b) suggested building edges between nodes with similar representations, this approach may introduce ambiguity in community structure due to its relaxed homogeneity-based constraint. For instance, the spread of fraudulent information can create two-hop associations between otherwise unrelated nodes. Under this relaxed constraint, such nodes might form connections, particularly at community boundaries. Hence, it is essential to introduce a compact constraint to ensure that the reconstructed structure reasonably reflects the inherent community characteristics.

Subsequently, we will introduce strategies in detail to address these two key challenges.

Inference Component of CT-VAE and CT-CAVAE

As discussed earlier, past events can positively or negatively influence the occurrence of future events. Capturing the timeliness and causality between nodes is particularly crucial in continuous dynamic networks. The Hawkes process, a classic self-exciting method, is well-suited for this scenario. Given two nodes u and v that interact at time t, their conditional intensity function is described as follows:

$$\begin{aligned}\lambda_{u,v}(t) = \mu_{u,v} &+ \sum_{v',t' \in N_t(v)} \alpha_{u,v'}(t)\mu_{u,v'}\kappa(t-t') \\ &+ \sum_{u',t' \in N_t(u)} \alpha_{v,u'}(t)\mu_{v,u'}\kappa(t-t')\end{aligned} \tag{1}$$

where the base intensity $\mu_{u,v}$ is typically measured by the similarity or distance between node representations. In this context, we define it as $\mu_{u,v} = -\|z_u - z_v\|^2$. $\kappa(t-t') = \exp(-\delta(t-t'))$ suggests that the larger the interval between the historical time and the current time, the smaller the impact on the current node, where δ is a learnable parameter. $\alpha(t)$ is the normalized causal weight, which quantifies the influence of historical interactions on the current node, as follows:

$$\alpha_{u,v'}(t) = \frac{exp(\mu_{u,v'})}{\sum_{v \in N_t(u)} \exp(\mu_{u,v})} \tag{2}$$

From Eq. (1), it is evident that the Hawkes process infers the interaction probability between two nodes at time t through the conditional density function. For a pair of nodes at time t, a higher conditional density function indicates greater similarity in their representations. Intuitively, this is analogous to the message propagation mechanism of GCN. For connected nodes, propagating and aggregating neighboring information through one or two hops results in these nodes having similar representations. Therefore, in the inference component of CT-VAE, we learn the latent representation of nodes by optimizing the log-likelihood function of the conditional density function for interactions between all nodes, as follows:

$$\begin{aligned}q(\mathbf{Z}|\mathcal{E}, \mathcal{T}) &:= \sum_{u \in \mathcal{V}} \log q(\mathbf{Z}|u, N_u(t)) \\ &= \sum_{u \in \mathcal{V}} \sum_{v \in N_t(u)} [\log \phi(\lambda_{u,v}(t)) - \sum \mathbb{E}_{v' \in P(u)} \log \phi(\lambda_{u,v'}(t))]\end{aligned} \tag{3}$$

where we adopt the negative sampling technique in (Zuo et al. 2018) to mitigate the significant computational overhead caused by the summation in Eq. (3). $P(u) \propto |N(u)|$ is the negative sampling distribution, $\mathbf{Z} \in \mathbb{R}^{|\mathcal{V}| \times d}$ is the node representation matrix, and $\phi(\cdot)$ is the sigmoid function.

Generation Component of CT-VAE and CT-CAVAE

The generation component ensures that the learned node representations capture the overall characteristics of the input data as accurately as possible. Since we treat the input of continuous dynamic networks as event streams rather than an adjacency matrix, traditional reconstruction approaches are challenging to apply directly. Recently, (Liu et al. 2024b) proposed a reconstruction approach suitable for this scenario. Specifically, given an interaction triplet (u, v, t), neighbor nodes $h \in N_u(t)$ and negative sampling nodes $v' \in P(u)$, the reconstruction objective is described as follows:

$$\begin{aligned}\min\{|1 - \cos(\mathbf{Z}_u, \mathbf{Z}_v)| + |1 - \cos(\mathbf{Z}_u, \mathbf{Z}_h)| \\ + |0 - \cos(\mathbf{Z}_u, \mathbf{Z}'_v)|\}\end{aligned} \tag{4}$$

where $\cos(\mathbf{Z}_u, \mathbf{Z}_v) = \frac{\mathbf{Z}_u \mathbf{Z}_v^T}{\|\mathbf{Z}_u\| \cdot \|\mathbf{Z}_v\|}$ is cosine similarity, which indicates the probability that node u and v have an edge. Intuitively, minimizing Eq. (4) implies a high probability of edges existing between nodes with similar representations, while there is a high probability that edges do not exist between unrelated nodes (negative sampling nodes). However, this approach, which relies on relaxed homogeneity constraints, is likely to introduce ambiguity into the community structure. This issue is particularly pronounced at nodes located on the boundaries of communities, where the resulting overly close connections between communities can significantly reduce the accuracy of community detection. Therefore, we propose a compact constraint based on Eq. (4) to ensure that the reconstruction network effectively reflects the inherent characteristics of the community, as follows:

$$\begin{aligned} \min\{|\mathbf{X}_{u,v} - \cos(\mathbf{Z}_u, \mathbf{Z}_v)| + |\mathbf{X}_{u,h} - \cos(\mathbf{Z}_u, \mathbf{Z}_h)| \\ + |0 - \cos(\mathbf{Z}_u, \mathbf{Z}'_v)|\} \end{aligned} \tag{5}$$

where $\mathbf{X}$ is the pseudo-label matrix. Assume that the community membership vectors of nodes u and v at a given epoch are $c(u)$ and $c(v)$, respectively. Determining if these two nodes belong to the same community is equivalent to checking whether the product of their community membership vectors is zero. Therefore, $\mathbf{X}_{u,v} = 1$ if and only if $c(u)c(v)^T = 1$, 0 otherwise.

It is worth noting that our approach is motivated by the characteristic of community homogeneity. That is, nodes within the same community not only exhibit dense connections but also share similar representations. For nodes u and v, an edge is established if and only if they belong to the same community and have similar node representations. Its advantage lies in avoiding the need to reconstruct all interaction events, as the complexity and diversity of the network mean that not all edges need to be considered. For nodes with infrequent interactions that are connected by hub nodes, this approach minimizes the negative impact of hub nodes on community boundaries.

ELBO of CT-VAE

Based on the Eq. (3) and Eq. (5), the ELBO of CT-VAE can be expressed as:

$$\begin{aligned} \mathcal{L} = - \sum_{\substack{u,v,t \in \mathcal{T} \\ h \in N_t(u) \\ v' \in P(u)}} \{|\mathbf{X}_{u,v} - \cos(\mathbf{Z}_u, \mathbf{Z}_v)| + |\mathbf{X}_{u,h} - \cos(\mathbf{Z}_u, \mathbf{Z}_h)| \\ + |0 - \cos(\mathbf{Z}_u, \mathbf{Z}'_v)|\} - KL(q(\mathbf{Z}|\mathcal{E}, \mathcal{T}) \| p(\mathbf{Z})) \end{aligned} \tag{6}$$

It is important to note that the approach of calculating the KL divergence by directly introducing $p(z) \sim \mathcal{N}(0, 1)$ (Kipf and Welling 2016) in the aforementioned equation is no longer applicable. The primary reason is that the posterior distribution $q(\cdot)$ cannot be modeled by a Gaussian distribution. Thus, the optimization objective of EBLO need shift from optimizing the parameters of the posterior distribution to directly optimizing the node representations.

To address this issue, we introduce the Student's t-distribution (van der Maaten and Hinton 2008) as the prior distribution. The prior distribution of node u in the k-th community can be formulated as:

$$p(\mathbf{Z}_u) = \frac{(1 + \|\mathbf{Z}_u - \theta_k\|^2/\varepsilon)^{-\frac{\varepsilon+1}{2}}}{\sum_{c_i \in \mathcal{C}} (1 + \|\mathbf{Z}_u - \theta_{c_i}\|^2/\varepsilon)^{-\frac{\varepsilon+1}{2}}} \tag{7}$$

where θ_k is the cluster center of k-th community initialized by K-means, and ε represents the degrees of freedom, with a default value of 1. Our goal is to ensure that the posterior distribution remains as close as possible to the center of the Student's t-distribution via KL divergence, thereby aiding in the determination of the community structure. Additionally, using the Student's t-distribution as a prior distribution mitigates the impact of outliers on node representations, as it is less sensitive to outliers compared to the Gaussian distribution.

The Scalable Variant of CT-VAE

The proposed CT-VAE can be regarded as a method similar to VGAE applied to continuous dynamic networks. Additionally, CT-VAE can also achieve end-to-end community detection by incorporating the Gaussian mixture distribution (GMM), which results in a scalable variant known as CT-CAVAE. Specifically, while keeping the generation component unchanged, we impose a community-oriented objective on the node representation $\mathbf{Z}$ during the inference process of CT-VAE, as follows:

$$q(\mathbf{Z}, \mathbf{C}|\mathcal{E}, \mathcal{T}) = q(\mathbf{Z}|\mathcal{E}, \mathcal{T}) p(\mathbf{C}|\mathbf{Z}) \tag{8}$$

where the distribution $p(\mathbf{C}, |\mathbf{Z})$ is defined as follows (Choong, Liu, and Murata 2018):

$$p(\mathbf{C}|\mathbf{Z}) = \frac{p(\mathbf{Z}|\mathbf{C}) p(\mathbf{C})}{\sum_k p(\mathbf{Z}|\mathbf{C}_k) p(\mathbf{C}_k)} \tag{9}$$

where $p(\mathbf{Z}|\mathbf{C}) \sim \mathcal{N}(\mathbf{Z}|\mu_c, \sigma_c^2)$, and $p(\mathbf{C}) \sim Cat(\pi)$ is categorical distribution with prior probability π, where $\sum_k \pi_k = 1$. Subsequently, Eq.(6) is updated as follows:

$$\begin{aligned} \mathcal{L} &= \mathbb{E}_{(\mathbf{Z},\mathbf{C}) \sim q(\mathbf{Z},\mathbf{C}|\mathcal{E},\mathcal{T})} \log \frac{p(\mathbf{Z}, \mathbf{C}, \mathcal{T})}{q(\mathbf{Z}, \mathbf{C} \mid \mathcal{E}, \mathcal{T})} \\ &= - \sum_{\substack{u,v,t \in \mathcal{T} \\ h \in N_t(u) \\ v' \in P(u)}} \{|\mathbf{X}_{u,v} - \cos(\mathbf{Z}_u, \mathbf{Z}_v)| + |\mathbf{X}_{u,h} - \cos(\mathbf{Z}_u, \mathbf{Z}_h)| \\ &+ |0 - \cos(\mathbf{Z}_u, \mathbf{Z}'_v)|\} - KL(q(\mathbf{Z}|\mathcal{E}, \mathcal{T}) \| p(\mathbf{Z})) \\ &- KL(q(\mathbf{C} \mid \mathbf{Z}) \| p(\mathbf{C} \mid \mathbf{Z}))) \end{aligned} \tag{10}$$

Intuitively, the final term of the KL divergence in Eq.(10) can be understood as a constraint that aligns the communities identified by the downstream manner with those identified by the end-to-end manner as closely as possible. Given the latter distribution, $q(\mathbf{C} \mid \mathbf{Z})$ is then used to describe the communities identified by downstream manner, which is defined as follows:

$$q(\mathbf{C} \mid \mathbf{Z}) = \frac{(1 + \|\mathbf{Z} - \mathbf{W}\|^2)^{-1}}{\sum_{c_i \in \mathcal{C}} (1 + \|\mathbf{Z} - \mathbf{W}_{c_i}\|^2)^{-1}} \tag{11}$$

where $\mathbf{W}$ is the learnable cluster center, initialized by K-means. By maximizing Eq. (10), the $KL(q(\mathbf{C} \mid \mathbf{Z}) \| p(\mathbf{C} \mid \mathbf{Z}))$ will gradually approach zero, ensuring that the community structures obtained through the two different approaches are as consistent as possible.

Experiments

Experimental Settings

Datasets and baselines: To comprehensively evaluate our methods, several continuous dynamic networks from various real-world domains are used. Specifically, these include a co-author network (DBLP), two citation networks (arXivCS and arXivAI), a brain network (Brain), a patent citation network (Patent), and a high school student interaction network (School). Table in the Appendix summarizes the statistics of these datasets, where the respective node homogeneity ratios $\mathcal{H}_n$ and edge homogeneity ratios $\mathcal{H}_e$ are calculated using (Pei et al. 2020) and (Zhu et al. 2020), respectively. For a detailed introduction to these datasets, refer to (Liu et al. 2024b).

Several state-of-the-art baselines have been selected for comparison, including the highly influential method node2vec (Grover and Leskovec 2016). In addition, eight representative continuous dynamic network learning methods have been considered: HTNE (Zuo et al. 2018), JODIE (Kumar, Zhang, and Leskovec 2019), TGAT (Xu et al. 2020), TGN (Rossi et al. 2020), MTNE (Huang et al. 2020), M^2DNE (Lu et al. 2019), TREND (Wen and Fang 2022) and TGC (Liu et al. 2024b), with TGC being the only method specifically designed for community detection.

Experimental setup and metrics: For CT-VAE and CT-CAVAE, the Adam is employed for training, with the learning rate selected from $\{1e^{-2}, 1e^{-3}, 2e^{-4}, 5e^{-5}\}$. Additionally, we set the negative sampling size to 2 and the history window to 1 in the Hawkes process. To ensure the fairness of our experiments, we initialize the representation $\mathbf{Z}$ using node2vec for all methods that require initialization. Furthermore, our proposed methods are trained for 100 epochs with a batch size of 1024. Other baseline methods adopt the settings described in their respective papers or undergo fine-tuning. For baseline methods that do not require initialization of representations, the data is referenced from (Liu et al. 2024b). All trials have been conducted on Intel Core i7-6700 CPUs and NVIDIA RTX 3090 GPUs. We employ four widely used metrics to measure performance of all methods: Accuracy (ACC), Normalized Mutual Information (NMI), Adjusted Rand Index(ARI) and F1-macro (F1).

Performance Comparison and Discussion

On each dataset, we conduct a comprehensive comparison of CT-VAE and CT-CAVAE with baseline methods, and the results are summarized in Table 1.

CT-VAE clearly outperforms most baseline methods across various datasets. Specifically, CT-VAE demonstrates significant performance advantages on the School, arXivAI and arXivCS. On the Patent dataset, CT-VAE achieves the best results in NMI, ARI and F1 scores, with improvements of 5.41%, 4.6% and 0.76% over the runner-up, respectively. Additionally, CT-VAE shows substantial average improvements in ACC, NMI, ARI and F1, with gains of 4.82%, 14.68%, 14.69% and 8.48%, respectively. On the DBLP and Brain datasets, CT-VAE secures either the best or runner-up positions in most evaluation metrics. For instance, on DBLP, it achieves the best results in ACC and NMI, and runner-up in ARI and F1, with ARI and F1 scores only 0.14% and 0.13% lower than the best. On the Brain, CT-VAE achieves the best results in NMI and ARI, while ACC and F1 are only marginally lower than the best scores, particularly in ACC.

Discussion. Both node homogeneity and edge homogeneity measure the probability that nodes within the same community interact. From the perspective of homogeneity, it can be observed that CT-VAE demonstrates significant average performance improvements on highly homogeneous networks such as arXivAI and School. Simultaneously, it also shows notable performance enhancements on the Brain dataset, which has lower homogeneity. We attribute this advantage to the compactness constraint introduced by CT-VAE during the generation process, which alleviates the impact of heterogeneity to some extent. For example, in the Brain dataset, many nodes that interact and have similar representations do not belong to the same community. During the reconstruction process of CT-VAE, these heterogeneous edges can be effectively ignored, thereby enhancing the accuracy of community detection. In addition, we observed that on the sparse Patent dataset, CT-VAE exhibited promising improvements in NMI and ARI, with increases of 11.26% and 11.73%, respectively, compared to TGC, a method specifically designed for community detection. We attribute this improvement to the design of the prior distribution $p(\mathbf{Z})$. Mathematically, the Student's t-distribution is less sensitive to outlier nodes (such as isolated nodes and low-degree nodes) compared to the Gaussian distribution, which helps to accurately determine the community structure.

CT-CAVAE performs better on non-academic network datasets. Specifically, CT-CAVAE improves ACC, NMI, ARI and F1 by 10.15%, 6.68%, 10.71% and 5.12%, respectively, on the Patent dataset compared to the runner-up. Additionally, it achieves better or comparable performance results than CT-VAE on the Brain and School datasets. Notably, on academic network datasets (e.g., DBLP, arXivAI and arXivCS), although CT-CAVAE does not achieve the best results in every metric, it also achieves average performance improvements on academic networks. For instance, on the DBLP dataset, compared to all baseline methods, ACC, NMI, ARI and F1 show average improvements of 7.8%, 5.84%, 2.21% and 8.24%, respectively.

Discussion. From the perspective of datasets, CT-CACAE is more adept at handling event-driven non-academic networks compared to CT-VAE. This is because interactions in these networks are typically strongly causal, making it easier for the Hawkes process to model the impact of past events on future interactions. Under a community-oriented objective, community information can be better integrated into the node representation $\mathbf{Z}$ during the inference process. For instance, on the Patent dataset, ACC, NMI, ARI and F1 are improved by 14.38%, 1.91%, 6.11% and 4.83%, respectively. In contrast, the performance of CT-CAVAE in academic networks such as DBLP, arXivAI and arXivCS is hindered by multiple factors. Firstly, academic networks exhibit complex, diverse, and often asynchronous interactions. For example, the uncertainty of the paper publication cycle obscures the causality of interaction patterns. Additionally, the intermittent nature of interactions makes it challenging

Datasets	Metrics	node2vec	HTNE	JODIE	TGAT	TGN	MTNE	M^{2}DNE	TREND	TGC	CT-VAE	Avg.	CT-CAVAE	Avg.
DBLP	ACC	46.76	47.13	20.79	36.76	19.78	46.69	46.04	46.82	47.91	**48.40**	8.55↑	47.84	7.80↑
	NMI	34.97	35.23	11.67	28.98	9.82	34.46	33.89	36.56	36.74	**37.25**	8.10↑	35.24	5.84↑
	ARI	20.71	20.53	11.32	17.64	5.46	18.99	19.94	**22.83**	22.23	22.69	4.95↑	20.17	2.21↑
	F1	44.01	43.45	13.23	34.22	10.66	43.96	39.07	**44.54**	43.91	44.41	9.18↑	43.51	8.24↑
Brain	ACC	42.02	42.3	19.14	41.43	17.40	41.48	42.54	39.83	42.88	42.02	5.46↑	**46.94**	10.38↑
	NMI	46.55	50.08	10.50	48.72	8.04	48.74	50.07	45.64	50.32	50.42	10.57↑	**51.25**	10.97↑
	ARI	26.72	29.45	5.00	23.64	4.56	27.72	29.56	22.82	29.91	30.54	8.39↑	**32.92**	10.34↑
	F1	**46.74**	41.36	11.12	41.13	13.49	41.87	41.47	33.67	41.68	40.37	5.64↑	39.48	5.46↑
Patent	ACC	38.09	45.98	30.82	38.26	38.77	42.93	47.90	38.72	50.36	46.13	4.82↑	**60.51**	18.30↑
	NMI	28.85	26.3	9.55	19.74	8.24	30.89	31.53	14.44	25.04	36.30	14.68↑	**38.21**	15.76↑
	ARI	18.36	18.50	7.46	13.31	6.01	20.83	25.94	13.45	18.81	30.54	14.69↑	**36.65**	19.44↑
	F1	30.70	38.33	20.83	26.97	21.40	34.26	39.16	28.41	38.69	39.45	8.48↑	**44.28**	12.34↑
School	ACC	90.84	99.08	65.64	80.54	31.71	99.69	98.78	94.18	99.39	**100.00**	15.57↑	**100.00**	14.55↑
	NMI	92.17	98.09	63.82	73.25	19.45	99.36	97.48	89.55	98.74	**100.00**	18.68↑	**100.00**	17.81↑
	ARI	90.25	98.00	71.94	80.04	32.12	99.33	97.45	87.5	98.71	**100.00**	16.07↑	**100.00**	14.99↑
	F1	93.44	99.05	68.53	79.56	29.50	99.69	98.68	94.18	99.35	**100.00**	15.34↑	**100.00**	14.61↑
arXivAI	ACC	64.98	63.29	30.71	48.69	31.25	64.07	70.10	32.79	74.96	**76.66**	23.23↑	61.63	6.91↑
	NMI	36.20	36.51	32.16	32.12	24.74	36.43	42.97	19.92	44.34	**46.33**	12.40↑	37.80	2.74↑
	ARI	40.39	39.04	33.47	30.34	11.91	39.93	49.18	25.37	51.71	**53.33**	17.63↑	36.05	-1.09↓
	F1	52.80	51.25	19.91	43.62	21.93	51.95	57.48	23.09	59.20	**60.05**	17.69↑	45.70	2.54↑
arXivCS	ACC	29.43	37.64	11.27	20.53	20.10	39.82	38.55	19.94	36.39	**42.65**	14.46↑	29.49	-0.16↓
	NMI	41.65	43.85	15.50	38.64	16.21	44.14	43.91	25.58	44.55	**45.75**	10.86↑	41.57	6.22↑
	ARI	20.56	35.02	25.74	15.54	18.63	36.12	35.26	23.48	34.12	**36.58**	9.42↑	20.08	-8.86↓
	F1	21.60	23.39	12.71	13.23	22.67	24.17	23.11	14.55	25.64	**27.61**	7.49↑	22.50	1.71↑

Table 1: Community detection result %. (Bold: best; "_": runner-up)

to utilize timing information effectively. Specifically, during the inference process, the Hawkes process imposes a decay factor on historical events, meaning that interactions that appear recent may actually have occurred a long time ago relative to the current moment. Secondly, well-known scholars often have multiple research interests and interact frequently with many nodes simultaneously, yet these nodes may belong to only one community. This can lead to negative effects when applying a community-oriented objective during the inference process of CT-CAVAE.

Ablation Analysis

We designed two ablation scenarios to verify the effectiveness of our proposed strategies. Specifically, we use /o to denote the application of Eq. (4) for reconstructing the original network in the generation component. Additionally, considering that CT-CAVAE performs community alignment in Eq. (10) and Eq. (11), we aim to explore whether utilizing K-means in CT-CAVAE can also achieve favorable results. Here, we use /k to represent the scalable variant. The results of the ablation experiments are reported in Table 2, and we make the following observations:

In general, the experimental data demonstrate that our proposed reconstruction strategy can effectively mitigate the negative impact of relaxing the homogeneity constraint. Specifically, by comparing CT-VAE and CT-VAE/o, it can be observed that the former exhibits performance improvements across most datasets. For instance, on the Patent dataset, ACC, NMI, ARI and F1 increased by 1.45%, 2.09%, 0.5% and 0.5%, respectively. Notably, despite the lower homogeneity characteristics of the Brain dataset, which are likely to hinder the performance of existing methods, CT-VAE still performs well under our compact reconstruction constraints. When comparing CT-CAVAE and CT-CAVAE/o, more significant performance improvements are evident. On the same Patent dataset, ACC, NMI, ARI and F1 increased by 5.79%, 1.53%, 1.98% and 3.17%, respectively.

Comparing CT-VAE and CT-VAE/k, we found that CT-VAE achieved favorable results on some academic networks by utilizing community alignment. Specifically, on the DBLP dataset, CT-VAE/k improved by 0.11%, 0.08%, 0.19% and 0.02% in ACC, NMI, ARI and F1, respectively, compared to CT-VAE. Additionally, on some non-academic networks, we observed that CT-VAE/k also achieved certain performance improvements in non-core community detection metrics (i.e., ACC and F1). For instance, on the Patent dataset, ACC improved by 7.76%, and on the Brain dataset, F1 improved by 1.08%. Overall, the strategy of detecting communities using K-means after community alignment appears to be more conducive to improving ACC and F1, but it is lacking in NMI and ARI.

Analysis of the Negative Sampling

In the inference component, we use negative sampling to reduce the time overhead associated with calculating the sub-equation in Eq. (3). In this section, we explore the impact of different numbers of negative samples (i.e., $|P(n)| \in \{2, 4, 6, 8, 10\}$) on CT-VAE, and the results are shown in Figure 3. We have the following observations:

In general, the number of negative samples does not exhibit a clear pattern in the performance of CT-VAE across the six datasets. For example, it is observed that four metrics on the DBLP dataset increase with the number of negative samples, while on the arXivAI dataset, they generally

Datasets	Metrics	CT-VAE	/o	/k	CT-CAVAE	/o
DBLP	ACC	48.40	48.20	**48.51**	**47.84**	47.30
	NMI	37.25	37.02	**37.33**	**35.24**	34.94
	ARI	22.69	21.86	**22.88**	**20.17**	19.75
	F1	44.41	44.04	**44.43**	**43.51**	42.91
Brain	ACC	42.02	41.72	**43.06**	**46.94**	39.96
	NMI	**50.42**	50.39	50.36	**51.25**	49.84
	ARI	**30.54**	29.87	30.31	**32.92**	29.87
	F1	40.37	39.55	**41.45**	**39.48**	31.30
Patent	ACC	46.13	44.68	**53.89**	**60.51**	54.72
	NMI	**36.30**	34.21	32.59	**38.21**	36.68
	ARI	**30.54**	30.04	26.28	**36.65**	34.67
	F1	39.45	38.95	**39.72**	**44.28**	41.11
School	ACC	**100**	99.39	**100**	**100**	97.55
	NMI	**100**	98.74	**100**	**100**	95.62
	ARI	**100**	98.71	**100**	**100**	94.97
	F1	**100**	99.35	**100**	**100**	97.37
arXivAI	ACC	**76.66**	76.52	76.45	**61.63**	61.50
	NMI	**46.33**	46.13	46.03	**37.80**	37.14
	ARI	**53.33**	53.11	52.99	**36.05**	35.57
	F1	**60.05**	59.90	59.97	**45.70**	45.48
arXivCS	ACC	42.65	**43.55**	43.09	29.49	**29.51**
	NMI	**45.75**	45.68	45.56	41.57	**41.58**
	ARI	36.58	36.26	**38.12**	**20.08**	**20.08**
	F1	27.61	27.35	**27.77**	**22.50**	**22.50**

Table 2: Community detection result % (Bold: best). /o denote the ablation variants where Eq. (4) is used to replace Eq. (5) in the generation process. /k represents a scalable variant that utilizes K-means for community detection.

show a decreasing trend. Although the number of negative samples can have a positive or negative impact on the performance of CT-VAE, these effects are subtle except for the Patent dataset. For instance, the maximum numerical variance of four metrics in datasets other than the Patent dataset is no more than 0.008%. For the Patent dataset, the numerical variances of ACC, NMI, ARI and F1 are 0.37%, 0.02%, 0.06% and 0.24%, respectively, with ACC and ARI showing more noticeable fluctuations, while NMI and F1 are less affected. In other words, on the Patent dataset, the impact of varying the number of negative samples on community division results remains manageable. It is worth noting that although we set the default number of negative samples for CT-VAE to 2 (a non-optimal value) across all datasets in our experimental settings, it still achieves impressive performance. This indicates that our proposed method does not require extensive time for parameter fine-tuning.

Related Works

In recent years, many methods based on the Hawkes process have been proposed to address the challenge of learning continuous dynamic network representations. Specifically, (Zuo et al. 2018) were the first to model the complete temporal formation process of continuous dynamic networks, characterized by sequential interactive events, using the Hawkes process. They proposed inputting the interaction of low-dimensional vectors into the Hawkes process as the base rate and time influence, respectively. Subsequently, (Lu et al. 2019) posited that both micro and macro dynamics are key

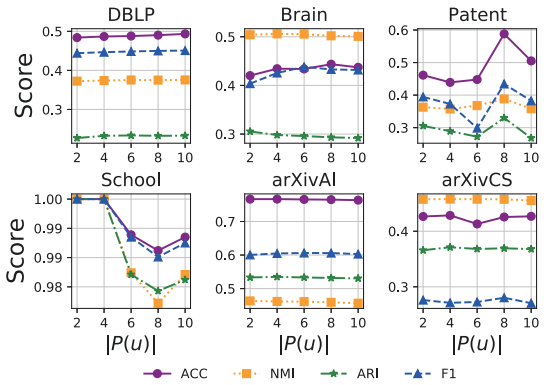

Figure 3: Impact of negative sample count on performance.

factors in network evolution, and thus incorporated these two types of dynamic evolution information into the learning process of node representation. Building on this, (Huang et al. 2020) proposed studying the embedding of node representations through specific ternary motifs from a mesoscopic perspective, in combination with the work of (Zhou et al. 2018). Additionally, inspired by the message propagation mechanism of GNN, some works focus on applying this mechanism to continuous dynamic networks or combining it with the Hawkes process. Specifically, (Trivedi et al. 2019) proposed learning the latent representation of nodes from three perspectives: localized embedding propagation, self-propagation and exogenous drive. (Chen et al. 2024) proposed a self-supervised continuous dynamic network representation learning framework that simultaneously learns the structure and evolution characteristics of dynamic graphs by defining a temporal subgraph contrast learning task. (Wen and Fang 2022) concentrated on the integration of the Hawkes process with the message propagation mechanism. They specifically proposed using self-information as the base intensity of the Hawkes process and first-order neighbor information as the historical influence, deriving the conditional intensity function through a transfer function. Building on this, (Liu et al. 2024a) not only used historical neighbor sequence information for modeling but also considered structural knowledge at both local and global levels.

Conclusion

In this paper, we aim to extend variational autoencoder to community detection in continuous dynamic networks. However, the non-adjacent structure of continuous dynamic networks, along with their inherent timeliness and causality, necessitates careful consideration in designing the inference and generation components within variational autoencoder. To address these challenges, we propose CT-VAE and its scalable variant, CT-CAVAE. By incorporating the Hawkes process into the inference process and integrating pseudo-labeling and compact constraint strategies in the generation phase, our methods enable the efficient application of extensions within the variational autoencoder framework. Extensive experiments on multiple real-world datasets demonstrate that our method outperforms existing methods.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62477016 and 62077045, Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515011758, National Key Research and Development Program of China (Research and Demonstration Application of Key Technologies for Personalized Learning Driven by Educational Big Data) under Grant 2023YFC3341200, and Scientific Research Innovation Project of Graduate School of South China Normal University under Grant 2024KYLX067.

References

Bhattacharya, S.; Liu, Z.; and Maiti, T. 2024. Comprehensive study of variational Bayes classification for dense deep neural networks. *Stat. Comput.*, 34(1): 17.

Blei, D. M.; Kucukelbir, A.; and McAuliffe, J. D. 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518): 859–877.

Chen, K.; Liu, L.; Jiang, L.; and Chen, J. 2024. Self-Supervised Dynamic Graph Representation Learning via Temporal Subgraph Contrast. *ACM Trans. Knowl. Discov. Data*, 18(1): 18:1–18:20.

Cheng, J.; He, C.; Han, K.; Chen, G.; Liang, W.; and Tang, Y. 2024. Unveiling community structures in static networks through graph variational Bayes with evolution information. *Neurocomputing*, 576: 127349.

Cheng, J.; He, C.; Han, K.; Ma, W.; and Tang, Y. 2023. How Significant Attributes are in the Community Detection of Attributed Multiplex Networks. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, 2057–2061. ACM.

Choong, J. J.; Liu, X.; and Murata, T. 2018. Learning Community Structure with Variational Autoencoder. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, 69–78. IEEE Computer Society.

Grover, A.; and Leskovec, J. 2016. node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 855–864. ACM.

Huang, H.; Fang, Z.; Wang, X.; Miao, Y.; and Jin, H. 2020. Motif-Preserving Temporal Network Embedding. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, 1237–1243. ijcai.org.

Khoshraftar, S.; and An, A. 2024. A Survey on Graph Representation Learning Methods. *ACM Trans. Intell. Syst. Technol.*, 15(1): 19:1–19:55.

Kipf, T. N.; and Welling, M. 2016. Variational Graph Auto-Encoders. *NIPS Workshop on Bayesian Deep Learning*.

Kumar, S.; Zhang, X.; and Leskovec, J. 2019. Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, 1269–1278. ACM.

Li, D.; Li, D.; and Lian, G. 2023. Variational Graph Autoencoder with Adversarial Mutual Information Learning for Network Representation Learning. *ACM Trans. Knowl. Discov. Data*, 17(3): 45:1–45:18.

Li, T.; Wang, W.; Jiao, P.; Wang, Y.; Ding, R.; Wu, H.; Pan, L.; and Jin, D. 2023. Exploring Temporal Community Structure via Network Embedding. *IEEE Trans. Cybern.*, 53(11): 7021–7033.

Liang, K.; Liu, Y.; Zhou, S.; Tu, W.; Wen, Y.; Yang, X.; Dong, X.; and Liu, X. 2023. Knowledge graph contrastive learning based on relation-symmetrical structure. *IEEE Transactions on Knowledge and Data Engineering*, 36(1): 226–238.

Liang, K.; Meng, L.; Liu, M.; Liu, Y.; Tu, W.; Wang, S.; Zhou, S.; Liu, X.; Sun, F.; and He, K. 2024. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Liu, M.; Liang, K.; Zhao, Y.; Tu, W.; Zhou, S.; Gan, X.; Liu, X.; and He, K. 2024a. Self-Supervised Temporal Graph Learning With Temporal and Structural Intensity Alignment. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.

Liu, M.; Liu, Y.; LIANG, K.; Tu, W.; Wang, S.; sihang zhou; and Liu, X. 2024b. Deep Temporal Graph Clustering.

Lu, Y.; Wang, X.; Shi, C.; Yu, P. S.; and Ye, Y. 2019. Temporal Network Embedding with Micro- and Macro-dynamics. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, 469–478. ACM.

Mrabah, N.; Bouguessa, M.; and Ksantini, R. 2024. A contrastive variational graph auto-encoder for node clustering. *Pattern Recognit.*, 149: 110209.

Paranjape, A.; Benson, A. R.; and Leskovec, J. 2017. Motifs in Temporal Networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, February 6-10, 2017*, 601–610. ACM.

Pei, H.; Wei, B.; Chang, K. C.; Lei, Y.; and Yang, B. 2020. Geom-GCN: Geometric Graph Convolutional Networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Rossi, E.; Chamberlain, B.; Frasca, F.; Eynard, D.; Monti, F.; and Bronstein, M. M. 2020. Temporal Graph Networks for Deep Learning on Dynamic Graphs. *CoRR*, abs/2006.10637.

Salha-Galvan, G.; Lutzeyer, J. F.; Dasoulas, G.; Hennequin, R.; and Vazirgiannis, M. 2022. Modularity-aware graph autoencoders for joint community detection and link prediction. *Neural Networks*, 153: 474–495.

Su, X.; Xue, S.; Liu, F.; Wu, J.; Yang, J.; Zhou, C.; Hu, W.; Paris, C.; Nepal, S.; Jin, D.; Sheng, Q. Z.; and Yu, P. S. 2024. A Comprehensive Survey on Community Detection

With Deep Learning. *IEEE Trans. Neural Networks Learn. Syst.*, 35(4): 4682–4702.

Trivedi, R.; Farajtabar, M.; Biswal, P.; and Zha, H. 2019. DyRep: Learning Representations over Dynamic Graphs. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

van der Maaten, L.; and Hinton, G. E. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9: 2579–2605.

Wen, Z.; and Fang, Y. 2022. TREND: TempoRal Event and Node Dynamics for Graph Representation Learning. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, 1159–1169. ACM.

Xu, D.; Ruan, C.; Körpeoglu, E.; Kumar, S.; and Achan, K. 2020. Inductive representation learning on temporal graphs. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Zhou, L.; Yang, Y.; Ren, X.; Wu, F.; and Zhuang, Y. 2018. Dynamic Network Embedding by Modeling Triadic Closure Process. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, 571–578. AAAI Press.

Zhu, J.; Yan, Y.; Zhao, L.; Heimann, M.; Akoglu, L.; and Koutra, D. 2020. Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Zuo, Y.; Liu, G.; Lin, H.; Guo, J.; Hu, X.; and Wu, J. 2018. Embedding Temporal Network via Neighborhood Formation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, 2857–2866. ACM.