MSR: A Multifaceted Self-Retrieval Framework for Microscopic Cascade Prediction

Dongsheng Hong[1], Chao Chen[2], Xujia Li[3], Shuhui Wang[1], Wen Lin[4], Xiangwen Liao[1]*

[1] College of Computer and Data Science, Fuzhou University, China
[2] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), China
[3] Hong Kong University of Science and Technology, Hong Kong SAR, China
[4] College of Computer and Control Engineering, Minjiang University, China
{221027039, 221027214, liaoxw}@fzu.edu.cn, cha01nbox@gmail.com, leexujia@ust.hk, 2216@mju.edu.cn

Abstract

The microscopic cascade prediction task has wide applications in downstream areas like "rumor detection". Its goal is to forecast the diffusion routines of information cascade within networks. Existing works typically formulate it as a classification task, which fails to well align with the *Social Homophily* assumption, as it just use the features of "infected" users while neglecting those of "uninfected" users in representation learning. Moreover, these methods focus primarily on social relationships, thereby dismissing other vital dimensions like users' historical behavior and the underlying preferences behind it. To address these challenges, we introduce the **MSR** (Multifaceted Self-Retrieval) framework. During encoding, in addition to the existing social graph, we construct a preference graph to represent "behavioral preferences" and further propose a modified multi-channel GRAU for multi-view analysis of cascade phenomenon. For decoding, our approach diverges from classification-based methods by reformulating the task as an information retrieval problem that predicts the target user with similarity measures. Empirical evaluations on public datasets demonstrate that this framework significantly outperforms baselines on Hits@κ and MAP@κ, affirming its enhanced ability.

1 Introduction

In the digital era, the landscape of information sharing has been revolutionized (Iamnitchi et al. 2023), necessitating a nuanced understanding of "information cascade", which sheds light on how information diffuses through social networks (Bikhchandani, Hirshleifer, and Welch 1992; Jalili and Perc 2017). Cascade prediction, a vital task in this field, concentrates on delineating the trajectories of information (Gong et al. 2023; Guille et al. 2013). It is subdivided into two key levels: macroscopic and microscopic. Macroscopic cascade prediction focuses on predicting the total number of retweets over a set period (Cheng et al. 2014; Cao et al. 2017). Conversely, microscopic cascade prediction focuses on the probability of a user to retweet a tweet (Galuba et al. 2010; Qiu et al. 2016). The microscopic task offers a deeper insight on the intricacies of information, which is essential for the tasks like "rumor detection" and "viral marketing", balancing online freedom and responsibility. (Khan, Soroya, and Mahmood 2024; Griffin 2023; Zhou et al. 2021). We primarily focus on this aspect of cascade prediction.

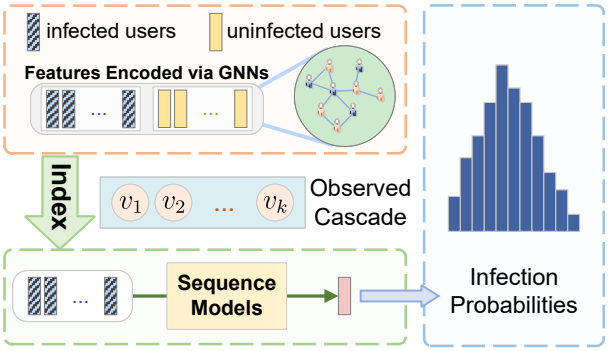


Figure 1: The encode-decode paradigm of recent methods.

The assumption of *Social Homophily*, which posits that individuals with similar characteristics or social ties tend to display analogous behaviors (McPherson, Smith-Lovin, and Cook 2001; Kossinets and Watts 2009), holds considerable significance in this domain. To leverage this assumption, recent studies have constructed social graphs and adopted Graph Neural Networks (GNNs) for prediction. For microscopic cascade prediction, given an observed cascade (users involved in the cascade are designated as "infected"), our goal is to predict its next participant from those not yet involved users (called "uninfected"). As illustrated in Figure 1, these GNN-based methods (Yuan et al. 2021; Sun et al. 2022; Wang et al. 2022; Li et al. 2023; Yunbing et al. 2023) typically consist of a three-step process: **(1)** GNNs are employed to extract user features from graphical data. **(2)** The features of "infected" users are indexed and sequentially encoded into a vector representation. **(3)** These methods commonly adopt a classification strategy, such as linear layers coupled with a softmax function, to output a probability distribution of those "uninfected" users to get "infected".

In the framework presented in Figure 1, two principal challenges are identified: **(1)** The model initially learns representations for all users. However, in the subsequent steps, only the features of "infected" users are indexed, while the features of those "uninfected" users are excluded from fur-

*Corresponding author.

ther operations, potentially disregarding the latent value of these features. **(2)** Relevant studies in other domains such as marketing highlight various factors that shaping individual decisions, such as recommendations from friends or their personal preferences. (Dietrich 2010; Acciarini, Brunetta, and Boccardelli 2021; Nakao, Ohira, and Northoff 2012; Nibbelink and Brewer 2018), emphasizing the critical need for a multi-viewpoint analysis to understand individual behaviors. However, as illustrated in Figure 1, a notable limitation is evident in the initial step of existing research. In the construction of user features, current studies tend to primarily focus on extrinsic factors, such as social relationships, represented as edges on a graph, while overlooking some other intrinsic factors, like users' personal preferences that can be derived from their historical participation records, leading to an incomplete representation.

To overcome these limitations, we propose the **MSR** (Multifaceted Self-Retrieval) framework. In this framework: **(1)** During encoding phase, inspired by research advocating for a multifaceted analysis of individual behaviors, in addition to existing social graph, we construct a preference graph from users' historical participation records to represent their "behavioral preferences" and propose a novel multi-channel GRAU module to support this multi-viewpoint analysis. This module not only enhances model's accuracy but also ensures robustness and scalability as the number of channels increases. **(2)** For decoding, diverging from classification-based approaches, we innovatively formulate this task as an information retrieval problem. Instead of merely applying linear layers to output logits for prediction, we predict the target user based on the features of "uninfected" users using similarity measures. To this end, we propose a novel self-retrieval decoder to more effectively utilize Social Homophily. Primary contributions of our work include:

- **A novel self-retrieval framework for microscopic cascade prediction**: This study introduces an innovative self-retrieval decoder to re-frame microscopic cascade prediction within a similarity-based context, enabling a deeper exploration of information cascades.
- **A multi-viewpoint analysis of users' participation**: Our research offers a multi-viewpoint analysis for users' participation, providing a more granular understanding of the participation in sharing information.
- **Empirical Validation**: Our method shows significant improvements over baseline models, with an average increase of 6.0% on Hits@κ and 4.4% on MAP@κ across three public datasets. Additionally, the ablation studies, visualizations, and parameter analysis further validate the effectiveness and robustness of each component.

2 Related Work

2.1 Cascade Prediction

Cascade prediction has been approached through three principal methodologies: feature engineering-based, generative-based and deep learning-based.

Feature engineering-based methods rely on expert knowledge to cover a wide array of aspects, including temporal dynamics (Cheng et al. 2014), attributes of users (Dow, Adamic, and Friggeri 2021; Yuan et al. 2016), and the characteristics of content (Tsagkias, Weerkamp, and De Rijke 2009). While offering insights into intricate patterns, these methods may not completely grasp the subtleties of social networks since they rely on manually curated features and extensive annotation (Zhou et al. 2021).

Generative-based approaches strive to anticipate diffusion trajectories with probabilistic data generation models (Crane and Sornette 2008; Kempe, Kleinberg, and Tardos 2003; Granovetter 1978). Despite sophisticated modeling, they are often susceptible to outliers (Mishra, Rizoiu, and Xie 2016) and usually make strong assumptions on fixed parameters that may limits their generality (Du et al. 2016).

Deep learning-based methods have advanced predictive analytics. DeepDiffuse (Islam et al. 2018) utilizes Recurrent Neural Networks (RNNs) for prediction and Topo-LSTM (Wang et al. 2017) incorporates structural data for forecasting user participation. Advancements in GNNs incorporate the social relationships, Inf-VAE (Sankar et al. 2020) and DyHGCN (Yuan et al. 2021) utilize GNNs for dynamic graph learning. MS-HGAT (Sun et al. 2022) and GRASS (Li et al. 2023) introduce hyper-graph for prediction. Despite their effectiveness, they still suffer from the neglect of "uninfected" users' features and the incomplete representation of "individual preference" in the cascade phenomenon.

2.2 GRAU & Structural Spreading

The GRASS (Li et al. 2023) model is composed of two key components: GRAU module and Structural Spreading (SS) mechanism. The GRAU module integrates the Gated Recurrent Unit (GRU) (Chung et al. 2014) with an attention mechanism (Vaswani et al. 2017) to address the "position-hopping" problem in cascade data. By expanding the receptive field, GRAU improves the capture of one-step infection relationships between non-consecutive users. The SS mechanism addresses the "branch-independency" problem by leveraging hypergraph-based features to filter relevant users (Bai, Zhang, and Torr 2021) . It controls the generation of hidden states, ensuring that the prediction of the next "infected" user is influenced by users with strong correlations, thereby enhancing prediction accuracy.

3 Problem Formulation

This section outlines the data format of the cascade and the formal definitions of the task at hand.

Information Cascade Let $V_{user} = \{v_1, v_2, \cdots, v_N\}$ be a set of N users, an information cascade with k participants is defined as a sequence $c_k = \{(v_i, t_i) \mid v_i \in V_{user}, t_i \in [0, +\infty), 0 < i \leq k\}$. Here, v_i signifies the i-th participant of cascade c_k, and forwarded the information at time t_i.

Microscopic Cascade Prediction Given an ongoing cascade c_k, the goal of microscopic cascade prediction is to predict the $(k+1)$-th participant v_{k+1}. Formally,

$$v_{k+1} = \underset{v \in V_{\text{user}} \setminus \{v_1, \ldots, v_k\}}{\arg\max} P(v|c_k), \tag{1}$$

where $P(v|c_k)$ represents the probability of user v being "infected" given k engaged users.

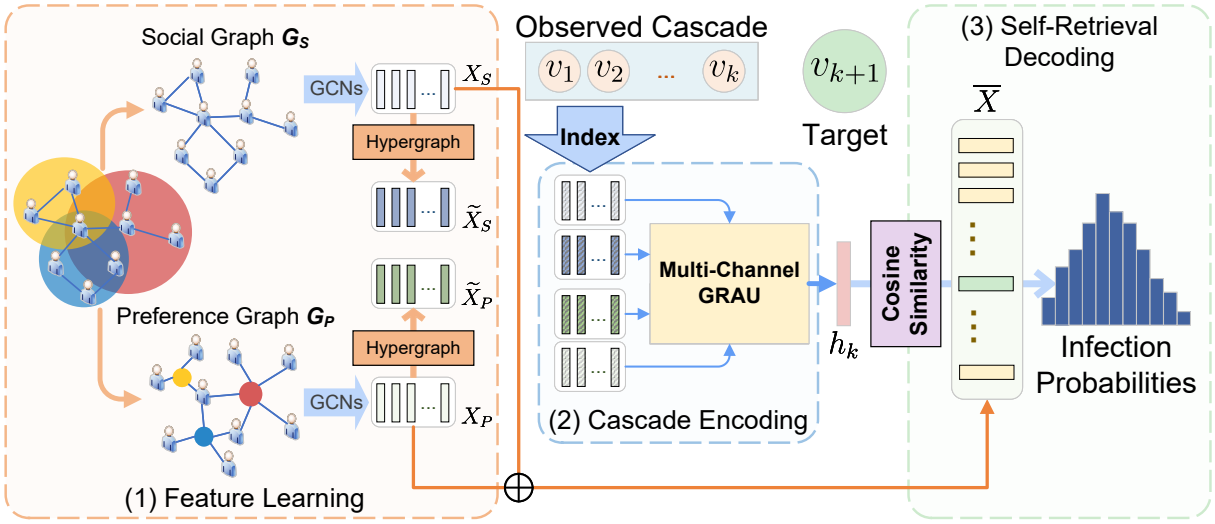


Figure 2: Architectural overview: (1) GNNs extract user features for diverse factors. (2) Process the features of "infected" users via a multi-channel GRAU. (3) A self-retrieval decoder utilizes cosine similarity to predict the target from "uninfected" users.

4 Method

In this section, we present a comprehensive overview of the framework's pipeline (Sec. 4.1) and elaborate on its two principal components: Multi-channel encoding (Sec. 4.2) and Self-retrieval decoding (Sec. 4.3).

4.1 The Pipeline of MSR Framework

This subsection delineates the proposed self-retrieval framework shown in Figure 2. Specifically, the pipeline consists of three key steps: learning features from various factors with GNNs; indexing and encoding the features relevant to infected users by a multi-channel GRAU; making predictions for the next participant by a self-retrieval decoder.

Graph Construction We consider Ω as the set of diverse factors influencing information diffusing in online social media, and build a corresponding graph G_ω to explore user features specifying to every factor $\omega \in \Omega$. Two critical factors are considered: $\Omega = \{S, P\}$ where S refers to the external factor of *social relationships* and P refers to the internal factor of *behavioral preferences*.

Concerning *social relationships*, the **social graph** $G_S = (V_{user}, E_S)$ consists of all the users. An edge from v_b to v_a in E_S indicates that user v_a is a follower of user v_b or succeeds user v_b in a cascade. One could follow many other works (Yuan et al. 2021; Sun et al. 2022; Wang et al. 2022; Li et al. 2023) to instantlize the modeling.

Additionally, we construct a **preferences graph** G_P to capture users' *behavioral preferences* based on their historical activities. Specifically, we introduce an additional node set $V_{cas} = \{v'_1, v'_2, \cdots, v'_M\}$, consisting of M nodes that collectively represent all historical cascades in the training set. The preferences graph is then defined as $G_P = (V_{user} \cup V_{cas}, E_P)$ where an edge from v'_b to v_a in E_P indicates user v_a participated in cascade v'_b. For example, a cascade could correspond to a tweet, with the edge signifying that the user either liked or retweeted this tweet.

Feature Learning With GNNs As shown in the first portion (in orange) of Figure 2, we first extract features for users from each factor ω by a L-layer GCN (Kipf and Welling 2017), respectively. The input of the l-th layer GCN contains: a feature set $X_\omega^{(l)} \in \mathbb{R}^{N_\omega \times d}$ and an adjacency matrix $A_\omega \in \mathbb{R}^{N_\omega \times N_\omega}$ (which contains N_ω nodes). Note that the input features of 1st-layer GCN $X_\omega^{(1)}$ is randomly initialized and N_ω varies for different G_ω, for example, $N_S = N$ and $N_P = (N + M)$. Since only the user part of the final GCN layer's output will be used in subsequent operations, we further define the user part of $X_\omega^{(L+1)}$ to $X_\omega \in \mathbb{R}^{N \times d}$.

To understand the complex relationships among users at a higher level, we construct a hypergraph (Bai, Zhang, and Torr 2021) based on the training dataset and then derive an incidence matrix $H \in \{0, 1\}^{N \times M}$, where $H_{a,b} = 1$ signifies user v_a participated in cascade v'_b, and 0 otherwise. Subsequently, hypergraph convolution (Hyper-Conv) (Bai, Zhang, and Torr 2021) is used to build hyper-features for users. Formally, for each factor ω, the hyper-features $\tilde{X}_\omega$ given by Hyper-Conv is defined as:

$$\tilde{X}_\omega = D^{-1} H B^{-1} H^T X_\omega \tilde{W}_\omega, \tag{2}$$

where $D \in \mathbb{R}^{N \times N}$ and $B \in \mathbb{R}^{M \times M}$ are the diagonal degree matrices of users and cascades derived from the incidence matrix H, respectively. $\tilde{W}_\omega \in \mathbb{R}^{d \times d}$ is the trainable weight.

Cascade Encoding (Sec. 4.2) We propose a multi-channel GRAU module based on GRAU (Li et al. 2023) to handle those features from various factors, i.e. X_ω and $\tilde{X}_\omega$, and it subsequently outputs a hidden state h_k to represent these "infected" users.

Self-retrieval Decoding (Sec. 4.3) We propose a novel self-retrieval decoder that makes predictions within a retrieval-based context by calculating *cosine similarity* values. Subsequently, these values are utilized to generate a probability distribution for infection.

4.2 Multi-channel Encoding

Extensive works in other domains (Nakao, Ohira, and Northoff 2012; Dietrich 2010; Acciarini, Brunetta, and Boccardelli 2021; Nibbelink and Brewer 2018) show that the diffusion of information in online social media is influenced by various factors. In response, we build a corresponding graph G_ω to explore user features specifying to factor $\omega \in \Omega$.

After the employment of GNNs, the user features are encoded separately, leading to the dismiss of connections among the "infected" users. To alleviate the issue, a sequence model, such as GRU, is typically used to encode the cascade c_k into a single vector. However, these sequence models are limited in integrating *multiple* sources of features, as they are typically designed for a *single* source.

To take **multifaceted factors** into account, based on the insights of GRAU (Li et al. 2023), we propose the multi-channel version for encoding. Preliminarily, this module expands the receptive field by an attention (Vaswani et al. 2017) mechanism to obtain context features. Formally,

$$\Phi = Softmax(\frac{(QW_Q)^T(KW_K)}{\sqrt{d}} + Mask)(VW_V), \quad (3)$$

where W_Q, W_K and W_V are trainable weights while the $Mask \in \{-\infty, 0\}^{|c_k|\times|c_k|}$ prevents future leakage by setting $Mask_{a,b}$ to $-\infty$ if $a > b$, and 0 otherwise. $\Phi \in \mathbb{R}^{|c_k|\times d}$ denotes the outputted context features.

The Structural Spreading mechanism (Li et al. 2023) suggests that incorporating hypergraph-based features could significantly improve attention mechanisms by capturing the nuanced relationships among users within a cascade. Consequently, in the definition of Q, K, and V, they are set as:

$$Q = K = [X_{i\le|c_k|}; \tilde{X}_{i\le|c_k|}], \quad V = X_{i\le|c_k|}, \quad (4)$$

Borrowing insights from research on pooling methods (Schweidtmann et al. 2023; Xu et al. 2019; Gholamalinezhad and Khosravi 2020) indicating that sum is more powerful than mean and max as it can better retain information, we integrate features from different channels through a simple yet effective operation: summation with affine transformation, e.g. $\sum_\omega W_\omega X_\omega$. When only one channel is considered, i.e. $|\Omega| = 1$, Eq. (5) recovers the original GRAU.

$$\begin{cases} r_i = \sigma(W_1^r h_{i-1} + \sum_{\omega\in\Omega} W_\omega^r X_{i,\omega}), \\ z_i = \sigma(W_1^z h_{i-1} + \sum_{\omega\in\Omega} W_\omega^z X_{i,\omega}), \\ \tilde{h}_i = Tanh(r_i \odot W_1^n h_{i-1} + \sum_{\omega\in\Omega} W_\omega^n \Phi_{i,\omega}), \\ h_i = z_i \odot h_{i-1} + (1 - z_i) \odot \tilde{h}_i, \end{cases} \quad (5)$$

Specifically, in this work, we utilize social relationships and behavioral preferences as examples for implementation. As illustrated in Figure 3, the left section utilizes attention operation and Structural Spreading mechanism to extract context features for social relationships (in the blue box) and behavioral preferences (in the green box). Then, the right segment employs Multi-Channel GRAU to recursively encode the features of infected users.

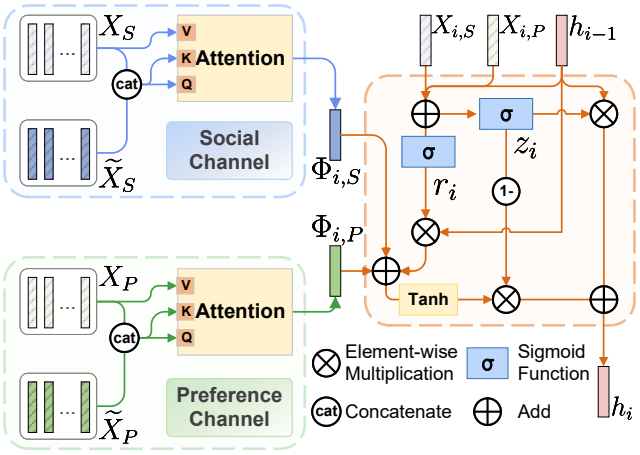


Figure 3: Instantiation of multi-channel encoding structure.

While our focus is on these two channels, this module is **inherently scalable** and can be extended to incorporate additional channels in larger networks. Ablation experiments in Sec. 5.3 demonstrate that considering any single channel alone yields sub-optimal results. Furthermore, the proposed design achieves better results than simply stacking the graphs corresponding to the two factors into a single graph, **underscoring its robustness**.

4.3 Self-retrieval Decoding

Shortage in Classification As illustrated in Eq. (1), the objective is to predict the participant most likely to be infected given h_k. Previous studies (Yuan et al. 2021; Sun et al. 2022; Wang et al. 2022; Li et al. 2023) typically formulate this task as an N-class classification problem, where each label corresponds to a user. Commonly, these works employ Multilayer Perceptrons (MLPs) to map h_k to an N-dimensional probability distribution $\hat{y}$:

$$\hat{y} = softmax(W * z(h_k) + b), \quad (6)$$

where $z(\cdot)$ represents a transformation into an internal representation through neural layers. This setup facilitates the application of the weight matrix W, where each column contributes to the corresponding category's score in $\hat{y}$. Notably, each row of W operates independently. Consequently, this classification strategy can lead to a decline in prediction accuracy due to **the lack of semantic association** among the parameters within the weight matrix W.

Shifting to Retrieval While classification setup suffers from a "**semantic irrelavant**", leveraging the rich semantic features provided by GNNs for decoding may help making prediction. Since GNNs excel at producing similar features for neighboring nodes, and based on the Social Homophily assumption, there is a high probability that the feature of the target user (v_{k+1}) might inherently resemble h_k. Exploiting this similarity could improve the effectiveness of prediction.

Consequently, we formulate microscopic cascade prediction as a retrieval task that identifies the most similar feature of user $v \in V_{user}$ given h_k. Formally, the optimal selection

is found by maximizing the similarity score from a measurement $sim(\cdot,\cdot)$, where v^*_{k+1} represents the user predicted to be the next participant and p_v, derived from GNNs, denotes the feature vector of each uninfected user v.

$$v^*_{k+1} = \underset{v \in V_{user} \setminus \{v_1,\ldots,v_k\}}{\arg\max} sim(p_v, h_k) \tag{7}$$

Instantiation of Self-retrieval Intuitively, we define the composite feature set $\overline{X} \in \mathbb{R}^{N\times d}$, where each row vector corresponds to an individual user and serves as p_v in Eq.(7). Those features reflecting different factors are aggregated to form $\overline{X}$. By default, aggregation is defined as the summation of features from distinct factors. Given that $\overline{X}$ and h_k share the same origin, i.e. X_ω , we term this retrieval formulation as "**self-retrieval**".

$$\overline{X} = \sum\nolimits_{\omega\in\Omega} X_\omega. \tag{8}$$

For each user, we calculate the *cosine similarity* between h_k and the respective row vector in $\overline{X}$ and consequently obtain $s \in [-1,1]^N$ that encapsulates the cosine-similarities between h_k and the composite features. Prior to calculating these similarities through matrix multiplication, $\overline{X}$ and h_k are normalized with the L2 norm, where $\overline{X}' \in \mathbb{R}^{N\times d}$ and $h'_k \in \mathbb{R}^d$ denote the normalized composite features and hidden state, respectively. Besides, $\overline{Mask} \in \{-\infty, 0\}^N$ excludes the users already engaged in the cascade for subsequent prediction. Specifically, for each user v_j, $\overline{Mask}_j$ is assigned a value of $-\infty$ if the user has previously participated in the cascade, and a value of 0 otherwise.

$$h'_k = \frac{h_k}{\|h_k\|_2}, \quad \overline{X}' = \left\{ \frac{\overline{X}_j}{\|\overline{X}_j\|_2} \mid \overline{X}_j \in \overline{X} \right\}, \tag{9}$$

$$s = (\overline{X}' \times h'_k) + \overline{Mask}. \tag{10}$$

Due to the narrow range of cosine similarity values, which fall in $[-1, 1]$, directly inputting these values into a softmax function could lead to training difficulties. Subsequently, for efficient training, we apply a temperature coefficient τ that scales s along with the softmax function and then obtain a probability distribution $\hat{y} \in [0,1]^N$.

$$\hat{y} = softmax(\frac{s}{\tau}). \tag{11}$$

Given the importance of the temperature coefficient τ, we also conducted a parameter selection experiment to evaluate its impact on the model's performance. As detailed in Sec. 5.4, the result indicates that the model's performance stabilizes when τ is within the range of 0.05 to 0.1, with the optimal performance observed at $\tau = 0.02$.

Optimization The goal of our method is to maximize the similarity between the hidden state h_k and the composite feature $\overline{X}_{k+1}$ corresponding to the target user ($(k+1)$-th participant) of cascade c[1]. Specifically,

$$\theta^* = \arg\max_\theta \sum_{k=1}^{|c|-1} \cos(\overline{X}_{k+1}, h_k; \theta), \tag{12}$$

where θ denotes the model parameters and cos is cosine similarity. The log-cross entropy loss is used for training, where the label y_{kj} is 1 if user v_j is the $(k+1)$-th participant, and 0 otherwise.

$$\mathcal{L} = -\sum_{k=1}^{|c|-1}\sum_{j=1}^{N} y_{kj}\log(\hat{y}_{kj}). \tag{13}$$

By focusing on the previously overlooked aspects of uninfected users' characteristics, the self-retrieval decoder discerns latent homophily among "infected" and "uninfected" users, **offering a more nuanced and effective strategy for microscopic cascade prediction**.

5 Experiments

5.1 Experiment Settings

Evaluation Metrics We evaluated models by two retrieval metrics: Mean Average Precision at top κ (MAP@κ) and Hit Score at top κ (Hits@κ), where κ is selected from $\{50, 100\}$.

Datasets We experimented with the following datasets.

- **Twitter** (Hodas and Lerman 2014): Includes retweet records from 2010, with cascades formed by retweets.
- **Memetracker** (Leskovec, Backstrom, and Kleinberg 2009): Forms cascades of identical memes from social media, without explicit follower relationships.
- **Douban** (Zhong et al. 2012): Comprises cascades based on comments for movie and book reviews.

Baselines We compared the following seven representative methods as baselines with our proposed method.

- **DeepDiffuse** (Islam et al. 2018): integrates recurrent neural networks with an attention mechanism.
- **Topo-LSTM** (Wang et al. 2017): captures topological information from cascades for prediction.
- **FOREST** (Yang et al. 2021): integrates reinforcement learning with RNNs for multi-scale prediction.
- **DyHGCN** (Yuan et al. 2021): extracts dynamic diffusion graphs from cascades to learn dynamic representations.
- **MS-HGAT** (Sun et al. 2022): integrates hypergraph techniques in cascade prediction based on time slices.
- **CEGCN** (Wang et al. 2022): builds additional nodes and a feature aggregator based on attention mechanism.
- **GRASS** (Li et al. 2023): proposes GRAU and SS mechanism for "position hopping" and "branch independence".

In alignment with previous studies (Sun et al. 2022; Wang et al. 2022; Li et al. 2023), we partitioned datasets into training, validation and test at a ratio of 8:1:1. Due to the absence of follower relationships in **Memetracker**, we excluded **Topo-LSTM** and **MS-HGAT** from its analysis.

[1]Here, c denotes a full cascade. In this domain, we split c into $(|c|-1)$ samples, i.e. c_k, for batch operation.

Models	Twitter				Memetracker				Douban			
	Hits@κ		MAP@κ		Hits@κ		MAP@κ		Hits@κ		MAP@κ	
	@50	@100	@50	@100	@50	@100	@50	@100	@50	@100	@50	@100
DeepDiffuse	8.80	13.39	3.79	3.85	26.50	34.77	8.69	8.80	14.93	19.13	5.07	5.13
Topo-LSTM	15.48	23.68	4.67	4.79	-	-	-	-	14.94	18.93	5.26	5.32
FOREST	40.95	50.39	17.88	18.02	47.41	56.77	17.21	17.34	24.79	31.25	8.38	8.47
DyHGCN	47.18	58.48	17.63	17.79	48.78	58.78	17.21	17.36	28.95	36.45	9.03	9.13
MS-HGAT	46.69	56.93	18.16	18.28	-	-	-	-	29.55	37.02	9.60	9.69
CEGCN	50.15	58.89	21.61	21.73	55.30	64.35	22.40	22.35	32.75	40.42	11.64	11.75
GRASS	51.57	60.41	22.47	22.57	61.13	68.34	27.86	27.97	37.67	44.92	15.06	15.16
MSR	**55.83**	**65.08**	**23.87**	**24.00**	**68.43**	**75.39**	**34.72**	**34.82**	**44.37**	**50.99**	**20.02**	**20.12**

Table 1: Experimental results on Hits@κ and MAP@κ metrics.

5.2 Main Result

We compared **MSR** with baselines across three datasets, with results detailed in Table 1. Best results are highlighted in bold font. Compared to the runner-up, GRASS, **MSR** is statistically significant with $p < 0.001$ on paired t-test.

The empirical data demonstrates our method's superior performance on all datasets. When compared with the second-best performing model, GRASS (Li et al. 2023), our method shows average improvements of 4.46%, 7.17%, and 6.38% in Hits@κ, and 1.41%, 6.85%, and 4.96% in MAP@κ across the datasets. Overall, the model enhances Hits@κ and MAP@κ by 6.0% and 4.4% respectively, underscoring its effectiveness. These enhancements are attributed to: **(1)** The adoption of the self-retrieval decoder, which capitalizes on the higher similarity between neighboring nodes in features obtained by GNNs, thereby overcoming limitations in capturing similarities in users' features. **(2)** The multi-viewpoint analysis of cascade prediction on various factors. As opposed to the unified approach of GRAU, our multi-channel GRAU predict the next participant from a more effective and multifaceted manner.

5.3 Ablation Study

To validate the effectiveness of individual components of our proposed method, we conducted ablation experiments on the **Twitter** and **Memetracker** datasets. These tested variants are listed below:

- **SR+**G_S: This variant employs the encoding structure of original GRAU with social graph G_S and utilizes self-retrieval decoder during decoding phase.
- **SR+**G_P: This variant employs the encoding structure of original GRAU with preference graph G_P and utilizes self-retrieval decoder during decoding phase.
- **SR+**G_U: Directly stack G_S and G_P to derive an union graph G_U. Employing the encoding structure same to the above two variants.
- **w/o SR**: Here, the multi-channel GRAU is used to obtain the hidden state h_k, which is then decoded via only a linear layer and a softmax function.
- **GRASS**: Employing G_U with GRAU for encoding and a classification strategy for decoding.

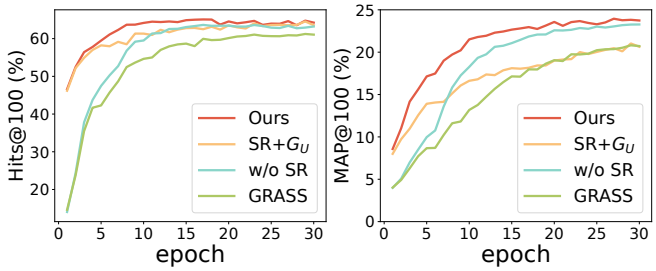
Models	Twitter		Memetracker	
	Hits@100	MAP@100	Hits@100	MAP@100
MSR	**65.08**	**24.00**	**75.39**	**34.82**
SR+G_U	64.88	22.61	75.38	33.84
SR+G_S	64.03	21.51	60.30	19.76
SR+G_P	60.66	15.88	74.29	32.83
w/o SR	63.07	23.37	71.18	31.17
GRASS	60.41	22.47	68.34	27.97

Table 2: Comparative results on Twitter and Memetracker.

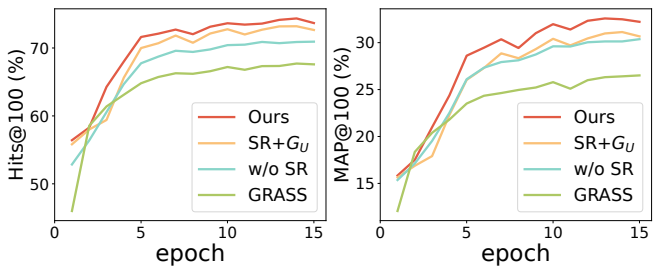
Comparative Results of Model Variants The comparative results, presented in Table 2, indicate that our proposed method outperforms all variants in every metric. For variants utilizing only the self-retrieval decoder, they generally show a decline in performance. Notably, the variant that employs the original GRAU with the union graph G_U (SR+G_U) proves to be the most effective. This is attributed to its combination of social relationships and behavioral preferences, leading to a relatively smaller decrease in performance. However, our **MSR** framework still yields superior performance compared to SR+G_U, affirming its robustness.

For Twitter dataset, the exclusion of self-retrieval (w/o SR) results in a decrease of 2.01% in Hits@100 and 0.63% in MAP@100. While for Memetracker dataset, its absence leads to a more significant reduction of 4.21% in Hits@100 and 3.65% in MAP@100, emphasizing its essential role. Notably, the Twitter dataset demonstrates sensitivity to social relationships, making the SR+G_S variant relatively effective. Conversely, the Memetracker dataset is more influenced by behavioral preferences, where G_P becomes crucial. Such variations highlight the necessity of customizing features to match unique attributes under analysis.

Performance of Variants Across Epochs We selected 3 variants (SR+G_U, w/o SR, GRASS) along with our method to analyze the performance across epochs to evaluate its training efficiency. Figures 4a and 4b illustrate the test performance of these variants on the Twitter dataset and Memetracker dataset, respectively. The analysis covers the first 30 epochs for Twitter and the initial 15 for Memetracker.



(a) Performance trends on Twitter



(b) Performance trends on Memetracker

Figure 4: Performance trends of variants across epochs on Twitter and Memetracker datasets.

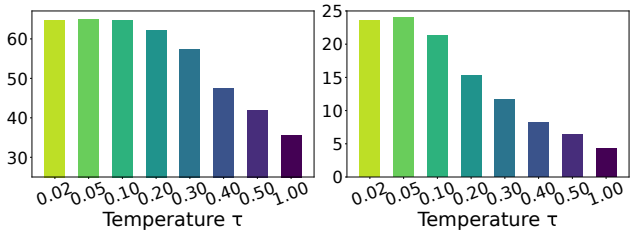


Figure 5: Performance on Hits@100 (Left) and MAP@100 (Right) with varying temperature coefficient τ.

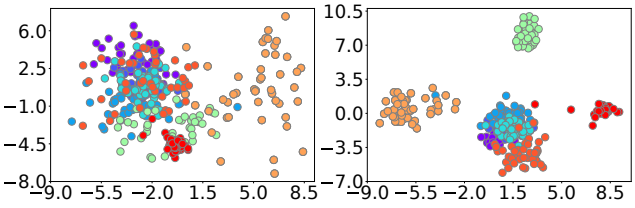


Figure 6: Visualization by PCA. The sub-figures show the distribution for GRASS (Left) and MSR (Right).

For Twitter, the self-retrieval decoder improves Hits@100 but only matches GRASS in MAP@100, while converging rapidly. In contrast, multi-channel GRAU initially underperforms but eventually excels in MAP@100. This discrepancy likely stems from MAP@κ's emphasis on rank. The multi-factor integration in multi-channel GRAU potentially facilitates more accuracy in rank. Integrating it with the self-retrieval decoder yields rapid and notable improvements.

In Memetracker, all variants show similar convergence speeds. Models with either the self-retrieval decoder or multi-channel GRAU surpass GRASS since the 4th epoch, and their integration further enhances performance. Despite Memetracker's differences from Twitter, likely due to no explicit follower relationships, excluding multi-channel GRAU significantly reduces MAP@100 compared to Hits@100, highlighting its rank sensitivity.

5.4 Parameter Analysis

Given the narrow range of cosine similarity within $[-1, 1]$, scaling these values using τ is essential for facilitating effective learning. Recognizing its importance, we conducted a parameter selection experiment on the **Twitter** dataset to analyze the impact of τ on the performance. The τ was selected within a spectrum ranging from 0.02 to 1.00.

The results shown in Figure 5 indicate a clear trend: as τ decreases from 1.00 to 0.10, both MAP@100 and Hits@100 show improvement. Most notably, when τ is within the range of [0.02, 0.10], the method's performance appears to plateau, suggesting an optimal efficiency in this range. Consequently, 0.05 is identified as yielding the best performance and is thus selected as the default setting for our method.

5.5 Visualization Analysis of Cascades

We employ Principal Component Analysis (PCA) to visualize user features and assess the efficacy of the proposed method. In this analysis, a more clustered distribution after dimensional reduction indicates a better fit to the assumption of *Social Homophily*. In the Twitter dataset, we randomly select 10 cascades with 40-60 participants each, using checkpoints to generate user features, which are color-coded by same cascade. We compare our method against the second-best baseline model, GRASS (Li et al. 2023). For GRASS, features derived from GCNs are used for PCA, while for our proposed method, we utilize the composite features $\overline{X}$.

After applying PCA, the distribution depicted in Figure 6 shows that the features learned by our method are more closely clustered compared to those derived from GRASS. This proximity suggests that our similarity-based approach more effectively captures *Social Homophily*, offering significant advantages in microscopic cascade prediction.

6 Conclusion

In this study, we introduced a novel multifaceted self-retrieval style microscopic cascade prediction framework that integrates the factors of both social relationships and behavioral preferences. During the encoding phase, we proposed a multi-channel GRAU module to dissect the impacts stemming from diverse factors. During the decoding phase, we proposed a self-retrieval decoder to optimize the application of the Social Homophily assumption, enhancing the precision for microscopic cascade prediction. Empirical evaluations on three publicly available datasets demonstrate that our method achieves substantial improvements in both MAP@κ and Hits@κ metrics, affirming its efficacy.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62476060). We appreciate all the co-workers' constructive comments, which significantly contributed to the development of this paper.

References

Acciarini, C.; Brunetta, F.; and Boccardelli, P. 2021. Cognitive biases and decision-making strategies in times of change: a systematic literature review. *Management Decision*.

Bai, S.; Zhang, F.; and Torr, P. H. 2021. Hypergraph convolution and hypergraph attention. *Pattern Recognition*.

Bikhchandani, S.; Hirshleifer, D.; and Welch, I. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*.

Cao, Q.; Shen, H.; Cen, K.; Ouyang, W.; and Cheng, X. 2017. Deephawkes: Bridging the gap between prediction and understanding of information cascades. In *CIKM*.

Cheng, J.; Adamic, L.; Dow, P. A.; Kleinberg, J. M.; and Leskovec, J. 2014. Can cascades be predicted? In *WWW*.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NeurIPS Workshop on Deep Learning*.

Crane, R.; and Sornette, D. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS*.

Dietrich, C. 2010. Decision making: Factors that influence decision making, heuristics used, and decision outcomes. *Inquiries Journal*.

Dow, P. A.; Adamic, L.; and Friggeri, A. 2021. The Anatomy of Large Facebook Cascades. In *ICWSM*.

Du, N.; Dai, H.; Trivedi, R.; Upadhyay, U.; Gomez-Rodriguez, M.; and Song, L. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *SIGKDD*.

Galuba, W.; Aberer, K.; Chakraborty, D.; Despotovic, Z.; and Kellerer, W. 2010. Outtweeting the {Twitterers—predicting} information cascades in microblogs. In *WOSN*.

Gholamalinezhad, H.; and Khosravi, H. 2020. Pooling methods in deep neural networks, a review. *arXiv*.

Gong, X.; Huskey, R.; Xue, H.; Shen, C.; and Frey, S. 2023. Broadcast information diffusion processes on social media networks: exogenous events lead to more integrated public discourse. *J. Commun.*

Granovetter, M. 1978. Threshold models of collective behavior. *Am. J. Sociol.*

Griffin, R. 2023. Rethinking rights in social media governance: human rights, ideology and inequality. *ELO*.

Guille, A.; Hacid, H.; Favre, C.; and Zighed, D. A. 2013. Information diffusion in online social networks: A survey. *ACM SIGMOD Record*.

Hodas, N. O.; and Lerman, K. 2014. The simple rules of social contagion. *Scientific reports*, 4(1): 4343.

Iamnitchi, A.; Hall, L. O.; Horawalavithana, S.; Mubang, F.; Ng, K. W.; and Skvoretz, J. 2023. Modeling information diffusion in social media: data-driven observations. *Frontiers in Big Data*.

Islam, M. R.; Muthiah, S.; Adhikari, B.; Prakash, B. A.; and Ramakrishnan, N. 2018. Deepdiffuse: Predicting the'who'and'when'in cascades. In *ICDM*.

Jalili, M.; and Perc, M. 2017. Information cascades in complex networks. *J. Complex Netw.*

Kempe, D.; Kleinberg, J.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *SIGKDD*.

Khan, A. M.; Soroya, S. H.; and Mahmood, K. 2024. Impact of information credibility on social media information adoption behavior: a systematic literature review. *Library Hi Tech*.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

Kossinets, G.; and Watts, D. J. 2009. Origins of homophily in an evolving social network. *Am. J. Sociol.*

Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Meme-tracking and the dynamics of the news cycle. In *SIGKDD*.

Li, H.; Xia, C.; Wang, T.; Wang, Z.; Cui, P.; and Li, X. 2023. GRASS: Learning Spatial–Temporal Properties From Chainlike Cascade Data for Microscopic Diffusion Prediction. *TNNLS*.

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.*

Mishra, S.; Rizoiu, M.-A.; and Xie, L. 2016. Feature driven and point process approaches for popularity prediction. In *CIKM*.

Nakao, T.; Ohira, H.; and Northoff, G. 2012. Distinction between externally vs. internally guided decision-making: operational differences, meta-analytical comparisons and their theoretical implications. *Front. Neurosci.*

Nibbelink, C. W.; and Brewer, B. B. 2018. Decision-making in nursing practice: An integrative literature review. *Journal of clinical nursing*.

Qiu, J.; Li, Y.; Tang, J.; Lu, Z.; Ye, H.; Chen, B.; Yang, Q.; and Hopcroft, J. E. 2016. The lifecycle and cascade of wechat social messaging groups. In *WWW*.

Sankar, A.; Zhang, X.; Krishnan, A.; and Han, J. 2020. Inf-VAE: A variational autoencoder framework to integrate homophily and influence in diffusion prediction. In *WSDM*.

Schweidtmann, A. M.; Rittig, J. G.; Weber, J. M.; Grohe, M.; Dahmen, M.; Leonhard, K.; and Mitsos, A. 2023. Physical pooling functions in graph neural networks for molecular property prediction. *Comput. Aided Chem. Eng.*

Sun, L.; Rao, Y.; Zhang, X.; Lan, Y.; and Yu, S. 2022. MS-HGAT: Memory-Enhanced Sequential Hypergraph Attention Network for Information Diffusion Prediction. In *AAAI*.

Tsagkias, M.; Weerkamp, W.; and De Rijke, M. 2009. Predicting the volume of comments on online news stories. In *CIKM*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *NeurIPS*.

Wang, D.; Wei, L.; Yuan, C.; Bao, Y.; Zhou, W.; Zhu, X.; and Hu, S. 2022. Cascade-Enhanced Graph Convolutional Network for Information Diffusion Prediction. In *DASFAA*.

Wang, J.; Zheng, V. W.; Liu, Z.; and Chang, K. C.-C. 2017. Topological recurrent neural network for diffusion prediction. In *ICDM*.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How Powerful are Graph Neural Networks? In *ICLR*.

Yang, C.; Wang, H.; Tang, J.; Shi, C.; Sun, M.; Cui, G.; and Liu, Z. 2021. Full-scale information diffusion prediction with reinforced recurrent networks. *TNNLS*.

Yuan, C.; Li, J.; Zhou, W.; Lu, Y.; Zhang, X.; and Hu, S. 2021. DyHGCN: A dynamic heterogeneous graph convolutional network to learn users' dynamic preferences for information diffusion prediction. In *ECML-PKDD*.

Yuan, N. J.; Zhong, Y.; Zhang, F.; Xie, X.; Lin, C.-Y.; and Rui, Y. 2016. Who will reply to/retweet this tweet? The dynamics of intimacy from online social interactions. In *WSDM*.

Yunbing, W.; Hang, G.; Weisen, Z.; and Aiying, Y. 2023. Integrating User Relation Representations and Information Diffusion Topology Features for Information Propagation Prediction. *FCST*.

Zhong, E.; Fan, W.; Wang, J.; Xiao, L.; and Li, Y. 2012. Comsoc: adaptive transfer of user behaviors over composite social network. In *SIGKDD*.

Zhou, F.; Xu, X.; Trajcevski, G.; and Zhang, K. 2021. A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Computing Surveys (CSUR)*.