Ranking on Dynamic Graphs: An Effective and Robust Band-Pass Disentangled Approach

Anonymous Author(s)*

ABSTRACT

Ranking is an essential and practical task on dynamic graphs, which aims to prioritize future interaction candidates for given queries. While existing solutions achieve promising ranking performance, they leverage a single listwise loss to jointly optimize candidate sets, which leads to the gradient vanishing issue; and they employ neural networks to model complex temporal structures within a shared latent space, which fails to accurately capture multi-scale temporal patterns due to the frequency aliasing issue. To address these issues, we propose BandRank, a novel and robust band-pass disentangled ranking approach for dynamic graphs in the frequency domain. Concretely, we propose a band-pass disentangled representation (BPDR) approach, which disentangles complex temporal structures into multiple frequency bands and employs non-shared frequency-enhanced multilayer perceptrons (MLPs) to model each band independently. We prove that our BPDR approach ensures effective multi-scale learning for temporal structures by demonstrating its multi-scale global convolution property. Besides, we design a robust Harmonic Ranking (HR) loss to jointly optimize candidate sets and continuously track comparisons between real and virtual candidates, where we theoretically guarantee its ability to alleviate the gradient vanishing issue. Extensive experimental results show that our BandRank achieves an average improvement of 21.31% against eight baselines while demonstrating superior robustness across different learning scenarios.

KEYWORDS

Dynamic graphs; Ranking; Disentangled representation learning

ACM Reference Format:

Anonymous Author(s). 2018. Ranking on Dynamic Graphs: An Effective and Robust Band-Pass Disentangled Approach. In *Proceedings of The Web Conference (WWW)*. ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

Statement of Relevance This paper proposes BandRank, a novel ranking approach for dynamic graphs in the frequency domain, which addresses key challenges in graph algorithms for the web. Our approach is highly relevant to the "Graph Algorithms and Modeling for the Web" track, as it enhances the accuracy and robustness of the ranking task in dynamic graphs, contributing to more effective graph-based web modeling.

WWW, April 28 – May 02, 2025, Sydney, AU

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

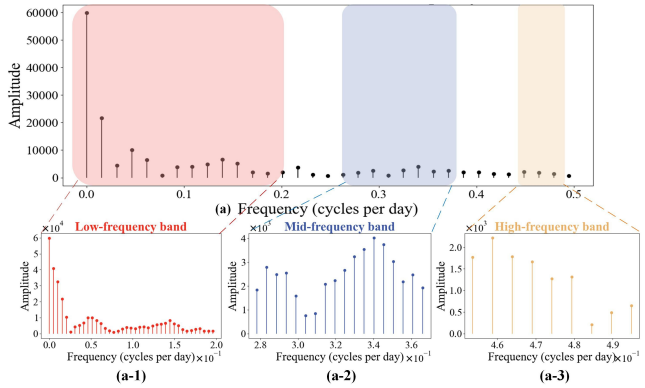


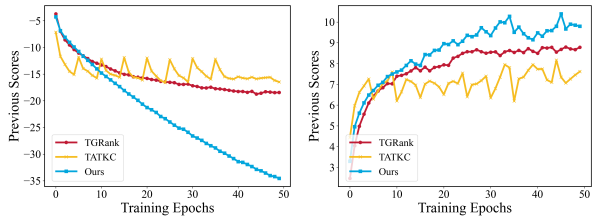
Figure 1: (a) Frequency spectrum of UCI dataset; and its low- mid- and high-frequency band illustrations.

1 INTRODUCTION

Dynamic graphs, which preserve evolving relationships (edges) between entities (nodes) over time, serve as fundamental structures for modeling complex real-world systems, such as social networks [13, 21, 42], recommendation systems [10, 34, 61], and financial markets [11, 28, 31]. In these contexts, a common and key task is predicting and prioritizing future interactions, which introduces link prediction and link ranking tasks. This paper focuses on link ranking which involves ranking interaction candidates for given nodes based on their likelihood. For example, on e-commerce platforms, users are exposed to *ranked* product lists based on their previous behaviors, which enhances personalized experiences.

Recent research has made progress in ranking on dynamic graphs. A pioneering work, TGRank [36], tackles the ranking problem by leveraging a listwise loss and employs a labeling trick to enhance expressivity. Although achieving promising performance in subgraph-level ranking, it is constrained by the gradient vanishing problem stemming from its reliance on a single listwise loss. This compromises the learning ability during model training. Later, TATKC [62] is designed to mirror the temporal Katz centrality (TKC) calculation for node ranking over dynamic graphs. It leverages a self-attention mechanism to learn node embeddings and uses a pairwise loss to train the model to approximate TKC, enabling graph-level node ranking. However, despite its focus on TKC approximation, TATKC employs a pairwise loss that independently optimizes over future interactions rather than jointly optimizing across candidate sets, resulting in poor generalization ability and sub-optimal performance. In light of this, we aim to further investigate the ranking problem over dynamic graphs by tackling the following challenges.

Challenge I: How to accurately model multi-scale temporal patterns over dynamic graphs? Dynamic graphs naturally exhibit multi-scale temporal patterns, encompassing dynamics that



(a) Virtual candidate scores (b) Real candidate scores

Figure 2: Previous scores during the training on UCI dataset.

range from long-term to short-term evolution. Existing ranking-driven approaches over dynamic graphs [36, 62] typically model multi-scale temporal patterns along with temporal structures within a shared latent space in the time domain. However, they easily overfit to long-term evolution (the majority pattern) while overlooking other dynamics, such as periodic behaviors, resulting in the frequency aliasing issue and thus compromising model robustness. Although there are a few explorations of dynamic graph learning in the frequency domain (e.g., FreeDyG [37]), they focus on the link prediction task and still model multi-scale temporal patterns within a shared latent space in the time domain, which also leads to the frequency aliasing issue. To further explore the frequency domain, we visualize the temporal patterns of the UCI dataset using the Fourier Transform. Fig. 1a reveals that the low-frequency signals dominate the frequency spectrum, reflecting the long-term evolution inherent in dynamic graphs. In contrast, mid- and high-frequency signals exhibit lower amplitudes, making it challenging for existing approaches to accurately capture them. Consequently, these approaches often overemphasize the 'highest amplitude' bands or become overly sensitive to high-frequency fluctuations, either of which incurs the *frequency aliasing* issue, ultimately leading to sub-optimal performance for ranking tasks (see Fig. 2).

Dynamic Graph Analysis in the Frequency Domain: According to graph signal theory [33], each frequency band preserves distinct temporal patterns in the frequency domain. Taking the UCI dataset as an example, we make the following analysis: (1) the low-frequency band (Fig. 1a-1) contains the highest energy (the largest amplitudes), representing the long-term evolution that dominates the dynamic graph; (2) the mid-frequency band (Fig. 1a-2) exhibits multiple peaks with varying amplitudes, which are less prominent in the overall spectrum, indicating medium-scale repetitive or periodic patterns, such as community interactions; and (3) the high-frequency band (Fig. 1a-3) contains several peaks with the lowest amplitudes, barely visible in the overall spectrum, highlighting abrupt fluctuations often triggered by unexpected events.

However, existing approaches overlook the properties of temporal dynamics in the frequency domain as well as the frequency aliasing issue, thus compromising model effectiveness and robustness. This motivates us to explicitly model multi-scale temporal patterns for the ranking task.

Challenge II: How to prevent the gradient vanishing issue of the ranking loss in dyanmic graph learning? For the ranking problem, the listwise loss has been shown to be effective for jointly optimizing candidate sets, as highlighted in [36]. To explore this further, we conduct experiments tracking the evolution of previous scores for virtual and real candidates across epochs on the UCI dataset using two ranking approaches (i.e., TGRank [36] and TATKC [62]), where the pairwise loss in TATKC is optimized with the listwise loss for consistent comparison. We can observe from Fig. 2 that previous scores of virtual candidates for both TGRank and TATKC tend to plateau and level off in Fig. 2a, suggesting that their gradients for virtual candidates vanish over time. This weakens their ability to differentiate between real and virtual candidates during training, leading to relatively inferior learning ability towards real candidates, as shown in Fig. 2b. We aim to prevent the gradient vanishing issue to enhance ranking performance.

In this paper, we propose BandRank, a novel and robust Band-pass disentangled Ranking approach for dynamic graphs in the frequency domain. To address **Challenge I** and based on our frequency analysis, we propose a novel band-pass disentangled representation (BPDR) approach that contains a band-pass filter with adaptive quality factor and multi-scale learning for temporal structures. Concretely, we first design quality factors to adaptively determine frequency centers based on the energy distribution. These quality factors are employed to disentangle dynamic graph features into multiple frequency bands via the band-pass filter. Non-shared frequency-enhanced MLPs are then used to capture temporal patterns and structures within each frequency band, generating band-specific embeddings while alleviating the frequency aliasing issue. These embeddings are subsequently integrated to generate high-quality node representations. We theoretically demonstrate that our BPDR approach achieves multi-scale global convolution property, which ensures a comprehensive understanding of multi-scale temporal patterns including long-term evolution, periodic patterns, and abrupt fluctuations. Besides, we design a robust Harmonic Ranking (HR) Loss that can track comparison between real and virtual candidates by our proposed harmonic term and jointly optimize candidate sets by the listwise loss. We prove that our HR loss mitigates the gradient vanishing issue by maintaining a consistent gradient signal throughout the training, thus addressing **Challenge II**. The main contributions are summarized as

- We provide insights into the properties of temporal patterns within dynamic graphs in the frequency domain. We propose a novel band-pass disentangled representation approach that adaptively disentangles these dynamics into distinct frequency bands, enabling effective multi-scale learning for temporal structures.
- We design a robust Harmonic Ranking loss, which optimizes candidate sets jointly while theoretically ensuring a consistent gradient signal throughout the training process, thereby preventing the gradient vanishing issue.
- Extensive experimental results validate the effectiveness of our proposed BandRank in the ranking task, achieving an average improvement of 21.31% against the best competitors. We empirically demonstrate its superior robustness from three key perspectives.

2 RELATED WORK

We briefly review the dynamic graph learning works and frequency-domain representation learning techniques.

2.1 Dynamic Graph Learning

Dynamic graph learning has gained significant attention recently due to its broad applicability. These approaches can be roughly categorized into two groups: classification-driven and ranking-driven.
The classification-driven line encompasses three main learning tasks: link prediction [4, 28, 30, 43, 44, 47, 57, 59, 66], node classification [12, 28, 35, 40, 47, 50, 63], and anomaly detection [7, 27, 48, 56, 58]. These works primarily focused on developing various temporal graph neural networks to enhance performance in these tasks. Some works [19, 38, 41, 47] studied temporal dependency modeling and structure learning to generate high-quality temporal embeddings, with TGN [32] serving as a general framework. Building on the general framework, co-neighbor encoder techniques were explored to capture co-occurrence events, especially improving performance in inductive learning settings. Additionally, accelerated techniques such as cache management were integrated into temporal graph neural networks to enhance model training efficiency[22–24, 51]. However, these approaches primarily operate in the time domain, making them challenging to capture long-term periodicities effectively. More recently, FreeDyG [37] introduced an adaptive filter in the frequency domain, applied before structure learning, allowing the model to capture periodic patterns and shifting phenomena. Despite this, it overlooked multi-scale temporal pattern modeling, which was critical for fully capturing the complexity of dynamic graphs, thus limiting model performance.
For the ranking-driven line, existing research explores both representation learning techniques and loss function, aiming to model a candidate set for each node during training. Specifically, TGRank [36] addressed the subgraph-level ranking task by applying a listwise ranking loss to optimize the relative ordering of candidate nodes. It modeled temporal structures within a shared space in the time domain using the MLPs. However, TGRank suffered from the gradient vanishing issue, which resulted in poor discrimination between real and virtual candidates. The recent TATKC [62] leveraged the self-attention mechanism and pairwise loss to mirror the calculation of temporal Katz centrality (TKC), facilitating graph-level node ranking in dynamic graphs. Nevertheless, TATKC failed to jointly optimize across candidate sets based on pairwise loss, leading to suboptimal performance. Additionally, ranking-driven approaches typically learn multi-scale temporal patterns within a shared space in the time domain, leading to the frequency aliasing issue that compromises the model's robustness.

2.2 Frequency-domain Representation Learning

The frequency domain inherently represents signals as a sum of various frequency components, making it a powerful tool in many fields such as computer vision [16, 25, 26], natural language processing [14, 46, 49, 53], and time series analysis [6, 8, 9, 37, 54, 55, 65]. For instance, several attempts, such as [39, 52, 60, 64, 65], integrated Fourier theory into deep neural networks to capture temporal and periodic patterns in data that may not be easily detectable in the time domain, such as in FourierGNN [54] for multivariate time series forecasting. Additionally, frequency filters [9, 37, 45] were extensively investigated to help models preserve important information while filtering out irreverent ones, enabling more effective learning for subsequent tasks, such as image deblurring [18]. Nevertheless, existing approaches often treated all frequency components uniformly, failing to accurately learn signals within specific frequency bands, which limited their ability to capture multi-scale patterns effectively. In addition, frequency-enhanced representation learning remains largely underexplored in dynamic graph learning. Thus, we aim to explore effective multi-scale temporal pattern modeling to enhance ranking performance on dynamic graphs.

3 PRELIMINARIES

3.1 Continuous-Time Dynamic Graphs

Definition 1 (Continuous-Time Dynamic Graph (CTDG)). *A continuous-time dynamic graph can be denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$, where $\mathcal{V}$ is the node set, $\mathcal{E}$ is the edge set, and $\mathcal{X}$ is the feature matrix. Each edge $e(t) \in \mathcal{E}$ represents a temporal interaction event occurring between nodes u and v at a timestamp t.*

In a CTDG, multiple interaction events can occur between two nodes at different timestamps, with the edge feature $\boldsymbol{x}_e$ potentially varying over time. For simplicity, we refer to CTDGs as dynamic graphs when the context is clear.

3.2 Problem Definition

We first define queries in ranking over dynamic graphs. Concretely, each query $q_i = (c_i, \mathbb{D}_i, t)$ occurs at time t, comprising a center node c_i and several associated ranking candidate nodes $\mathbb{D}_i = \{v_1, v_2, \ldots, v_{n_i}\}$. The next interaction of c_i is referred to as the *real candidate*, while others are *virtual candidates*. A dynamic subgraph $\mathcal{G}_i(t)$ is induced by the query q_i, recording its K-hop neighbors within $\mathcal{G}$ before timestamp t. Next, we present the problem definition.

Problem 1 (Ranking on Continuous-Time Dynamic Graphs). *Given a CTDG $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$ and a query set $\{q_1, q_2, \ldots, q_m\}$. For each query $q_i = (c_i, \mathbb{D}_i, t)$ and its induced subgraph $\mathcal{G}_i(t)$ with candidate nodes $\mathbb{D}_i = \{v_1, v_2, \ldots, v_{n_i}\}$, we aim to learn a function $\mathcal{F}$ that encodes induced subgraph $\mathcal{G}_i(t)$ and maps each query q_i to a ranking score for each candidate node, denoted as $\mathcal{F}(q_i|\mathcal{G}_i(t)) = \{y_{ij}\}_{j=1}^{n_i}$, where y_{ij} is a ranking score of v_j for c_i.*

3.3 Discrete Fourier Transform

Discrete Fourier Transform (DFT) lays the foundation for exploring representation learning in the frequency domain, which can convert time-domain signals into the frequency domain. The DFT allows us to analyze different frequency components and detect complex temporal patterns within dynamic graphs. Concretely, given a time-domain signal represented as a sequence of temporal interactions, the DFT is defined as

$$X(f) = \sum_{n=0}^{N_s-1} x(n) e^{-i2\pi f n / N_s}, \tag{1}$$

where $x(n)$ is the discrete time-domain signal, N_s is the length of the signal, and f represents the frequency components. To revert

the signal back to the time domain, the Inverse Discrete Fourier Transform (IDFT) is used:

$$x(n) = \frac{1}{N_s} \sum_{f=0}^{N_s-1} X(f) e^{i2\pi f n / N_s}. \quad (2)$$

In practice, the Fast Fourier Transform (FFT) is commonly used to compute the DFT efficiently by recursively breaking down the DFT computation. Similarly, the IDFT can be efficiently computed using the inverse Fast Fourier Transform (IFFT). For more derivations, please refer to [1].

4 THE PROPOSED BANDRANK

4.1 Overview

In this paper, we propose a band-pass disentangled ranking approach (BandRank) for dynamic graphs, which can accurately model multi-scale temporal patterns inherent in dynamic graphs. To achieve this, we first design a band-pass disentangled representation (BPDR) approach inspired by the concept of a band-pass filter, which adaptively disentangles the frequency components of temporal structures into N bands, allowing the model to selectively focus on specific frequency ranges during the learning. Here we leverage non-shared frequency-enhanced MLPs for each frequency band, theoretically ensuring comprehensive multi-scale learning within dynamic graphs. Additionally, we design a harmonic ranking loss based on listwise loss to jointly optimize candidate sets, which alleviates the gradient vanishing issue. We elaborate on each component below.

4.2 Band-Pass Disentangled Representation

4.2.1 Motivation. Dynamic graphs are complex data structures that evolve over time, inherently exhibiting multi-scale temporal patterns, including (1) long-term evolution, representing smooth changes in graph structures over extended periods; (2) periodic patterns, where structures recur at regular intervals; and (3) abrupt fluctuations, which involve abrupt changes or anomalies. Different temporal patterns are naturally projected into distinct frequency bands with individual energy distribution, as shown in Figs. 1a-1 to a-3. Looking at Fig. 1a, it becomes evident that the low-frequency energy distribution dominates the entire frequency spectrum. In light of the above, existing solutions tend to overemphasize the low-frequency patterns or become overly sensitive to high-frequency fluctuations, leading to the frequency aliasing issue. Motivated by our analysis, we aim to disentangle complex dynamics to accurately capture temporal structures for ranking-driven dynamic graph learning.

4.2.2 Overall of the Proposed BPDR. Here, we explore the potential of applying the concept of band-pass filtering [5] to dynamic graph representation learning and propose a novel Band-Pass Disentangled Representation (BPDR) approach in the frequency domain for dynamic graph ranking. It consists of three key modules: *adaptive quality factor computation*, *band-pass filter*, and *multi-scale learning for temporal structures*. Concretely, we define the quality factor based on the energy distribution, serving to adaptively determine the optimal bandwidth for frequency disentanglement. Based on it, we implement a band-pass filter to disentangle the graph signal into N frequency bands, which isolate specific temporal patterns within each band. For each frequency band, we formulate a structure learning process using the frequency-enhanced MLP that captures the unique characteristics of graph dynamics, generating band-specific embeddings. We then integrate these embeddings to update node embeddings. We formulate this process below.

4.2.3 Adaptive Quality Factor Computation. For each query $q_i = (c_i, \mathbb{D}_i, t)$, we have $\mathcal{G}_i(t) = (\mathcal{V}_i(t), \mathcal{E}_i(t), \mathcal{X}_i(t))$ to represent a candidate subgraph for the center node c_i before timestamp t. Our objective is to rank the candidate nodes based on their relevance to the center node. We initialize these nodes with learnable vectors, such as $\mathbf{m}_v^0(t) \in \mathbb{R}^d$. For each edge $e(t) = (c, v, t) \in \mathcal{E}_i(t)$, the l-layer time-domain information $\mathbf{h}_e^l(t)$ is constructed by concatenating previous node embeddings $\mathbf{m}_c^{l-1}(t')$, $\mathbf{m}_v^{l-1}(t')$ with the time encoding and features, capturing the temporal relationship between nodes c and v at time t by $\mathbf{h}_e^l(t) = [\mathbf{m}_c^{l-1}(t') \| \mathbf{m}_v^{l-1}(t') \| \phi(\Delta t) \| \boldsymbol{x}_e]$, where $\phi(\cdot)$ is the time encoding function [32, 47] and $\boldsymbol{x}_e$ is the features including node features and edge features (if available). Given the l-layer input with B subgraphs $\{\mathcal{G}_j(t)\}_{j=1}^B$ and time-domain temporal information $\mathbf{H}^l(t)$, we transform the batch of messages into the frequency domain by

$$\mathcal{H}^l(f) = \text{DomainConversion}(\mathbf{H}^l(t)), \quad (3)$$

where DomainConversion$(\cdot)$ denotes FFT in Eq. (1). Frequency-domain signal $\mathcal{H}^l(f)$ preserves the underlying temporal structures from each subgraph. Then we aim to disentangle the complex temporal structures into N independent bands in the frequency domain. To achieve this, we define the adaptive quality factor based on the energy for frequency component disentanglement. Concretely, we first compute the energy for each frequency component f_i and then compute adaptive quality factor Q_k to determine the frequency center for the k-th frequency band, which can be formulated as

$$E(f_i) = |\mathcal{H}^l(f_i)|^2, i \in [1, F], \quad Q_k = \alpha_k \frac{(2k-1)\sum_i^F E(f_i)}{2N}. \quad (4)$$

Here, α_k is a scaling factor that is updated according to the gradients $\frac{\partial \mathcal{L}_{HR}}{\partial \alpha_k}$ during the training. It allows the model to adaptively adjust the bandwidth of each band for subsequent band-pass filter based on the energy distribution of the signal. Consequently, higher energy levels would produce higher Q_k, which corresponds to narrower bandwidths, focusing the band-pass filter on specific frequency components. Conversely, lower energy levels may yield smaller Q_k, corresponding to broader bandwidths, capturing a wider range of frequencies. By doing so, the quality factor can adaptively adjust its sensitivity to the signal during training, making the band-pass filtering process responsive to various energy components.

4.2.4 Band-Pass Filter. With computed quality factors $Q = [Q_1, Q_2, \ldots, Q_N]$, we aim to disentangle all frequency components of temporal structures into N frequency bands. For computation efficiency, we leverage the simple distance function to assign a specific band for each frequency component f_i of $\mathcal{H}^l(f)$, which is formulated as

$$B(f_i) = \arg\min_k \|E(\mathcal{H}^l(f)) - Q_k\|_2, \quad \forall k \in [1, N], \quad (5)$$

where $B(f_i)$ is the spectral mask, recording its corresponding frequency band. In this way, each frequency component is assigned to

its closest band based on the energy and quality factors. Thus, we achieve $\mathbf{B} = \{\mathbf{B}_k\}_{k=1}^N$ for all components of $\mathcal{H}^l(f)$ and then apply the band-pass filter to disentangle the messages into N message representations by

$$\mathcal{H}_k^l(f) = \mathcal{H}^l(f) \odot \mathbb{I}(\mathbf{B} = k), \quad k \in [1, N]. \tag{6}$$

Here, $\odot$ denotes element-wise multiplication, and $\mathbb{I}(\mathbf{B} = k)$ is an indicator function that ensures only the frequency components corresponding to k-th band are kept. This filtering process disentangles frequency-domain temporal patterns into N frequency bands based on the energy distribution. By adaptively adjusting the quality factors Q_k, the band-pass filters effectively capture the most relevant frequency components for different temporal patterns, from long-term evolution in the low-frequency bands to short-term fluctuation in the high-frequency bands.

4.2.5 Multi-Scale Learning for Temporal Structures. With the disentangled inputs $[\mathcal{H}_1^l(f), \mathcal{H}_2^l(f), \ldots, \mathcal{H}_N^l(f)]$, we accomplish the multi-scale learning for temporal structures, including band-specific embedding generation and node embedding update. First, given a band-specific input $\mathcal{H}_k^l(f)$ in the k-th band, we leverage a simple frequency-enhanced MLP to model temporal patterns, generating band-specific embeddings. We then integrate these band-specific embeddings to generate temporal embeddings and transform them into the time domain for the node embedding update. We formulate these process as

$$\mathcal{Z}_k^l(f) = \text{FreMLP}_k\left(\mathcal{H}_k^l(f)\right), k \in [1, N], \tag{7}$$

$$\mathcal{Z}^l(f) = \psi\left(\sum_{k=1}^N \mathcal{Z}_k^l(f)\right), \tag{8}$$

$$\mathbf{Z}^l(t) = \text{DomainInversion}\left(\mathcal{Z}^l(f)\right), \tag{9}$$

where DomainInversion$(\cdot)$ denotes IFFT in Eq. (2), $\mathcal{Z}_k^l(f)$ represents the band-specific embedding from the k-th frequency band and $\psi(\cdot)$ denotes a complex number linear function. $\mathcal{Z}^l(f)$ is the frequency-domain temporal embedding aggregated from N frequency bands, which is transformed to the time domain to generate time-domain temporal embedding $\mathbf{Z}^l(t)$. Therefore, based on Eqs. (7) - (9), different temporal patterns are independently modeled in different latent spaces, allowing for accurate learning of multi-scale temporal patterns and thus alleviating the frequency aliasing issue.

Last, we update the embedding of each node by aggregating the temporal embeddings $\mathbf{Z}^l(t)$ from its neighbors within the batch. For each node u, we update its node embedding based on its neighbor $\mathcal{N}_u(t)$ by

$$\mathbf{m}_u^l(t) = \text{UPDATE}^l\left(\mathbf{m}_u^{l-1}(t), \left\{\mathbf{z}_{u,v}^l(t) | v \in \mathcal{N}_u(t)\right\}\right), \tag{10}$$

where we leverage a simple sum pooling together with a nonlinear function for the update function UPDATE$(\cdot)$. Based on Eq. (10), the node embeddings are iteratively refined, preserving multi-scale temporal and structural information. Leveraging our band-pass disentangled representation approach, we perform temporal structure learning across multiple latent spaces, effectively preventing the mutual interfere of distinct temporal patterns. Next, we demonstrate how our approach achieves multi-scale learning in the frequency domain by formulating the following theorem.

Theorem 1 (Multi-Scale Global Convolution Theorem). *Let $\mathcal{H}(f)$ be a frequency-domain dynamic graph signal, and assume that $\mathcal{H}(f)$ can be decomposed into N non-overlapping frequency bands via our band-pass filter. Let $\{\text{FreMLP}_k\}_{k=1}^N$ be a set of independent frequency-enhanced multilayer perceptrons (MLPs) for all frequency bands and $\{\mathbb{I}_k\}_{k=1}^N$ be a set of filter indicators based on $\{\mathbf{B}_k\}_{k=1}^N$. Thus, the learning process of band-pass disentangled representation approach in frequency domain is equivalent to $\sum_{k=1}^N \mathbb{I}_k \odot \mathcal{H}(f) \odot \mathbf{C}_k(f)$, where $\odot$ denotes element-wise multiplication and $\mathbf{C}_k(f)$ are the complex number weight.*

The proof of Theorem 1 is provided in Appendix B.1. Theorem 1 ensures the multi-scale learning capability of our band-pass disentangled representation learning approach, providing a theoretical foundation for modeling complex temporal structures over dynamic graphs.

4.3 The Proposed Harmonic Ranking Loss

Listwise loss has been widely used in ranking tasks due to its ability to jointly optimize candidate sets [2, 3, 36]. However, we found that it suffers from the gradient vanishing problem for dynamic graph ranking, particularly for virtual candidate nodes, as highlighted in **Challenge II** and Fig. 2. This weakens the discriminative ability between real and virtual candidates, thus leading to sub-optimal ranking performance over dynamic graphs. To address this, we aim to design a robust ranking loss that can jointly optimize candidate sets and continuously track gradient signals from candidate sets. Concretely, we first design a *harmonic term* to track the comparison between real and virtual candidates during the training. Given a query $q = (c, \mathbb{D}, t)$, we aim to select hard virtual candidates for comparison and formulate it as

$$\mathcal{L}_{\text{harmonic}} = -\log\left(\frac{1}{1 + \exp(-(y_r - y_n^{\text{hard}}))}\right), \tag{11}$$

where y_r is the score of unique real candidate node v_r that c is to connect to next. y_n^{hard} is selected as the hard candidates within $\mathbb{D}(t)$. The selection strategy is similar to [15]. By focusing on the "hard" candidate, our harmonic term remains sensitive to the critical distinctions between real and virtual candidates, maintaining a consistent gradient signal throughout the training process. Building on it, we then present our Harmonic Ranking (HR) loss that integrates our harmonic term with listwise loss, which can be formulated as

$$\mathcal{L}_{\text{HR}} = \mathcal{L}_{\text{listwise}} + \beta \mathcal{L}_{\text{harmonic}}, \tag{12}$$

where $\mathcal{L}_{\text{listwise}} = -\log\left(\frac{\exp(y_r)}{\sum_{j=1}^{|\mathbb{D}(t)|}\exp(y_j)}\right)$ and β controls the balance between the listwise loss and our harmonic term. Unlike listwise loss, where samples contribute less over time, our HR Loss consistently drives the model to distinguish between real and virtual candidates, leading to a robust optimization process that alleviates gradient vanishing and maintains model robustness throughout training. To support this, we provide a Theorem below.

Table 1: Dataset Statistics

Dataset	# Nodes & Attr.	# Edges & Attr.	# Time Span	# Transductive & Inductive Test Edges
Reddit	10,984 & 172	672,447 & 172	1 month	100,867 & 21,470
Wikipedia	9,227 & 172	157,474 & 172	1 month	23,621 & 11,715
MOOC	7,144 & 0	411,749 & 4	17 months	61,763 & 29,179
LastFM	1,980 & 0	1,293,103 & 0	1 month	193,966 & 98,442
Enron	184 & 0	125,235 & 0	3 years	18,785 & 4,206
UCI	1,899 & 0	59,835 & 0	196 days	8,976 & 5,932

Theorem 2 (Robustness of Harmonic Ranking Loss for Virtual Condidates). *Given a query $q = (c, \mathbb{D}, t)$ with its real candidate v_r and its virtual candidates $\mathbb{D} \setminus v_r$. We can obtain that the gradient of our Harmonic Ranking (HR) loss with respect to virtual candidate scores $\boldsymbol{y}_{\mathbb{D}\setminus v_r}$ is greater than those of Listwise loss, satisfying the following inequality:*

$$\frac{\partial L_{HR}}{\partial \boldsymbol{y}_{\mathbb{D}\setminus v_r}} - \frac{\partial L_{listwise}}{\partial \boldsymbol{y}_{\mathbb{D}\setminus v_r}} > 0. \tag{13}$$

The proof is provided in Appendix B.2. Theorem 2 ensures that our harmonic ranking loss maintains a greater gradient for virtual candidates compared to the listwise loss, thereby alleviating the gradient vanishing problem and enhancing ranking robustness.

5 EXPERIMENTS

5.1 Experimental Setting

5.1.1 Datasets. We collect six real-world dynamic graphs from [19, 20, 41] for ranking evaluation, covering domains such as social networks, user-item interactions, and music engagement. Table 1 presents detailed statistics of these datasets.

5.1.2 Baselines. We evaluate our BandRank against eight state-of-the-art dynamic graph learning approaches in the ranking setting. Selected baselines covers three groups: ranking-driven approaches (TGRank [36] and TATKC [62]), classification-driven approaches (DyRep [38], JODIE [19], TGAT [47], TGN [32], and CAWN [41]), and a frequency-enhanced classification-driven approach (FreeDyG [37]). To improve TATKC's performance in subgraph-level ranking, we replace its original pairwise loss with the listwise loss (called 'TATKC-LW'), as the original version does not yield satisfactory results. For a fair comparison, all comparative approaches are trained using the same subgraph sampling strategy from [36].

5.1.3 Evaluation Metrics. We evaluate the model under two learning scenarios: transductive and inductive learning. In the transductive setting, each dataset is chronologically divided into training $[0, T_{\text{train}}]$, validation $[T_{\text{train}}, T_{\text{val}}]$, and testing intervals $[T_{\text{val}}, T]$ following a 70%-15%-15% split, aligned with previous studies [32, 36, 47]. For the inductive setting, we reserve 10% of the nodes for inductive tasks through random sampling, ensuring that both the center nodes and their real candidates in the test queries remain entirely unseen during training. For each query $q = (c, \mathbb{D}, t)$, we extract the induced subgraph $\mathcal{G}_c(t)$ and perform node downsampling. For the evaluation metrics, we employ the mean reciprocal rank (MRR) and Hits@5, both of which are widely used metrics for ranking tasks [36, 45]. We repeat the experiments five times with random seeds and report the mean as well as the standard deviation.

5.1.4 Training Configurations. We run all experiments on a single machine with Intel(R) Core(TM) i9-10980XE 3.00GHz CPUs, NVIDIA RTX A6000, and 48 GB RAM memory. We run 8 baselines using their official codes. We set the batch size B as 128 and the band count as 3. We train the BandRank with Adam optimizer [17], with an empirical learning rate of 0.0001.

5.2 Effectiveness Evaluation

We evaluate the ranking performance of our BandRank across two learning scenarios under two metrics against eight baselines.

Exp-1: Temporal Interaction Ranking in Transductive and Inductive Settings. As shown in Table 2, we observe that (1) BandRank significantly outperforms the eight baselines in nearly all cases, achieving an average improvement of 21.24% in the transductive setting and 21.38% in the inductive setting compared to the best competitors. These results demonstrate the effectiveness of our band-pass disentangled representation learning approach and the Harmonic Ranking loss for ranking tasks over dynamic graphs. (2) Our BandRank outperforms the ranking-driven approaches (TGRank and TATKC) by an average of 30.85% in MRR and 16.24% in Hits@5. This improvement is attributed to the limitations of existing ranking-driven approaches, which employ a single listwise loss for ranking, making them suffer from the gradient vanishing issue that degrades model performance. Additionally, these approaches model multi-scale temporal patterns in a shared latent space, causing the frequency aliasing issue. TATKC, in particular, emphasizes approximating temporal Katz centrality ranking, resulting in high computational complexity and limited scalability. (3) Classification-driven approaches consistently show inferior performance compared to ranking-driven ones, highlighting the priority of ranking-driven mechanisms in achieving superior performance on ranking tasks over dynamic graphs. (4) Listwise-based approaches (e.g., BandRank, TGRank, and TATKC-LW) generally outperform pointwise-based ones, confirming the importance of jointly optimizing candidate sets for the ranking task. (5) Our BandRank outperforms FreeDyG by an average of 59.89% in MRR and 43.91% in Hit@5, suggesting that FreeDyG's adaptive filter in the frequency domain is insufficient for capturing multi-scale graph patterns accurately. FreeDyG is a link prediction approach that cannot jointly optimize candidate sets, further compromising ranking performance. In contrast, BandRank disentangles complex temporal structures into multiple frequency bands, allowing for accurate multi-scale learning and thus leading to superior performance.

5.3 Robustness Evaluation

To evaluate the robustness of our BandRank, we conduct experiments on two key learning scenarios: **Random Dropout Iterations** and **Different-level Noise** in Figs. 3 and 4 compared to the ranking-driven TGRank. We also conduct two variants of our BandRank to investigate their contribution to the model robustness, called 'w/o Distangle' and 'w/o HR loss', where we remove our frequency disentanglement and Harmonic loss term, respectively.

Exp-2: Robustness against Dropout Iterations. We plot virtual candidate score distributions over 100 dropout iterations for four

Table 2: Transductive and inductive temporal interaction ranking performance in MRR (mean ± std.) and Hits@5 (mean ± std.). TLE signifies that a 12-hour time limit exceeded during link ranking inference. * denotes results collected from [36]. The best and second-best results are marked in bold and underline, respectively.

Dataset	Reddit		Wikipedia		MOOC		LastFM		Enron		UCI	
Metric	MRR	Hits@5	MRR	Hits@5	MRR	Hits@5	MRR	Hits@5	MRR	Hits@5	MRR	Hits@5
Transductive												
DyRep*	0.519 ± 0.007	0.642 ± 0.011	0.473 ± 0.013	0.606 ± 0.009	0.122 ± 0.016	0.172 ± 0.013	0.065 ± 0.018	0.084 ± 0.010	0.200 ± 0.013	0.297 ± 0.018	0.029 ± 0.009	0.027 ± 0.010
JODIE*	0.376 ± 0.003	0.539 ± 0.008	0.287 ± 0.006	0.514 ± 0.008	0.110 ± 0.013	0.156 ± 0.020	0.031 ± 0.012	0.034 ± 0.016	0.104 ± 0.018	0.126 ± 0.021	0.023 ± 0.008	0.019 ± 0.007
TGAT*	0.413 ± 0.006	0.536 ± 0.007	0.497 ± 0.012	0.737 ± 0.006	0.188 ± 0.009	0.264 ± 0.008	0.098 ± 0.011	0.122 ± 0.014	0.252 ± 0.007	0.424 ± 0.012	0.198 ± 0.004	0.288 ± 0.005
TGN*	0.552 ± 0.004	0.678 ± 0.010	0.575 ± 0.015	0.795 ± 0.013	0.208 ± 0.012	0.329 ± 0.011	0.089 ± 0.008	0.095 ± 0.013	0.242 ± 0.008	0.397 ± 0.009	0.203 ± 0.001	0.280 ± 0.006
CAWN*	TLE	TLE	0.795 ± 0.022	0.868 ± 0.018	0.206 ± 0.003	0.265 ± 0.016	TLE	TLE	0.386 ± 0.011	0.614 ± 0.003	0.518 ± 0.005	0.683 ± 0.008
FreeDyG	0.325 ± 0.025	0.526 ± 0.004	0.616 ± 0.008	0.782 ± 0.003	0.057 ± 0.003	0.051 ± 0.009	TLE	TLE	0.254 ± 0.012	0.420 ± 0.013	0.322 ± 0.022	0.544 ± 0.014
TGRank*	0.663 ± 0.005	0.821 ± 0.009	0.792 ± 0.006	0.870 ± 0.001	0.321 ± 0.001	0.436 ± 0.002	0.165 ± 0.009	0.230 ± 0.010	0.414 ± 0.007	0.650 ± 0.006	0.685 ± 0.002	**0.781 ± 0.004**
TATKC	TLE	TLE	0.189 ± 0.036	0.247 ± 0.037	0.028 ± 0.009	0.026 ± 0.008	TLE	TLE	0.191 ± 0.035	0.327 ± 0.074	0.039 ± 0.010	0.042 ± 0.014
TATKC-LW	TLE	TLE	0.781 ± 0.008	0.901 ± 0.002	0.182 ± 0.007	0.303 ± 0.015	TLE	TLE	0.533 ± 0.023	0.653 ± 0.018	0.381 ± 0.004	0.584 ± 0.011
BandRank	**0.853 ± 0.111**	**0.921 ± 0.045**	**0.931 ± 0.001**	**0.944 ± 0.002**	**0.550 ± 0.010**	**0.536 ± 0.010**	**0.567 ± 0.008**	**0.557 ± 0.008**	**0.570 ± 0.005**	**0.685 ± 0.004**	**0.739 ± 0.002**	0.738 ± 0.002
Inductive												
DyRep*	0.388 ± 0.008	0.528 ± 0.007	0.474 ± 0.004	0.609 ± 0.007	0.125 ± 0.014	0.173 ± 0.008	0.112 ± 0.015	0.152 ± 0.009	0.137 ± 0.014	0.200 ± 0.016	0.035 ± 0.015	0.040 ± 0.008
JODIE*	0.042 ± 0.010	0.048 ± 0.009	0.264 ± 0.009	0.459 ± 0.008	0.069 ± 0.012	0.085 ± 0.013	0.055 ± 0.012	0.061 ± 0.017	0.119 ± 0.010	0.162 ± 0.009	0.029 ± 0.009	0.028 ± 0.011
TGAT*	0.374 ± 0.013	0.493 ± 0.013	0.523 ± 0.007	0.746 ± 0.010	0.169 ± 0.004	0.230 ± 0.007	0.117 ± 0.013	0.134 ± 0.011	0.201 ± 0.008	0.360 ± 0.011	0.221 ± 0.007	0.321 ± 0.004
TGN*	0.537 ± 0.009	0.650 ± 0.008	0.604 ± 0.007	0.786 ± 0.006	0.228 ± 0.005	0.355 ± 0.010	0.120 ± 0.010	0.126 ± 0.021	0.219 ± 0.005	0.376 ± 0.008	0.230 ± 0.006	0.315 ± 0.004
CAWN*	TLE	TLE	0.798 ± 0.008	0.847 ± 0.007	0.200 ± 0.010	0.261 ± 0.008	TLE	TLE	0.395 ± 0.013	0.609 ± 0.007	0.521 ± 0.004	0.681 ± 0.009
FreeDyG	0.321 ± 0.016	0.515 ± 0.016	0.627 ± 0.009	0.781 ± 0.004	0.056 ± 0.003	0.049 ± 0.008	TLE	TLE	0.222 ± 0.004	0.377 ± 0.017	0.325 ± 0.021	0.544 ± 0.013
TGRank*	0.608 ± 0.004	0.761 ± 0.008	0.797 ± 0.008	0.866 ± 0.009	0.299 ± 0.002	0.403 ± 0.005	0.194 ± 0.008	0.273 ± 0.005	0.432 ± 0.008	0.648 ± 0.006	0.684 ± 0.003	**0.782 ± 0.002**
TATKC	TLE	TLE	0.183 ± 0.039	0.238 ± 0.040	0.028 ± 0.008	0.026 ± 0.007	TLE	TLE	0.173 ± 0.032	0.288 ± 0.060	0.044 ± 0.012	0.049 ± 0.016
TATKC-LW	TLE	TLE	0.784 ± 0.008	0.896 ± 0.007	0.177 ± 0.003	0.291 ± 0.005	TLE	TLE	0.553 ± 0.032	0.687 ± 0.009	0.379 ± 0.002	0.583 ± 0.012
BandRank	**0.826 ± 0.130**	**0.905 ± 0.054**	**0.919 ± 0.002**	**0.937 ± 0.002**	**0.541 ± 0.002**	**0.530 ± 0.001**	**0.568 ± 0.006**	**0.539 ± 0.007**	**0.607 ± 0.006**	**0.713 ± 0.002**	**0.733 ± 0.001**	0.735 ± 0.002

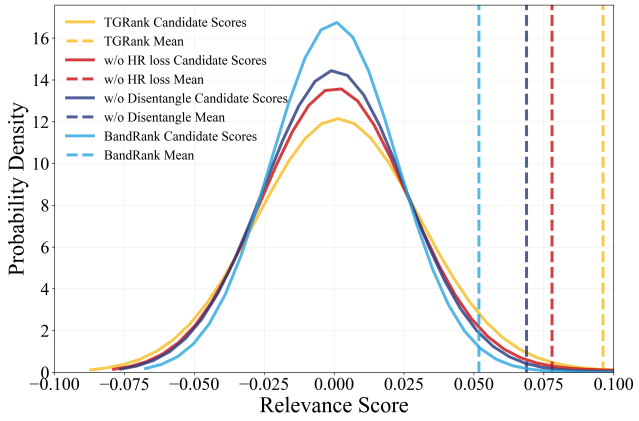
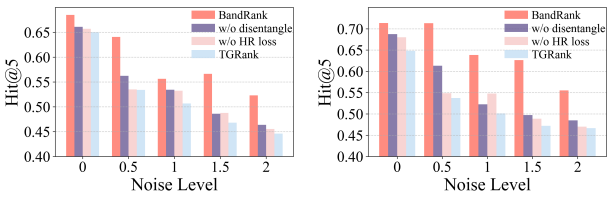


Figure 3: Comparison of virtual candidate score distribution across 100 dropout iterations. Narrower and taller distributions indicate stronger robustness.

comparative approaches in Fig. 3, using kernel density estimation (KDE) [29]. Our BandRank exhibits the most concentrated distribution and the least mean score compared to TGRank and its two variants. This suggests the best model robustness, as it consistently produces stable results despite the stochastic nature of dropout. In contrast, TGRank demonstrates the widest and flattest distribution and the largest mean score, which implies that its candidate scores fluctuate significantly across dropout iterations, revealing weaker robustness. Between our two variants, 'w/o HR loss' has a wider distribution and relatively greater mean score, further confirming that our harmonic term plays a crucial role in improving robustness under random dropout conditions. This is aligned with Theorem 2. Overall, these findings validate the robustness of BandRank, showing it is less sensitive to training randomness and more reliable in generating stable prediction across different conditions.



(a) Transductive setting (b) Inductive setting

Figure 4: Comparison of Hit@5 performance under different noise levels on Enron dataset.

Exp-3: Robustness against Different-level Noise. We present the results of four comparative approaches under varying noise levels by adding Gaussian noise of different intensities to the dynamic graphs. As shown in Fig. 4, the performance of TGRank decreases rapidly with increasing noise. In contrast, our BandRank demonstrates relatively stable performance. Notably, BandRank without the frequency disentanglement ('w/o disentangle') shows reduced noise robustness compared to the full BandRank model but still outperforms TGRank, as it can partially mitigate noise during complex structure learning in the frequency domain. As noise levels increase, the performance gap between BandRank and the other three comparative approaches widens by up to 24.66% in Hit@5, highlighting the superior noise robustness of our band-pass disentangled representation learning.

5.4 Ablation Study

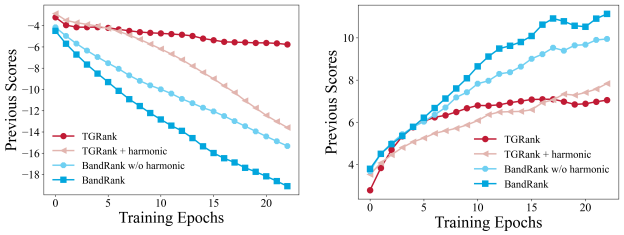
We investigate the contribution of different modules within our BandRank by evaluating its performance when each module is removed individually. The following analysis is conducted from two key perspectives: ranking accuracy and training robustness.

Exp-4: Ranking Accuracy Perspective. We conduct three variants to assess their impact on ranking performance across three datasets in two learning scenarios. Specifically, we evaluate: (1)

Table 3: Ablation study on three datasets under MRR metric in transductive and inductive settings.

Setting	Transductive MRR			Inductive MRR		
Dataset	UCI	Enron	Wikipedia	UCI	Enron	Wikipedia
BandRank	**0.739**	**0.570**	**0.931**	**0.733**	**0.607**	**0.919**
w/o disentangle	0.715	0.524	0.916	0.708	0.551	0.906
w/o HR Loss	0.703	0.508	0.907	0.695	0.536	0.902
w/o Frequency	0.688	0.544	0.892	0.681	0.574	0.871

removing the frequency disentanglement ('w/o disentangle'), (2) removing the harmonic term ('w/o HR Loss'), and (3) replacing the BPDR approach with a time-domain MLP ('w/o Frequency'). As shown in Table 3, we observe that (1) The frequency disentanglement enhances ranking performance by an average of 4.50% in terms of MRR, as it alleviates the frequency aliasing issue during structure learning and accurately models multi-scale temporal patterns. (2) The 'w/o HR Loss' variant performs worse than BandRank, with an average decrease of 6.18%. This highlights the importance of the harmonic term, which integrates a hard comparison between real and virtual candidates, thereby mitigating the gradient vanishing issue. Additionally, 'w/o HR Loss' has a more significant impact on overall model performance than 'w/o disentangle' in both transductive and inductive learning scenarios since it can enhance model discriminative ability. (3) Frequency-enhanced structure learning in our BPDR approach, is well-suited for structure learning over dynamic graphs, improving ranking performance by up to 7.09%. This improvement is attributed to the ability to project complex dynamics into distinct frequency bands, enabling the model to selectively learn multi-scale dynamics and structure evolution.



(a) Virtual candidate scores (b) Real candidate scores

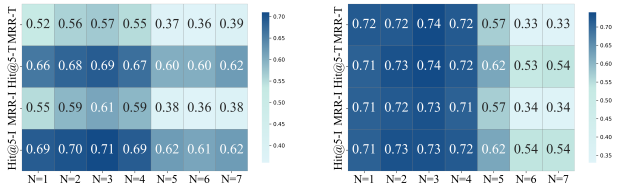
Figure 5: Comparison of previous scores during training on Enron dataset.

Exp-5: Training Robustness Perspective. To evaluate training robustness, we plot the previous score curves for real and virtual candidates across increasing training epochs on the Enron dataset. Here, we introduce a variant of BandRank by removing the harmonic term (called 'BandRank w/o harmonic'). As shown in Fig. 5, BandRank with the harmonic term illustrates a sharper slope in the candidate score curves of both real and virtual candidates, indicating that our HR Loss is more effective at distinguishing between real and virtual candidates compared to the listwise loss. This suggests that our HR Loss can facilitate more robust model training in the ranking task, thereby enhancing overall ranking performance as evidenced in Table 3.

Exp-6: Effectiveness of Our Harmonic Term. We also plug our harmonic term into the listwise loss of the existing TGRank model (referred to as 'TGRank + harmonic') and plot the candidate score curves on the Enron dataset in Fig. 5. We can observe from Fig. 5a that the harmonic term effectively mitigates the gradient vanishing issue for virtual candidates. Additionally, it enhances the learning ability of real candidates, with the scores of real candidates in 'TGRank + harmonic' surpassing those of TGRank in final epochs, as illustrated in Fig. 5b. These results further demonstrate that our HR loss is well-suited for temporal interaction ranking on dynamic graphs.

5.5 Hyperparameter Sensitivity Analysis

During multi-scale learning, we disentangle all frequency components into N distinct frequency bands, where N is a tunable hyperparameter. To evaluate its effect, we conduct experiments by varying N from 1 to 7 on the Enron and UCI datasets.



(a) Enron Dataset (b) UCI Dataset

Figure 6: Effect of hyperparameter (N) on two datasets in transductive (T) and inductive (I) settings under two metrics.

Exp-7: Effect of Hyperparameter N. As shown in Fig. 6, (1) our BandRank achieves optimal performance when the frequency components are disentangled into three bands: low, mid, and high frequencies. In this configuration, BandRank effectively captures long-term evolution, periodic patterns, and abrupt fluctuations. When N is set to 2 or 4, the ranking performance slightly decreases but still surpasses the 'w/o disentangle' case ($N = 1$), confirming the effectiveness of frequency disentanglement for dynamic graph learning. However, when N exceeds 5, the ranking performance progressively deteriorates. This is likely due to over-disentangling, which can overly separate similar temporal patterns into multiple frequency bands, distorting temporal integrity and thus degrading performance. Throughout this paper, we set $N = 3$.

6 CONCLUSION

In this paper, we investigate the ranking task over dynamic graphs and propose an effective and robust band-pass disentangled ranking approach, comprising two key modules: the band-pass disentangled representation learning and the harmonic ranking loss. Specifically, we propose a novel band-pass disentangled representation approach that adaptively disentangles complex temporal structures into multiple frequency bands based on our defined quality factors and band-pass filter, enabling accurate multi-scale learning for temporal structures to generate high-quality node embeddings. Additionally, we design a Harmonic Ranking loss to jointly optimize candidate sets while preventing the gradient vanishing issue. Comprehensive experimental results validate the effectiveness and robustness of our BandRank across various learning scenarios from several perspectives.

REFERENCES

[1] Georg Bruun. 1978. z-transform DFT filters and FFT's. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, 1 (1978), 56–63.

[2] Maarten Buyl, Paul Missault, and Pierre-Antoine Sondag. 2023. RankFormer: Listwise learning-to-rank using listwide labels. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3762–3773.

[3] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. 2007. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the international conference on Machine learning*. 129–136.

[4] Ke Cheng, Linzhi Peng, Junchen Ye, Leilei Sun, and Bowen Du. 2024. Co-Neighbor Encoding Schema: A Light-cost Structure Encoding Method for Dynamic Link Prediction. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 421–432.

[5] Lawrence J Christiano and Terry J Fitzgerald. 2003. The band pass filter. *International economic review* 44, 2 (2003), 435–465.

[6] Jonathan Crabbé, Nicolas Huynh, Jan Stanczuk, and Mihaela van der Schaar. 2024. Time series diffusion in the frequency domain. *arXiv preprint arXiv:2402.05933* (2024).

[7] Leyan Deng, Defu Lian, Zhenya Huang, and Enhong Chen. 2022. Graph Convolutional Adversarial Networks for Spatiotemporal Anomaly Detection. *IEEE Trans. Neural Networks Learn. Syst.* 33, 6 (2022), 2416–2428.

[8] Jufang Duan, Wei Zheng, Yangzhou Du, Wenfa Wu, Haipeng Jiang, and Hongsheng Qi. [n. d.]. MF-CLR: Multi-Frequency Contrastive Learning Representation for Time Series. In *Forty-first International Conference on Machine Learning*.

[9] Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. 2024. Tslanet: Rethinking transformers for time series representation learning. *arXiv preprint arXiv:2404.08472* (2024).

[10] Ziwei Fan, Zhiwei Liu, Jiawei Zhang, Yun Xiong, Lei Zheng, and Philip S Yu. 2021. Continuous-time sequential recommendation with temporal graph collaborative transformer. In *Proceedings of the ACM international conference on information & knowledge management*. 433–442.

[11] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems* 37, 2 (2019), 1–30.

[12] Alessio Gravina, Giulio Lovisotto, Claudio Gallicchio, Davide Bacciu, and Claas Grohnfeldt. 2024. Long Range Propagation on Continuous-Time Dynamic Graphs. In *The International Conference on Machine Learning*. OpenReview.net.

[13] Pankaj Gupta, Venu Satuluri, Ajeet Grewal, Siva Gurumurthy, Volodymyr Zhabiuk, Quannan Li, and Jimmy Lin. 2014. Real-time twitter recommendation: Online motif detection in large dynamic graphs. *Proceedings of the VLDB Endowment* 7, 13 (2014), 1379–1380.

[14] Ziwei He, Meng Yang, Minwei Feng, Jingcheng Yin, Xinbing Wang, Jingwen Leng, and Zhouhan Lin. 2023. Fourier transformer: Fast long range modeling by removing sequence redundancy with fft operator. *arXiv preprint arXiv:2305.15099* (2023).

[15] Alexander Hermans, Lucas Beyer, and Bastian Leibe. 2017. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737* (2017).

[16] Jiazhen Ji, Huan Wang, Yuge Huang, Jiaxiang Wu, Xingkun Xu, Shouhong Ding, ShengChuan Zhang, Liujuan Cao, and Rongrong Ji. 2022. Privacy-preserving face recognition with learnable privacy budgets in frequency domain. In *European Conference on Computer Vision*. Springer, 475–491.

[17] Diederik P Kingma. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[18] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. 2023. Efficient Frequency Domain-Based Transformers for High-Quality Image Deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5886–5895.

[19] Srijan Kumar, Xikun Zhang, and Jure Leskovec. 2019. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery & data mining*. 1269–1278.

[20] Jérôme Kunegis. 2013. Konect: the koblenz network collection. In *Proceedings of the international conference on world wide web*. 1343–1350.

[21] Peter Laflin, Alexander V Mantzaris, Fiona Ainley, Amanda Otley, Peter Grindrod, and Desmond J Higham. 2013. Discovering and validating influence in a dynamic online social network. *Social Network Analysis and Mining* 3 (2013), 1311–1323.

[22] Haoyang Li and Lei Chen. 2021. Cache-based GNN System for Dynamic Graphs. In *The ACM International Conference on Information and Knowledge Management*. ACM, 937–946.

[23] Yiming Li, Yanyan Shen, Lei Chen, and Mingxuan Yuan. 2023. Orca: Scalable Temporal Graph Neural Network Training with Theoretical Guarantees. *Proc. ACM Manag. Data* 1, 1 (2023), 52:1–52:27.

[24] Yiming Li, Yanyan Shen, Lei Chen, and Mingxuan Yuan. 2023. Zebra: When temporal graph neural networks meet temporal personalized PageRank. *Proceedings of the VLDB Endowment* 16, 6 (2023), 1332–1345.

[25] Yuyang Long, Qilong Zhang, Boheng Zeng, Lianli Gao, Xianglong Liu, Jian Zhang, and Jingkuan Song. 2022. Frequency domain model augmentation for adversarial attack. In *European conference on computer vision*. Springer, 549–566.

[26] Tianyu Luan, Yuanhao Zhai, Jingjing Meng, Zhong Li, Zhang Chen, Yi Xu, and Junsong Yuan. 2023. High Fidelity 3D Hand Shape Reconstruction via Scalable Graph Frequency Decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16795–16804.

[27] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z. Sheng, Hui Xiong, and Leman Akoglu. 2023. A Comprehensive Survey on Graph Anomaly Detection With Deep Learning. *IEEE Trans. Knowl. Data Eng.* 35, 12 (2023), 12012–12038.

[28] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao Schardl, and Charles Leiserson. 2020. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 5363–5370.

[29] Emanuel Parzen. 1962. On estimation of a probability density function and mode. *The annals of mathematical statistics* 33, 3 (1962), 1065–1076.

[30] Farimah Poursafaei, Shenyang Huang, Kellin Pelrine, and Reihaneh Rabbany. 2022. Towards better evaluation for dynamic link prediction. *Advances in Neural Information Processing Systems* 35 (2022), 32928–32941.

[31] Hao Qian, Hongting Zhou, Qian Zhao, Hao Chen, Hongxiang Yao, Jingwei Wang, Ziqi Liu, Fei Yu, Zhiqiang Zhang, and Jun Zhou. 2024. MDGNN: Multi-Relational Dynamic Graph Neural Network for Comprehensive and Dynamic Stock Investment Prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 14642–14650.

[32] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. 2020. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637* (2020).

[33] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. 2013. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine* 30, 3 (2013), 83–98.

[34] Weiping Song, Zhiping Xiao, Yifan Wang, Laurent Charlin, Ming Zhang, and Jian Tang. 2019. Session-based social recommendation via dynamic graph attention networks. In *Proceedings of the ACM international conference on web search and data mining*. 555–563.

[35] Junwei Su, Difan Zou, and Chuan Wu. 2024. PRES: Toward Scalable Memory-Based Dynamic Graph Neural Networks. In *The International Conference on Learning Representations*. OpenReview.net.

[36] Susheel Suresh, Mayank Shrivastava, Arko Mukherjee, Jennifer Neville, and Pan Li. 2023. Expressive and Efficient Representation Learning for Ranking Links in Temporal Graphs. In *Proceedings of the ACM Web Conference 2023*. 567–577.

[37] Yuxing Tian, Yiyan Qi, and Fan Guo. 2024. FreeDyG: Frequency Enhanced Continuous-Time Dynamic Graph Model for Link Prediction. In *The International Conference on Learning Representations*.

[38] Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. 2019. Dyrep: Learning representations over dynamic graphs. In *International conference on learning representations*.

[39] Jingyuan Wang, Ze Wang, Jianfeng Li, and Junjie Wu. 2018. Multilevel wavelet decomposition network for interpretable time series analysis. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2437–2446.

[40] Xuhong Wang, Ding Lyu, Mengjian Li, Yang Xia, Qi Yang, Xinwen Wang, Xinguang Wang, Ping Cui, Yupu Yang, Bowen Sun, et al. 2021. Apan: Asynchronous propagation attention network for real-time temporal graph embedding. In *Proceedings of the international conference on management of data*. 2628–2638.

[41] Yanbang Wang, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, and Pan Li. 2021. Inductive representation learning in temporal networks via causal anonymous walks. *arXiv preprint arXiv:2101.05974* (2021).

[42] Yanbang Wang, Pan Li, Chongyang Bai, and Jure Leskovec. 2021. Tedic: Neural modeling of behavioral patterns in dynamic social interaction networks. In *Proceedings of the Web Conference 2021*. 693–705.

[43] Zhihao Wen and Yuan Fang. 2022. TREND: TempoRal Event and Node Dynamics for Graph Representation Learning. In *Proceedings of the ACM Web Conference 2022*. ACM, 1159–1169.

[44] Yuxia Wu, Yuan Fang, and Lizi Liao. 2024. On the Feasibility of Simple Transformer for Dynamic Graph Modeling. In *Proceedings of the ACM on Web Conference 2024*. ACM, 870–880.

[45] Yiqing Wu, Ruobing Xie, Zhao Zhang, Xu Zhang, Fuzhen Zhuang, Leyu Lin, Zhanhui Kang, and Yongjun Xu. 2024. DFGNN: Dual-frequency Graph Neural Network for Sign-aware Feedback. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3437–3447.

[46] Huiru Xiao, Xin Liu, Yangqiu Song, Ginny Y Wong, and Simon See. 2022. Complex hyperbolic knowledge graph embeddings with fast fourier transform. *arXiv preprint arXiv:2211.03635* (2022).

[47] Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. 2020. Inductive representation learning on temporal graphs. *arXiv preprint arXiv:2002.07962* (2020).

[48] Ronghui Xu, Hao Miao, Senzhang Wang, Philip S. Yu, and Jianxin Wang. 2024. PeFAD: A Parameter-Efficient Federated Framework for Time Series Anomaly

Detection. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 3621–3632.
[49] Yang Xu and David Reitter. 2017. Spectral analysis of information density in dialogue predicts collaborative task performance. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 623–633.
[50] Yuanyuan Xu, Wenjie Zhang, Xiwei Xu, Binghao Li, and Ying Zhang. 2024. Scalable and Effective Temporal Graph Representation Learning With Hyperbolic Geometry. *IEEE Transactions on Neural Networks and Learning Systems* (2024).
[51] Yuanyuan Xu, Wenjie Zhang, Ying Zhang, Maria Orlowska, and Xuemin Lin. 2024. TimeSGN: Scalable and Effective Temporal Graph Neural Network. In *Proc. IEEE Int. Conf. Data Eng.*
[52] Zhangjing Yang, Weiwu Yan, Xiaolin Huang, and Lin Mei. 2022. Adaptive Temporal-Frequency Network for Time-Series Forecasting. *IEEE Trans. Knowl. Data Eng.* 34, 4 (2022), 1576–1587.
[53] Zuhao Yang, Yingfang Yuan, Yang Xu, Shuo Zhan, Huajun Bai, and Kefan Chen. 2024. Face: Evaluating natural language generation with fourier analysis of cross-entropy. *Advances in Neural Information Processing Systems* 36 (2024).
[54] Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. 2024. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. *Advances in Neural Information Processing Systems* 36 (2024).
[55] Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. 2023. Frequency-domain MLPs are More Effective Learners in Time Series Forecasting. In *Advances in Neural Information Processing Systems*.
[56] Minji Yoon, Bryan Hooi, Kijung Shin, and Christos Faloutsos. 2019. Fast and accurate anomaly detection in dynamic graphs with a two-pronged approach. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 647–657.
[57] Haonan Yuan, Qingyun Sun, Xingcheng Fu, Cheng Ji, and Jianxin Li. 2024. Dynamic Graph Information Bottleneck. In *Proceedings of the ACM on Web Conference 2024*. ACM, 469–480.
[58] Daniele Zambon, Cesare Alippi, and Lorenzo Livi. 2018. Concept Drift and Anomaly Detection in Graph Streams. *IEEE Trans. Neural Networks Learn. Syst.* 29, 11 (2018), 5592–5605.
[59] Guozhen Zhang, Tian Ye, Depeng Jin, and Yong Li. 2023. An Attentional Multi-scale Co-evolving Model for Dynamic Link Prediction. In *Proceedings of the ACM Web Conference 2023*. ACM, 429–437.
[60] Liheng Zhang, Charu Aggarwal, and Guo-Jun Qi. 2017. Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*. 2141–2149.
[61] Mengqi Zhang, Shu Wu, Xueli Yu, Qiang Liu, and Liang Wang. 2022. Dynamic graph neural networks for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2022), 4741–4753.
[62] Tianming Zhang, Junkai Fang, Zhengyi Yang, Bin Cao, and Jing Fan. 2024. TATKC: A Temporal Graph Neural Network for Fast Approximate Temporal Katz Centrality Ranking. In *Proceedings of the ACM on Web Conference 2024*. 527–538.
[63] Yongjian Zhong, Hieu Vu, Tianbao Yang, and Bijaya Adhikari. 2024. Efficient and Effective Implicit Dynamic Graph Neural Network. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 4595–4606.
[64] Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al. 2022. Film: Frequency improved legendre memory model for long-term time series forecasting. *Advances in neural information processing systems* 35 (2022), 12677–12690.
[65] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*. PMLR, 27268–27286.
[66] Tao Zou, Yuhao Mao, Junchen Ye, and Bowen Du. 2024. Repeat-Aware Neighbor Sampling for Dynamic Graph Learning. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 4722–4733.

A REPRODUCIBILITY

The source codes of the proposed BandRank and eight selected baselines are available at

https://anonymous.4open.science/r/BandRank-61E7.

B PROOF

B.1 Proof of Theorem 1

Given the time-domain representation of the temporal messages $\mathbf{H}(t)$ at time t, with its frequency components represented by $\mathcal{H}(f)$. We first prove the global convolution theorem for the case where the disentanglement parameter $N = 1$. This case serves as the foundation for Theorem 1, which extends the result to $N \geq 2$.

Case 1: $N = 1$. In this case, we can consider the process as disentangling the frequency components into a single frequency band. We aim to prove that our band-pass disentangled representation approach in the frequency domain is equivalent to a global convolution in the time domain based on the theorem 2 in [55]. Concretely, suppose that we conduct Fourier transform $\mathcal{F}$ in the d-dimension, then we have

$$\mathcal{F}(\mathbf{H}(t) * \mathbf{C}(t)) = \int_{-\infty}^{\infty} (\mathbf{H}(t) * \mathbf{C}(t)) e^{-i2\pi f t} dt. \tag{14}$$

According to the convolution theorem, $\mathbf{H}(t) * \mathbf{C}(t) = \int_{-\infty}^{\infty} \mathbf{H}(\tau)\mathbf{C}(t-\tau)d\tau$, we have

$$\mathcal{F}(\mathbf{H}(t) * \mathbf{C}(t)) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \mathbf{H}(\tau)\mathbf{C}(t-\tau)d\tau \right) e^{-i2\pi f t} dt, \tag{15}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{C}(t-\tau) e^{-i2\pi f t} dt \mathbf{H}(\tau) d\tau. \tag{16}$$

Next, we perform a change of variables $x = t - \tau$:

$$\mathcal{F}(\mathbf{H}(t) * \mathbf{C}(t)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{C}(x) e^{-i2\pi f (x+\tau)} dx \mathbf{H}(\tau) d\tau, \tag{17}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{C}(x) e^{-i2\pi f x} e^{-i2\pi f \tau} dx \mathbf{H}(\tau) d\tau, \tag{18}$$

$$= \int_{-\infty}^{\infty} \mathbf{H}(\tau) e^{-i2\pi f \tau} d\tau \int_{-\infty}^{\infty} \mathbf{C}(x) e^{-i2\pi f x} dx, \tag{19}$$

$$= \mathcal{H}(f) \odot \mathbf{C}(f). \tag{20}$$

This indicates that convolving $\mathbf{H}(t)$ and $\mathbf{C}(t)$ in the time domain is equivalent to multiplying their Fourier transforms in the frequency domain. Therefore, in the case where $N = 1$, the frequency-domain learning process can be viewed as applying a global convolution in the time domain, that is, $\text{FreMLP}(\mathcal{H}(f)) = \mathcal{F}(\mathbf{H}(t) * \mathbf{C}(t)) = \mathcal{H}(f) \odot \mathbf{C}(f)$.

Case 2: $N \geq 2$. Based on Case 1, we proceed to prove that our band-pass representation approach is equivariant to the multi-scale global convolution in the time domain when $N \geq 2$. Concretely, taking a single-layer network ($l = 1$) as an example, we disentangle the frequency signal $\mathcal{H}(f)$ into N distinct and non-overlapping frequency bands, which is done by the spectral mask. For each frequency band $\mathcal{H}_k(f)$, we leverage frequency-enhanced MLP to generate band-specific embeddings $\mathcal{Z}_k(f)$, which can formulated as

$$\mathcal{H}_k(f) = \mathbb{I}_k \odot \mathcal{H}(f), \tag{21}$$

$$\mathcal{Z}_k(f) = \text{FreMLP}_k(\mathcal{H}_k(f)) \tag{22}$$

$$= \mathcal{H}_k(f) \odot \mathbf{C}_k(f), \tag{23}$$

$$= \mathbb{I}_k \odot \mathcal{H}(f) \odot \mathbf{C}_k(f), \tag{24}$$

where $\mathcal{H}_k(f)$ represents the frequency-domain representation of the k-th frequency band at layer l, $\mathbf{C}_k(f)$ is the corresponding complex number weight in the frequency domain. The filter indicators $\mathbb{I}_k$ in Eq. (21) can adaptively isolate the frequency components based on quality factors defined in Eq. (5). Since the frequency masks are mutually exclusive, the operations performed on each

frequency band are independent. Thus, for the k-th frequency band, we perform global convolution in the frequency domain based on Eqs. 21- 24. That means we perform N independent global convolutions over the frequency spectrum. The final output of our band-pass disentangled representation approach in the frequency domain is the sum of the contributions from all N frequency bands based on Eq. (25), which can be represented as:

$$\mathcal{Z}(f) = \sum_{k=1}^{N} \mathbb{I}_k \odot \mathcal{H}(f) \odot \mathbf{C}_k(f), \tag{25}$$

where $\mathcal{Z}^l(f)$ is the multi-scale convolution result. Overall, the proof is completed.

B.2 Proof of Theorem 2

Given a query $q = (c, \mathbb{D}, t)$ and its scores of all the candidates in the subgraph $\{y_j\}_{j=1}^{|\mathbb{D}(t)|}$, which includes real candidate score y_{r} and virtual candidate scores $\boldsymbol{y}_{\mathbb{D}\setminus v_r}$, we first present the loss functions of listwise and our harmonic, which are formulated as

$$\mathcal{L}_{\text{listwise}} = -\log\left(\frac{\exp(y_{\mathrm{r}})}{\sum_{j=1}^{|\mathbb{D}(t)|}\exp(y_j)}\right), \tag{26}$$

$$\mathcal{L}_{\text{HR}} = -\log\left(\frac{\exp(y_{\mathrm{r}})}{\sum_{j=1}^{|\mathbb{D}(t)|}\exp(y_j)}\right) - \beta\log\left(\frac{1}{1+\exp(-(y_r - y_n^{\text{hard}}))}\right). \tag{27}$$

Then we compute the gradients of our harmonic ranking loss and listwise loss. Here, the gradients involve two cases: a hard virtual candidate and normal virtual candidates, which are discussed respectively. For the hard virtual candidate score y_n, we obtain the gradient of listwise loss as

$$\frac{\partial \mathcal{L}_{\text{listwise}}}{\partial y_n} = \frac{\exp(y_n)}{\sum_{j=1}^{|\mathbb{D}(t)|}\exp(y_j)}. \tag{28}$$

Then we compute the gradient of our harmonic ranking loss $\mathcal{L}_{\text{HR}}$. For the hard candidate score y_n, we can achieve its gradient as:

$$\frac{\partial \mathcal{L}_{\text{HR}}}{\partial y_n} = \frac{\exp(y_n)}{\sum_{j=1}^{|\mathbb{D}(t)|}\exp(y_j)} + \beta\frac{1}{1+\exp(y_{\mathrm{r}} - y_n)}. \tag{29}$$

To compare its robustness in learning virtual candidates, we compute the difference between their gradients, denoted as ΔGradient:

$$\Delta\text{Gradient} = \frac{\partial \mathcal{L}_{\text{HR}}}{\partial y_n} - \frac{\partial \mathcal{L}_{\text{listwise}}}{\partial y_n}, \tag{30}$$

$$= \beta\frac{1}{1+\exp(y_{\mathrm{r}} - y_n)}, \tag{31}$$

$$> 0, \tag{32}$$

where v_r and $v_n \in \mathbb{D}(t)$. For normal virtual candidates $v_n \in \mathbb{D} \setminus \{v_r, v_n^{\text{hard}}\}$, the gradient of our HR Loss is equivalent to the original listwise loss. Thus, we can conclude that

$$\frac{\partial \mathcal{L}_{\text{HR}}}{\partial \boldsymbol{y}_{\mathbb{D}\setminus v_r}} > \frac{\partial \mathcal{L}_{\text{listwise}}}{\partial \boldsymbol{y}_{\mathbb{D}\setminus v_r}}. \tag{33}$$

Overall, the proof of Theorem 2 is completed.

C ADDITIONAL EXPERIMENTAL SETTINGS

C.1 Datasets Descriptions

We present the detailed descriptions for the datasets used in our experiments below.

- **Reddit** is a bipartite interaction graph spanning 30 days, involving users and subreddits. Here, an edge is added whenever a user interacts with a subreddit. Each edge is annotated with continuous timestamps and textual attributes derived from the user's posts.
- **Wikipedia** is a bipartite interaction graph spanning 30 days, where users interact with wiki pages by editing them. These interactions include continuous timestamps and textual attributes related to the edits.
- **MOOC** is a bipartite interaction graph that covers 17 months, representing user actions on an online education platform. Here, nodes correspond to students and course content. Edges are created based on various interactions between students and content, which contain continuous timestamps.
- **LastFM** is a bipartite interaction graph that covers 1 month where edges represent user interactions with songs on a music platform. These edges capture user-song interactions with continuous timestamps but do not include any additional attributes.
- **Enron** is a communication network based on email exchanges between employees over 3 years. Here, edges represent email interactions annotated with timestamps but without any additional attributes.
- **UCI** is a communication network formed by message exchanges in an online student community over 196 days, similar to the Enron dataset but on a different scale.

C.2 Baselines

We compare our proposed BandRank against eight baselines include:

- **JODIE** [19] employs dual recurrent neural networks (RNNs) to update the representations of both nodes involved in each edge interaction, incorporating node prediction into the training process. It is designed for the link prediction task.
- **DyRep** [38] focuses on learning dynamic node embeddings with the self-attention mechanism to handle node and edge-level events in evolving graphs. It is designed for the link prediction task.
- **TGAT** [47] leverages the vanilla self-attention mechanism for dynamic graph modeling, together with a time encoding function. It is designed for the link prediction task.
- **TGN** [32] is a general learning framework from the dynamic graph, including two main modules: memory module and message passing. It is designed for the link prediction task.
- **CAWN** [41] extracts causal anonymous walks by temporal random walks and then encodes them to generate temporal embeddings. It is designed for the link prediction task.
- **FreeDyG** [37] is a frequency-enhanced dynamic graph model that incorporates frequency domain processing to

detect periodic patterns and abrupt events. It is designed for the link prediction task.

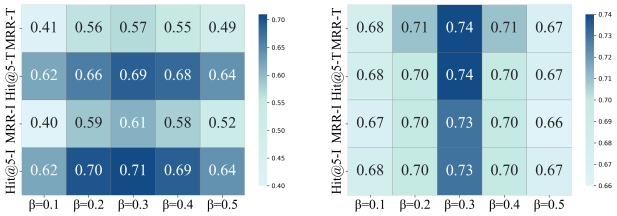
- **TGRank** [36] addresses ranking tasks by optimizing the relative order of candidate nodes through listwise ranking loss. It is designed for the subgraph-level ranking task.
- **TATKC** [62] focuses on modeling hierarchical structures in temporal graphs, employing attention mechanisms to capture multi-level dependencies. It is designed for the graph-level ranking task based on the temporal Katz centrality computation.

C.3 Additional Implementation Details

We set the embedding dimension to 128 for all datasets and fix the time encoding dimension at 128. The number of frequency bands is set to 3. We perform a grid search over the encoder layers, selecting from $\{1, 2, 3\}$. The dropout ratio is set to 0.1. We apply the HR Loss function with a parameter $\beta = 0.3$. Additionally, we incorporate layer normalization within the UPDATE module to ensure training stability.

D ADDITIONAL EXPERIMENTAL RESULTS

In our proposed Harmonic Ranking (HR) loss function, the hyperparameter β plays a crucial role in balancing the contributions of the listwise loss and the harmonic term, as defined in Eq. (12). Specifically, β controls the trade-off between encouraging the model to distinguish real candidates from virtual ones (via the harmonic term) and maintaining overall ranking performance (via the listwise loss). To analyze the effect of β, we conduct experiments on both the Enron and UCI datasets, varying β from 0.1 to 0.5.



(a) Enron Dataset **(b) UCI Dataset**

Figure 7: Effect of hyperparameter β on two datasets in transductive (T) and inductive (I) settings under two metrics.

Exp-8: Effect of Hyperparameter β. As shown in Fig. 7, we observe that the model performance increases as β rises from 0.1 to 0.3. This improvement occurs because the harmonic term progressively provides more gradient signals to optimize the virtual candidates, effectively distinguishing between real and virtual candidates and enhancing model performance, especially in the inductive setting. However, when β is set to higher values (e.g., $\beta = 0.4$ or $\beta = 0.5$), the harmonic term's influence becomes overly dominant in the loss function, which compromises the joint optimization ability among the candidate set. In this paper, we set β to 0.3 throughout the experiments. This choice strikes the best balance between the harmonic term and the listwise loss, ensuring the model's robustness and ranking effectiveness across different datasets.