THGNets: Constrained Temporal Hypergraphs and Graph Neural Networks in Hyperbolic Space for Information Diffusion Prediction

Yanchao Liu, Pengzhou Zhang, Wenchao Song, Yao Zheng, Deyu Li, Lei Shi*, Junpeng Gong*†

State Key Laboratory of Media Convergence and Communication, Communication University of China, China
{yanchaoliu, zhangpengzhou, Songwenchao, zhengyao, ldy, leiky_shi, jpgong}@cuc.edu.cn

Abstract

Information diffusion prediction aims to predict the next infected user in the information diffusion, which is a critical task to understand how information spreads on social platforms. Existing methods mainly focus on the sequences or topology structure in euclidean space. However, they fail to sufficiently consider the hierarchical structure or power-law structure of the underlying topology of information cascade graphs and social networks, resulting in distortion of user features. To tackle above issue, we propose an innovative Constrained **T**emporal **H**ypergraphs and **G**raph Neural **Net**work**s** (**THGNets**) framework that is tailored for information diffusion prediction. Specifically, we introduce hyperbolic temporal hypergraphs neural network to alleviate the distortion of user features by hyperbolic hierarchical learning in information cascades. Additionally, it also captures high-order dynamic interaction patterns between users and further integrates the time-consistency constraint mechanism to mitigate the instability and non-smoothness of user features in latent space. In parallel, we apply the hyperbolic graph neural network to investigate the hierarchical structure and user homogeneity on social networks, enhancing our understanding of social relationships. Moreover, hyperbolic gated recurrent units are employed to capture the potential dependency relationships between contextual users. Experiments conducted on four public datasets demonstrate that the proposed THGNets significantly outperform the existing methods, thereby validating the superiority and rationality of our approach.

Introduction

Online social media platforms (Weibo, Twitter, etc.) produce a large amount of information every day, providing great convenience for our lives, work, and study. Information diffusion formed by users forwarding information, aka information cascades, is ubiquitous on social networks. Information diffusion prediction, aiming to predict the next infected user, is urgently needed to understand how information spreads. It has attracted great attention in both academia and industry with a wide range of real-world applications in user discovery (Oro et al. 2017), influence maximization (He et al. 2023), recommendation systems (Chen et al. 2024a,b), and so forth.

Recently, information diffusion prediction has gained considerable attention, generally divided into three categories: independent cascades-based methods, embedding-based methods, and deep learning representation methods. Independent cascades-based methods (Saito et al. 2009) are predefined models that are not in line with the complex real-world environment. Embedding-based methods (Bourigault, Lamprier, and Gallinari 2016) parameterize user features and apply the distance or similarity between the embedding vectors of the user to predict the next infected users. Afterward, deep learning representation methods became mainstream methods with their powerful implicit feature extraction capabilities. Researchers first focus primarily on the sequence (Li et al. 2017; Wang et al. 2017; Islam et al. 2018), and recurrent neural networks are employed to learn the evolution patterns. However, such methods tend to ignore the significant topological structure, which undoubtedly potentially limits the models performance. To address this dilemma, researchers utilize graph neural networks to capture topological structures (Yang et al. 2021a; Yuan et al. 2021; Jiao et al. 2024), which achieve remarkable results.

Although existing approaches have achieved notable performance, they still suffer from some limitations: **Distortion of user features.** Recent studies (Meng and Zhou 2023) show that real information cascade graphs and social networks have complex structures, including hierarchical and power-law structures. RotDiff (Qiao et al. 2023) learns the hierarchical structure or the power-law structure by calculating the distance between directly connected users in a hyperbolic space, but it focuses only on low-order interaction patterns and ignores high-order interaction patterns between users. Thus, the distortion of user features has not been fully addressed. **Instability and non-smoothness of user features.** Information diffusion is a dynamic process in which user preferences evolve over time, so the position of user features may undergo significant changes in the latent space. However, existing work (Jiao et al. 2024) ignores this important phenomenon, leading to instability and nonsmoothness of user features with no difficulty. The above limitations are an urgent problem that needs to be solved due to limiting the ability of user feature expression.

To resolve above limitations, we propose an innova-

*Corresponding author
†Senior author

tive **THGNets** framework, which is specifically tailored for information diffusion prediction. To effectively alleviate the distortion of user features and accurately obtain user dynamic features, we propose the **H**yperbolic **T**emporal **H**ypergraph **N**eural **N**etwork(**HTHNN**) to capture the hierarchical structure or power-law structures, and the high-order dynamic interaction patterns between users. And we further design a time-consistency constraint to mitigate the instability and non-smoothness of user features in the latent space, enhancing the ability of user features expression. Moreover, we exploit the **H**yperbolic **G**raph **N**eural **N**etwork (**HGNN**) to capture the hierarchical structure and user homogeneity on social networks. In addition, we advance the **H**yperbolic **G**ated **R**ecurrent **U**nits (**HGRUs**) to explore the potential dependency between contextual users from interaction relationships and social relationships in hyperbolic space, respectively. We summarize the contributions of this paper as follows.

- We first introduce hyperbolic hypergraph and graph neural networks for information diffusion prediction. THGNets can effectively learn the hierarchical structure or power-law structures implicitly present in information cascade graphs and social networks.
- We propose the HTHNN to effectively capture the hierarchical structure and high-order dynamic interaction patterns, alleviating the distortion of user features. We further design a time-consistency constraint to ensure the stability and smoothness of user features.
- We present the HGNN to learn the hierarchical structure and user homogeneity on social networks. Meanwhile, we advance two HGRUs to mine the potential dependencies from interaction relationships and social relationships, respectively, effectively improving user feature expression ability.
- Extensive experiments conducted on four public real-world datasets demonstrate that THGNets significantly outperform existing methods, consequently validating the superiority and rationality.

Problem Formulation

Information diffusion prediction. According to historical information cascades, information diffusion prediction aims to predict the next infected user. Concretely, given the information cascade s_m and the social network G, our goal is to build the model $f(s_m, G)$ to predict the possibility $p(u_{m,n})$ of the next infected user $u_{m,n}$, which is mathematically formally defined as $f(s_m, G)\Theta{\rightarrow}p(u_{m,n})$. Table 1 shows the commonly used symbols.

Symbol	Meaning
G	Social networks.
U	User sets.
s_m^k	The $k-th$ subgraph of the $m-th$ cascade.
E	The social relationship sets.
Θ	The model parameter.
c	The curvature.
$D^{n,c}$	The hyperbolic space.
$\Gamma_x D^{n,c}$	The tangent space.
S	The cascades sets.
K	The number of subgraph intervals.
s_m	The $m-th$ information cascade.
n	The length of information cascade.
ε_m	The $m-th$ information diffusion sets.
d	The embedding dimension.

Table 1: Commonly used symbols.

Method

In this section, we comprehensively introduce our THGNets model as shown in Figure 1, which primarily includes four components.

User dynamic representation learning. We exploit the HTHNN to effectively capture the hierarchical structure and higher-order dynamic interaction patterns, alleviating the distortion of user features. To capture the dynamic of user features, we design a gate mechanism to iteratively fuse user features of contiguous snapshots. To further mitigate the instability and non-smoothness of user features, we design the time-consistency constraint mechanism to constraint the distance of user features in the latent space.

User static representation learning. We mainly leverage HGNN to jointly learn user homogeneity and hierarchical structure on social networks.

User dependency relationship representation learning. We employ two HGRUs to learn the dependency between users from interaction relationships and social relationships, respectively. Meanwhile, the designed gate mechanism fuses user static and dynamic features, improving the expression ability of user features.

Prediction. The fused user features are applied to predict the next infected user.

An information cascade $s_m = (U_m, \varepsilon_m, T_m)$ is divided into several cascade subgraphs $s_m^k = \left(U_{m,k}, \varepsilon_{m,k}\right), k \in (0, K], s_m^k \in s_m$, which means that all users participate in forwarding the $m-th$ information items before the k interval. We further construct hypergraphs for each cascade subgraph, randomly initialize all user features $X_0^E \in \Re^{|U|\times d}$ with a normal distribution, and obtain user features $X_0^H = proj^c\left(exp_o^c(X_0^E)\right)$ in a hyperbolic space.

User Dynamic Representation Learning

Hyperbolic hypergraph neural network. We exploit hyperbolic hypergraph neural network to learn the hypergraph of each cascade subgraphs, capturing high-order dynamic interaction patterns between users and the hierarchical structure of information cascades, alleviating the distortion of user features. The hyperedges and vertices denote the information items corresponding to information cascades and users, respectively. The learning process of the hyperbolic hypergraph neural network principally includes two stages.

Vertex to Hyperedge. For the given cascade subgraph ${s_m}^k$, we update the hyperedge representation $y_{m,k}$ by aggregating the user sets features X_k^H, who participate in forwarding the $m-th$ information item before the k interval,

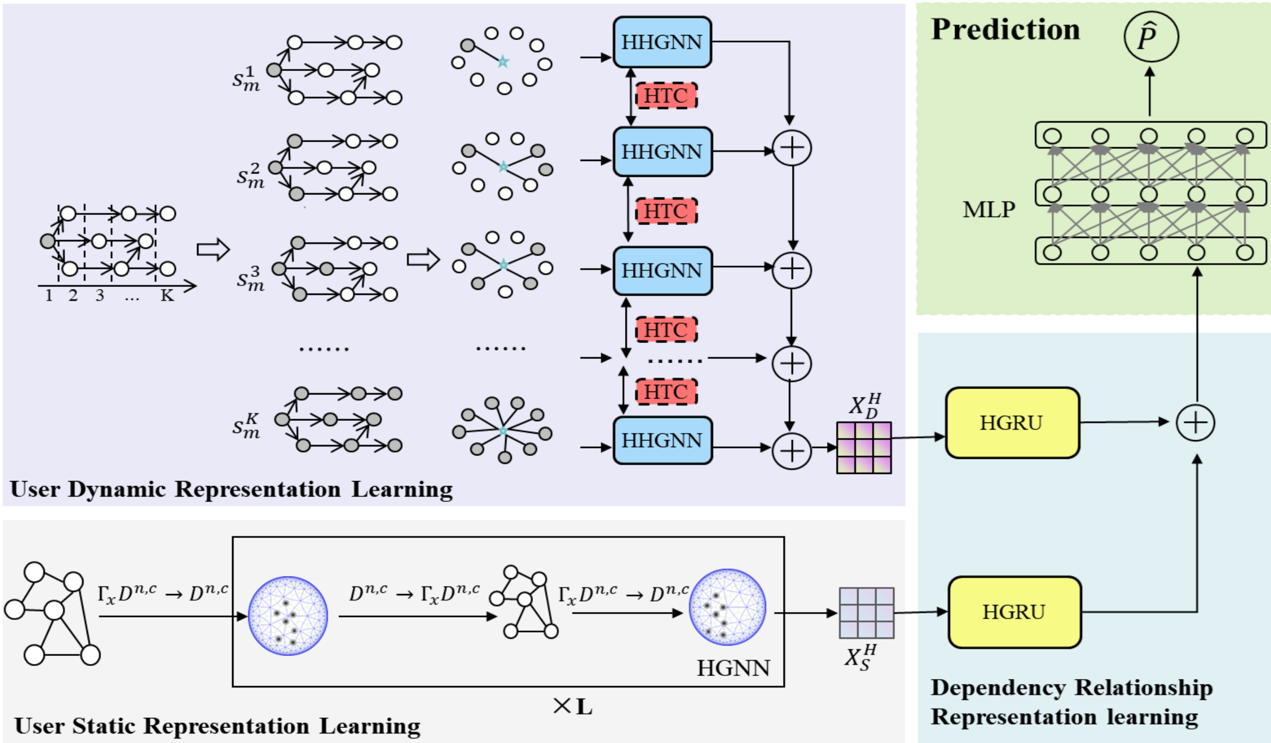

Figure 1: The overall architecture of the proposed THGNets framework.

which is defined as follows.

$$y_{m,k} = \sigma(w_{e_m^k} \sum_{u_{i,k} \in N_v(e_m^k)} \frac{\log_o^c u_{i,k}}{|N_v(e_m^k)|}), \tag{1}$$

where $u_{i,k} \in X_k^H$, $\sigma(\cdot)$ represents a non-linear activation function $\mathrm{Re}lu$, $N_v(e_m^k)$ denotes user sets connected by the hyperedge e_m^k, and $w_{e_m^k}$ is the hyperedge weight. Following previous research (Jiao et al. 2024) and ignoring differences in the importance and influence of different information cascades, therefore $w_{e_m^k}$ is set to 1.

Hyperedge to Vertex. Based on learned hyperedges feature $y_{m,k}$, user feature representation $u_{i,k}$ is learned by aggregating related hyperedges, meaning that the user $u_{i,k}$ forwards multiple information items, which is defined as follows.

$$u_{i,k} = proj^c\left(exp_o^c(\sigma(\Upsilon \sum_{e_m^k \in N_e} \frac{y_{m,k}}{|N_e(u_{i,k})|}))\right), \tag{2}$$

where $N_e(u_{i,k})$ represents the hyperedges sets connected to the users $u_{i,k}$, Υ represents the learnable weight, and $proj^c(\cdot)$ restricts the feature representation within the sphere.

Hyperbolic fusion. User preferences change dynamically over time in information diffusion process. We design the gate mechanism to iteratively fuse dynamic user features from the output of contiguous subgraphs learned by the hypergraph neural network, considering the dynamic of user preferences and the importance of user preferences in different time intervals, which is defined as follows.

$$\rho = \frac{\exp\left(\sigma(W_1 \cdot \log_o^c\ X_k^H)\right)}{\sum_{q \in (k-1,k)} \exp\left(\sigma(W_1\ \cdot \log_o^c X_q^H)\right)}, \tag{3}$$

$$X_k^H = proj^c(exp_o^c(\rho \cdot log_o^c X_k^H + (1-\rho) \cdot log_o^c X_{k-1}^H)), \tag{4}$$

where W_1 signifies learnable parameters.

Hyperbolic temporal consistency constraints. Due to user preferences changing over time, the position of user features changes in the latent space, and thus user features may be unstable and non-smoothness. To mitigate this issue, we design a time-consistency constraint mechanism to constraint the distance of user features in the latent space. The loss function is defined as follows.

$$Loss_{ts} = \begin{cases} \frac{1}{K-1} \sum_{k=0}^{K-1} d_H^c\left(X_k^H, X_{k+1}^H\right), k > 1 \\ 0, k = 0, 1 \end{cases} \tag{5}$$

$$d_H^c\left(u_{i,k}, u_{i,k+1}\right) = \frac{2 \operatorname{arctan} h\left(\sqrt{c}\left\|-u_{i,k} \oplus_c u_{i,k+1}\right\|\right)}{\sqrt{c}}, \tag{6}$$

where $d_H^c(\cdot)$ is used to calculate the distance of user feature in hyperbolic space. Finally, we obtain user dynamic features X_D^H.

User Static Representation Learning

Users with similar preferences, named user homogeneity, may have followers or friend relationships on social networks, which has a significant potential impact on information diffusion.

Hyperbolic graph neural network. We apply HGNN (Chami et al. 2019) to capture the hierarchical structure and user homogeneity, alleviating the distortion of user features, which comprises three stages.

Hyperbolic linear transformation. It mainly includes linear transformation and bias terms. Since the special geometric properties in hyperbolic space, the operations in Euclidean space cannot be directly applied. For brevity, we first map the user features from the hyperbolic space to the tangent space with $\log_o^c(\cdot)$ for linear transformation and then map the calculated result to the hyperbolic space with $\exp_o^c(\cdot)$, which is defined as follows.

$$W_2 \otimes_c X_0^H = \exp_o^c\left(W_2 \cdot \log_o^c X_0^H\right), \tag{7}$$

where W_2 denotes learnable parameters. For the bias term, we first parallel the bias term to the tangent space and then map it to the hyperbolic space with $\exp_o^c(\cdot)$. Finally, the hyperbolic linear transformation is defined as follows.

$$b^H = proj^c\left(\exp_o^c\left(P_{o\to x}(b)\right)\right), \tag{8}$$

$$HLiner\left(X_0^H\right) = W_2 \otimes_c X_0^H \oplus_c b^H, \tag{9}$$

where $\otimes_c$ represents the multiplication operation and $\oplus_c$ is the addition operation in the hyperbolic space, respectively.

Hyperbolic aggregation. Similar to GAT, we update the features of the target user features with weighted neighboring features in tangent space, which is defined as follows.

$$HAg\left(u_i\right) = proj^c\left(\exp_o^c\left(\sum_{j\in N(i)} \alpha_{ij} \cdot \log_o^c u_j\right)\right), \tag{10}$$

$$\alpha_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum\limits_{j\in N(i)} \exp\left(e_{ij}\right)}, \tag{11}$$

$$e_{ij} = \sigma\left(W_3 \cdot \left(\log_o^c u_i || \log_o^c u_j\right)\right), \tag{12}$$

where $N(i)$ is the neighbors sets of the user u_i and W_3 symbolizes learnable parameters.

Hyperbolic activation. Based on the aggregation operation, we use the function $Relu$ to activate user features, which is defined as follows.

$$HAct\left(u_i\right) = proj^c\left(\exp_o^c\left(\sigma\left(\log_o^c u_i\right)\right)\right). \tag{13}$$

Finally, we obtain static user features X_S^H on social networks.

User Dependency Relationship Representation Learning

Inspired by GRU, we further apply two HGRUs (Yang et al. 2021b) to learn the potential dependency between contextual users from dynamic interaction relationships and social relationships in hyperbolic space. For a given cascade s_m, we obtain users static features $s^m_{X_S^H} = \left\{u^0_{X_S^H}, u^1_{X_S^H}, ..., u^{n-1}_{X_S^H}\right\}$ and user dynamic features $s^m_{X_D^H} = \left\{u^0_{X_D^H}, u^1_{X_D^H}, ..., u^{n-1}_{X_D^H}\right\}$ with the memory lookup mechanism from user dynamic features X_D^H and user static features X_S^H, respectively.

Hyperbolic recurrent neural network. We learn the user static features $X^m_{X_S^H}$ and the user dynamic features $X^m_{X_D{}^H}$ from the social relationships and interaction relationships using HGRUs. For example, $X^m_{X_S{}^H} = HGRU(s^m_{X_S{}^H})$, $X^m_{X_D{}^H} = HGRU(s^m_{X_D{}^H})$.

Hyperbolic fusion. We design a gate mechanism to fuse user static features $X^m_{X_S{}^H}$ and user dynamic features $X^m_{X_D{}^H}$ in information diffusion, which is defined as follows.

$$\delta = \frac{\exp\left(\sigma\left(W_4 \cdot \left(\log_o^c X^m_{X_S{}^H}\right)\right)\right)}{\sum\limits_{q\in(D,S)} \exp\left(\sigma\left(W_4 \cdot \left(\log_o^c X^m_{X_q{}^H}\right)\right)\right)}, \tag{14}$$

$$X_m^E = (1-\delta)\cdot \log_o^c X^m_{X_S^H} + \delta \cdot \log_o^c X^m_{X_D^H}, \tag{15}$$

$$X_m^H = proj^c\left(\exp_o^c X_m^E\right), \tag{16}$$

where δ denotes the weights of user-static features, W_4 means learnable parameters. Finally, we obtain user features X_m^H.

Prediction

We utilize multilayer perceptrons (MLPs) to predict the probability of the next infected user, which is defined as follows.

$$\widehat{p}_{mi} = softmax\left(MLP\left(\log_o^c X_m^H + Mask\right)\right). \tag{17}$$

The total loss of our model is defined as follows.

$$Loss = -\frac{1}{|S|}\sum_{i=1}^{|U|}(\widehat{p}_{mi} \cdot logp_{mi} + Loss_{ts}) + \beta \left\|\Theta\right\|_2, \tag{18}$$

where $\widehat{p}_{mi}$ denotes the predicted value, p_{mi} is the real value, β symbolizes the decay factor, $||\cdot||_2$ refers to the regularization term of the L2-norm, Θ expresses all learnable parameters of our model, Mask indicates the user mask, the value is set to negative infinity for users who have forwarded the information item, otherwise, set to 0.

Complexity Analysis

In this section, we primarily analyze the time complexity of the main components. The time complexity of the user dynamics representation learning module is $O\left(|U|d^2 + |\varepsilon_k|\, kd\right)$. The time complexity of the user static representation learning module is $O\left(|U|\, d^2 + |\varepsilon|\, d\right)$. The time complexity of the user dependency relationship representation learning module is $O\left(n\left(dd_h + d_h^2\right)\right)$, where d_h denotes the embedding dimension of hidden features.

Experiments

Datasets. Following previous work (Jiao et al. 2024), we choose four public available datasets, such as Christianity (Sankar et al. 2020), Android (Sankar et al. 2020), Douban (Zhong et al. 2012), and Memetracker (Leskovec, Backstrom, and Kleinberg 2009). In addition, we randomly split the dataset into training, validating, and testing datasets for each dataset with the ratio of 0.8:0.1:0.1.

Baseline Models. To verify the performance of our model, we choose nine baseline models, such as Topo LSTM(Wang et al. 2017), NDM(Yang et al. 2019), SNIDSA(Wang, Chen, and Li 2018), Inf-VAE(Sankar et al. 2020), FOREST(Yang et al. 2021a), DyHGCN(Yuan et al. 2021), DMT-LIC (Chen et al. 2019), TAN-DRUD (Liu et al. 2022), and MINDS(Jiao et al. 2024). Furthermore, we use two ranking metrics, such as Mean Average Precision on top ($MAP@k$) and Hits socres on top k ($Hits@k$), k =

$10, 50, 100$ (Sun et al. 2022; Yuan et al. 2021; Jiao et al. 2024), where the higher the values, the better the model performance.

Parameters Settings. THGNets adopt adam optimizer to optimize parameters with the learning rate 3e-4, the batch size is 64, the embedding size is 128, the length of the information cascades is 200, the curvature is 1.0, the dropout is 0.3, the layer numbers of the HGNN are 2, the layer numbers of the hyperbolic hypergraph neural network is 1, the number of cascade intervals are 6, and the regularization decay is 5e-3.

Performance Comparison

To verify the THGNets performance, we comprehensively compare it with baseline models, primarily focusing on information diffusion prediction at the micro level. The results are shown in Tables 2 and 3, and we observe the following conclusions.

THGNets consistently outperforms the baseline models on four public datasets. Comparing with MINDS model on the four public datasets, our model can learn the hierarchical structure of the information cascade graphs and social networks in hyperbolic space, alleviating the distortion of user features and achieving an average improvement of 9.3% and 7.4% on android dataset, respectively. Comparing with Topo LSTM, NDM, and DMT-LIC, our model outperforms much better, benefiting from the user homogeneity on social networks. Comparing with SNIDSA, Inf-VAE, DyHGCN, TAN-DRUD, and FOREST, our model is superior to them by comprehensively considering the higher-order interaction patterns between users.

Ablation Study

Some ablation studies are conducted to verify the effectiveness of the main components on Christianity and Memetracker datasets, in which the scale corresponds to the minimum and maximum for information cascades. We list five variant methods. The results are shown in Figure 2, and we observe the following conclusions.

THGNets achieves promising results in Figure 2, proving the effectiveness of the designed module. THGNets performance significantly decreases after removing HTC, so the time-consistency constraint mechanism can effectively ensure the stability and smoothness of user features in latent space. The experimental results show that the importance of user features learned from different snapshots is inconsistent, thus user preference changes dynamically, as shown by w/o HFusion_o. The w/o HFusion_t shows that the gating mechanism effectively and dynamically combines interaction relationships and social relationships. THGNets performance is greatly reduced due to removing social networks, benefiting from user homogeneity and hierarchical structure on social networks. THGNets can effectively learn the high-order interaction pattern in hyperbolic space, as demonstrated by w/o HHGNN.

Parameters Analysis

Some sensitivity analysis experiments are conducted about the key parameters on Christianity and Douban datasets, in which the scale of user sets corresponds to the minimum and maximum in information cascades, verifying how different hyperparameters affect our model performance, such as user embedding dimension, cascade intervals, the curvature, and the proportion of the training datasets, as shown in Figures 3,4,5, and 6. We summarize the following conclusions.

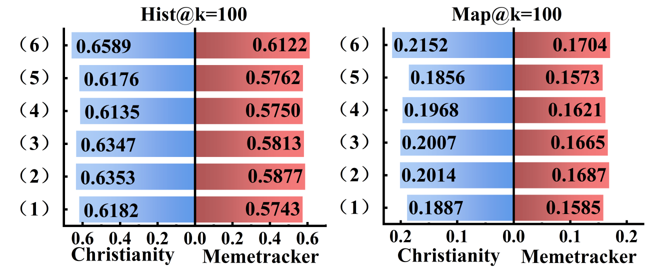

Figure 2: Ablation Study. (1)w/o HTC: removing the time-consistency constraint mechanism. (2)w/o HFusion_o: removing the fusion gating mechanism for capturing the dynamic of user preferences and replacing with the addition operation. (3)w/o HFusion_t: removing the fusion gating mechanism for fusing interaction relationships and social relationships and replacing with the addition operation. (4)w/o social network: removing social network. (5)w/o HHGNN: replacing the hyperbolic hypergraph with the hyperbolic graph neural network. (6)THGNets: Our model.

In Figure 3, experiments demonstrate that the performance of our model first increases and then decreases slightly with the rise of user features embedding dimension, indicating that the information cascades contain rich user features. However, other critical factors, contributing to damage our performance, are redundancy and noise. Figure 4 unsurprisingly finds that the model performance is improved by the advantage of cascade subgraphs for capturing the dynamic of user preference in the latent space. However, excessive isolated cascade subgraphs decrease model performance. In addition, compared to the literature (Jiao et al. 2024), our model contains all users before the interval k. Figure 5 shows that our model effectively learns the hierarchical structure on information cascade graphs and social networks, thus alleviating the distortion of user features. As the curvature increases, the performance first increases and then decreases. Experiments demonstrate that as the proportion of the training dataset increases, our model performance gradually improves, benefiting from the fact that the proposed model can learn more effective features for information diffusion prediction in Figure 6. Considering comprehensively, the user embedding dimension is 128, the cascade intervals are 6, the curvature is 1.0, and the proportion of the training dataset is 0.8.

Conclusion

In this paper, we propose a novel THGNets framework for information diffusion prediction. Concretely, we introduce hyperbolic temporal hypergraph neural networks and hyperbolic graph neural networks to capture user homogeneity and high-order interaction patterns with hierarchical learn-

Models	Christianity			Android			Douban			Memetracker		
	10	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**
Topo-LSTM	0.1559	0.3653	0.4777	0.0460	0.1318	0.2103	0.0306	0.0143	0.0184	0.1908	0.3687	0.4683
NDM	0.0464	0.1145	0.1461	0.0170	0.0423	0.0555	0.0388	0.0506	0.0528	0.0931	0.1228	0.1279
SNIDSA	0.0660	0.2098	0.3502	0.0271	0.0829	0.1299	0.0702	0.1807	0.2324	0.1395	0.2945	0.3977
Inf-VAE	0.0767	0.2569	0.3853	0.0318	0.0938	0.1452	0.1364	0.2361	0.3059	0.1165	0.3096	0.4200
DyHGCN	0.2380	0.4689	0.5923	0.0748	0.1746	0.2596	0.1438	0.2648	0.3329	0.2522	0.4603	0.5710
TAN-DRUD	0.1908	0.4406	0.5697	0.0281	0.1024	0.1658	0.0841	0.1604	0.2175	0.2139	0.4247	0.5383
FOREST	0.2746	0.4665	0.5603	0.0866	0.1739	0.2314	0.1106	0.1986	0.2559	0.2648	0.4502	0.5499
DMT-LIC	0.2768	0.4442	0.5669	0.0932	0.1639	0.2315	0.1465	0.2506	0.3054	0.2746	0.4619	0.5656
MINDS	0.3214	0.4978	0.6250	0.1096	0.1989	0.2766	0.1956	0.3087	0.3641	0.2819	0.4760	0.5790
THGNets	**0.3405**	**0.5212**	**0.6589**	**0.1171**	**0.2237**	**0.3023**	**0.2120**	**0.3445**	**0.3945**	**0.2876**	**0.4821**	**0.6122**

Table 2: Comparison experiments on four public datasets (*Hits@k*, *k*=10,50,100)

Models	Christianity			Android			Douban			Memetracker		
	10	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**	**10**	**50**	**100**
Topo-LSTM	0.0523	0.0619	0.0635	0.0166	0.0202	0.0231	0.0354	0.0824	0.0884	0.0870	0.0955	0.0969
NDM	0.0144	0.0177	0.0182	0.0059	0.0070	0.0072	0.0141	0.0824	0.0884	0.0463	0.0480	0.0481
SNIDSA	0.0246	0.0306	0.0326	0.0100	0.0122	0.0129	0.0371	0.0419	0.0148	0.0605	0.0674	0.0689
Inf-VAE	0.0172	0.0254	0.0272	0.0076	0.0103	0.0110	0.0543	0.0588	0.0598	0.0425	0.0509	0.0525
DyHGCN	0.1062	0.1167	0.1184	0.0392	0.0434	0.0446	0.0801	0.0856	0.0865	0.1410	0.1502	0.1518
TAN-DRUD	0.0752	0.1167	0.1184	0.0099	0.0130	0.0139	0.0359	0.0401	0.0409	0.0991	0.1086	0.1102
FOREST	0.1569	0.1658	0.1672	0.0628	0.0667	0.0675	0.0655	0.0694	0.0702	0.1429	0.1514	0.1528
DMT-LIC	0.1649	0.1728	0.1746	0.0622	0.0652	0.0662	0.0812	0.0856	0.0897	0.1496	0.1581	0.1595
MINDS	0.1955	0.2037	0.2054	0.0677	0.0716	0.0727	0.1142	0.1199	0.1213	0.1535	0.1623	0.1638
THGNets	**0.2056**	**0.2102**	**0.2152**	**0.0717**	**0.0756**	**0.0781**	**0.1202**	**0.1265**	**0.1310**	**0.1612**	**0.1635**	**0.1704**

Table 3: Comparison experiments on four public datasets (*Map@k*, *k*=10,50,100)

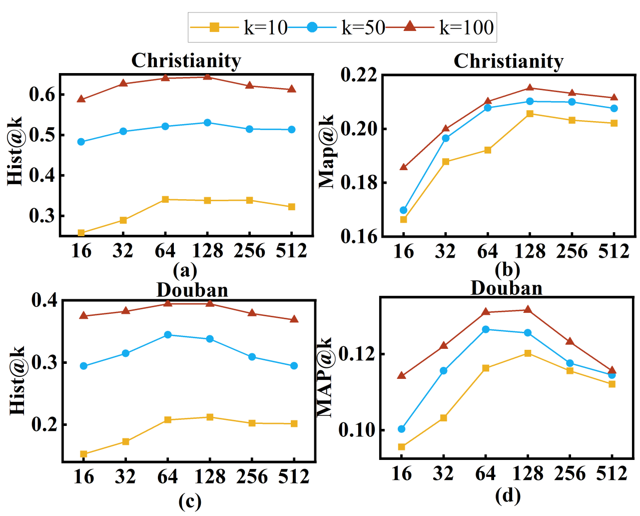

Figure 3: The embedding dimension

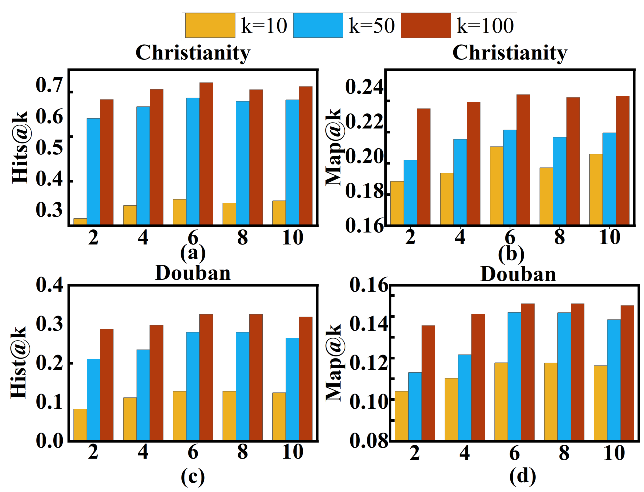

Figure 4: The number of cascade interval

ing in hyperbolic space, alleviating the distortion of the user features and enhancing the ability of user features expression. Additionally, we design a time-consistency constraint to alleviate the instability and non-smoothness of user features in latent space. Besides, we devise hyperbolic GRUs to mine potential dependencies between contextual users in hyperbolic space. Experiments conducted on four public datasets demonstrate that our model significantly outperforms existing methods, validating the effectiveness of hyperbolic geometry for information diffusion prediction. In the future, we will mainly focus on adopting different curvatures in different layers.

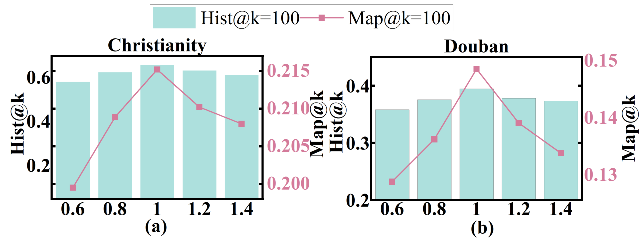

Figure 5: The curvature

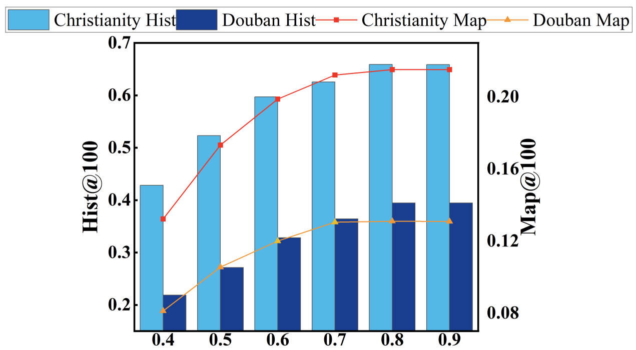

Figure 6: The proportion of training

Related Works

Infromation Diffusion Prediction. Previous studies mainly divide into three categories: **Independent cascades-based methods** contain independent cascade (IC) (Saito et al. 2009) (Saito, Nakano, and Kimura 2008) and linear threshold (LT) (Bourigault, Lamprier, and Gallinari 2016) (Rodriguez, Balduzzi, and Schölkopf 2011), as predefined models, assuming that users independently retweet information. However, such methods are not in line with the complex real-world environment. **Embedding-based methods** represent the user characteristics with parameterized vectors and predict the probabilities of the next infected users by calculating the distance or similarity between the embedding vectors of the users (Bourigault, Lamprier, and Gallinari 2016) (Feng et al. 2018) (Gao et al. 2017). Although embedding vectors can contain rich personalized features, these methods do not account for historical interaction behavior among users, limiting model performance. **Deep learning representation method.** Recently, deep learning with its powerful implicit feature extraction capabilities has been widely applied in information diffusion prediction. DeepCas(Li et al. 2017) is the first end-to-end framework. Building upon it, researchers further discuss DeepDiffuse(Islam et al. 2018), Topo-LSTM(Wang et al. 2017), SNIDSA(Wang, Chen, and Li 2018), NDM(Yang et al. 2018), etc., which primarily use recurrent neural networks and attention mechanisms to learn user features. However, the underlying topological structure and the homogeneity of the users are not utilized sufficiently, leading to degraded performance. To address this dilemma, researchers further establish FOREST(Yang et al. 2021a) and Inf-VAE(Sankar et al. 2020) models with graph structure learners to effectively capture the underlying topological structure and the homogeneity of the user. In addition, DyHGCN(Yuan et al. 2021), MS-HGAT(Sun et al. 2022), and MINDS(Jiao et al. 2024) are introduced to model the dynamic interaction relationship.The above methods have achieved promising results.

Hyperbolic Graph Representation Learning. Real social networks have non-Euclidean hierarchical structures, which effectively deal with Riemann hyperbolic geometric, such as the Poincaré ball (Nickel and Kiela 2017) and Lorentz (Nickel and Kiela 2018) models. Inspired by graph convolutional networks, researchers boast hyperbolic graph neural networks, such as HGCN (Chami et al. 2019) and HGAT (Zhang et al. 2021). RotDiff (Qiao et al. 2023) applies the Lorentz model to embed user features in a hyperbolic space.

Datasets

Douban comes from a new community website, which provides various services such as book recommendations and making friends based on common interests. Users can share the status of reading and follow other users. The co-occurrence relationship is regarded as the social relationship between users.

Android comes from Stack-Exchanges, which contains questions from different fields. The interaction patterns across various channels constitutes the social relationships between users.

Christianity comes from Stack-Exchanges, which collects information cascades related to the Christianity topic and user social relationship.

Memetracker collects 1 million news stories and blogs to analyze the migration of common memes among users, in which each meme is represented as an information entity, and each URL is considered as a user entity.

Baseline Models

Topo LSTM (Wang et al. 2017) extends the standard LSTM to Topo LSTM.

NDM (Yang et al. 2019) combines self-attention and CNN.

SNIDSA (Wang, Chen, and Li 2018) develops a sequence neural network with structured attention.

Inf-VAE (Sankar et al. 2020) introduces a variational autoencoder framework to jointly embed user homogeneity and temporal features.

FOREST (Yang et al. 2021a) integrates contextual topological structure into recurrent neural networks.

DyHGCN (Yuan et al. 2021) raises a novel dynamic heterogeneous GCN framework that jointly learns the topological structural on social networks and cascade graphs.

DMT-LIC (Chen et al. 2019) designs a shared presentation layer for learning the underlying topologies and user sequences.

TAN-DRUD (Liu et al. 2022) provides a novel neural network framework to capture dual user roles.

MINDS (Jiao et al. 2024) discusses hypergraph and GCN to model information cascades and user homogeneity, respectively.

Related Concepts

Information cascade. The infected users are arranged by forwarding time t, generating the information cascade as $s_m = \{(u_{m,0}, t_0), (u_{m,1}, t_1), ..., (u_{m,i}, t_i), ..., (u_{n-1}, t_{n-1})\}$, where $u_{m,i} \in U$. Information cascade set S contains multiple information cascades s_m.

Information cascade subgraph. Given an information cascade graph s_m and subgraphs intervals K, we divide the information cascade graph into several subgraphs, also named cascade snapshot, $s_m^k = (U_{m,k}, \varepsilon_{m,k}), k \in (0, K], s_m^k \in s_m$.

Social networks. Social network $G = (U, E)$ is a complete static graph with social relationships between users, the tuple $(u_i, u_j) \in E$ denotes a relationship between the user u_i and the user u_j.

Poincare Balls Geometry

Poincare balls. The Poincare Balls is the Riemann manifold $(D^{n,c}, g_x^c)$ with negative curvature $-c\,(c > 0)$, where $c = 0$ denotes the European space, $D^{n,c} \in R^n$. The expression of the formula of the Riemann manifold is defined as following. g_x^c denotes the metric tensor of the Riemann manifold, $\lambda_x^c = \frac{2}{1-c\|x\|^2}$, I_d is the identity matrix.

$$D^{n,c} = \left\{ x \in R^n : \|x\|^2 \frac{1}{c} \right\}, g_x^c = (\lambda_x^c)\, I_d, \tag{19}$$

Möbius addition. Given any point $x, y \in D^{n,c}$ in hyperbolic space, the Möbius addition operation $\oplus_c$ is defined as follows. $\langle\cdot\rangle$, $\|\cdot\|$ represents the inner product and the regularization in Euclidean space, respectively.

$$x \oplus_c y = \frac{\left(1 + 2c\langle x, y\rangle + c\|y\|^2\right) x + \left(1 - c\|x\|^2\right) y}{1 + 2c\langle x, y\rangle + c^2\|x\|^2\|y\|^2}, \tag{20}$$

Hyperbolic matrix-vector multiplication. Given $x \in D^{n,c}$ and a weight matrix M, the matrix multiplication operation is defined in the hyperbolic space as follows.

$$M \otimes_c x = exp_o^c\left(M \cdot log_0^c x\right), \tag{21}$$

Exponential and logarithmic mapping. The exponential function aims to map feature representations from Euclidean space to hyperbolic space. The logarithmic function aims to map feature representations from hyperbolic space to Euclidean space. The function of exponential and logarithmic maps is defined as follows. The function $proj^c(y)$ restricts the feature representation within the sphere. x, y represents any point in the hyperbolic space, $x, y \in D^{n,c}$.

$$\exp_o^c(y) = o \oplus_c \left(\tanh\left(\sqrt{|c|} \frac{\lambda_x^c \|y\|_2}{2} \right) \frac{v}{\sqrt{|c|}\, \|y\|_2} \right), \tag{22}$$

$$\log_o^c(y) = \frac{2}{\sqrt{|c|}\lambda_o^c} \tanh^{-1}\left(\sqrt{|c|}\, \|{-o} \oplus_c y\| \right) \frac{-o \oplus_c y}{\|{-o} \oplus_c y\|}, \tag{23}$$

$$proj^c(y) := \begin{cases} \frac{y}{\|y\|_2} \cdot norm_{\max}, \|y\|_2 \geq norm_{\max} \\ y, otherwise \end{cases} \tag{24}$$

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No.2022YFC3302101).

References

Bourigault, S.; Lamprier, S.; and Gallinari, P. 2016. Representation learning for information diffusion through social networks: an embedded cascade model. In *Proceedings of the Ninth ACM international conference on Web Search and Data Mining*, 573–582.

Chami, I.; Ying, Z.; Ré, C.; and Leskovec, J. 2019. Hyperbolic graph convolutional neural networks. *Advances in neural information processing systems*, 32.

Chen, W.; Wu, Y.; Zhang, Z.; Zhuang, F.; He, Z.; Xie, R.; and Xia, F. 2024a. FairGap: Fairness-aware Recommendation via Generating Counterfactual Graph. *ACM Transactions on Information Systems*, 42(4): 1–25.

Chen, W.; Yuan, M.; Zhang, Z.; Xie, R.; Zhuang, F.; Wang, D.; and Liu, R. 2024b. FairDgcl: Fairness-aware Recommendation with Dynamic Graph Contrastive Learning. *arXiv preprint arXiv:2410.17555*.

Chen, X.; Zhang, K.; Zhou, F.; Trajcevski, G.; Zhong, T.; and Zhang, F. 2019. Information cascades modeling via deep multi-task learning. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, 885–888.

Feng, S.; Cong, G.; Khan, A.; Li, X.; Liu, Y.; and Chee, Y. M. 2018. Inf2vec: Latent representation model for social influence embedding. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 941–952. IEEE.

Gao, S.; Pang, H.; Gallinari, P.; Guo, J.; and Kato, N. 2017. A novel embedding method for information diffusion prediction in social network big data. *IEEE Transactions on Industrial Informatics*, 13(4): 2097–2105.

He, G.; Xue, Z.; Jiang, Z.; Kang, Y.; Zhao, S.; and Lu, W. 2023. H2CGL: Modeling dynamics of citation network for impact prediction. *Information Processing & Management*, 60(6): 103512.

Islam, M. R.; Muthiah, S.; Adhikari, B.; Prakash, B. A.; and Ramakrishnan, N. 2018. Deepdiffuse: Predicting the'who'and'when'in cascades. In *2018 IEEE international conference on data mining (ICDM)*, 1055–1060. IEEE.

Jiao, P.; Chen, H.; Bao, Q.; Zhang, W.; and Wu, H. 2024. Enhancing Multi-Scale Diffusion Prediction via Sequential Hypergraphs and Adversarial Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8571–8581.

Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Memetracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 497–506.

Li, C.; Ma, J.; Guo, X.; and Mei, Q. 2017. Deepcas: An end-to-end predictor of information cascades. In *Proceedings of the 26th international conference on World Wide Web*, 577–586.

Liu, B.; Yang, D.; Shi, Y.; and Wang, Y. 2022. Improving information cascade modeling by social topology and dual role user dependency. In *International Conference on Database Systems for Advanced Applications*, 425–440. Springer.

Meng, X.; and Zhou, B. 2023. Scale-free networks beyond power-law degree distribution. *Chaos, Solitons & Fractals*, 176: 114173.

Nickel, M.; and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30.

Nickel, M.; and Kiela, D. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, 3779–3788. PMLR.

Oro, E.; Pizzuti, C.; Procopio, N.; and Ruffolo, M. 2017. Detecting topic authoritative social media users: a multilayer network approach. *IEEE Transactions on Multimedia*, 20(5): 1195–1208.

Qiao, H.; Feng, S.; Li, X.; Lin, H.; Hu, H.; Wei, W.; and Ye, Y. 2023. RotDiff: A hyperbolic rotation representation model for information diffusion prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 2065–2074.

Rodriguez, M. G.; Balduzzi, D.; and Schölkopf, B. 2011. Uncovering the temporal dynamics of diffusion networks. *arXiv preprint arXiv:1105.0697*.

Saito, K.; Kimura, M.; Ohara, K.; and Motoda, H. 2009. Learning continuous-time information diffusion model for social behavioral data analysis. In *Advances in Machine Learning: First Asian Conference on Machine Learning, ACML 2009, Nanjing, China, November 2-4, 2009. Proceedings 1*, 322–337. Springer.

Saito, K.; Nakano, R.; and Kimura, M. 2008. Prediction of information diffusion probabilities for independent cascade model. In *International conference on knowledge-based and intelligent information and engineering systems*, 67–75. Springer.

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 *Proceedings of the 13th international conference on web search and data mining*, 510–518.

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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 4156–4164.

Wang, J.; Zheng, V. W.; Liu, Z.; and Chang, K. C.-C. 2017. Topological recurrent neural network for diffusion prediction. In *2017 IEEE international conference on data mining (ICDM)*, 475–484. IEEE.

Wang, Z.; Chen, C.; and Li, W. 2018. A sequential neural information diffusion model with structure attention. In *Proceedings of the 27th ACM international conference on information and knowledge management*, 1795–1798.

Yang, C.; Sun, M.; Liu, H.; Han, S.; Liu, Z.; and Luan, H. 2018. Neural diffusion model for microscopic cascade prediction. *arXiv preprint arXiv:1812.08933*.

Yang, C.; Sun, M.; Liu, H.; Han, S.; Liu, Z.; and Luan, H. 2019. Neural diffusion model for microscopic cascade study. *IEEE Transactions on Knowledge and Data Engineering*, 33(3): 1128–1139.

Yang, C.; Wang, H.; Tang, J.; Shi, C.; Sun, M.; Cui, G.; and Liu, Z. 2021a. Full-scale information diffusion prediction with reinforced recurrent networks. *IEEE Transactions on Neural Networks and Learning Systems*, 34(5): 2271–2283.

Yang, M.; Zhou, M.; Kalander, M.; Huang, Z.; and King, I. 2021b. Discrete-time temporal network embedding via implicit hierarchical learning in hyperbolic space. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 1975–1985.

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 *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part III*, 347–363. Springer.

Zhang, Y.; Wang, X.; Shi, C.; Jiang, X.; and Ye, Y. 2021. Hyperbolic graph attention network. *IEEE Transactions on Big Data*, 8(6): 1690–1701.

Zhong, E.; Fan, W.; Wang, J.; Xiao, L.; and Li, Y. 2012. Comsoc: adaptive transfer of user behaviors over composite social network. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 696–704.