节点流行度预测

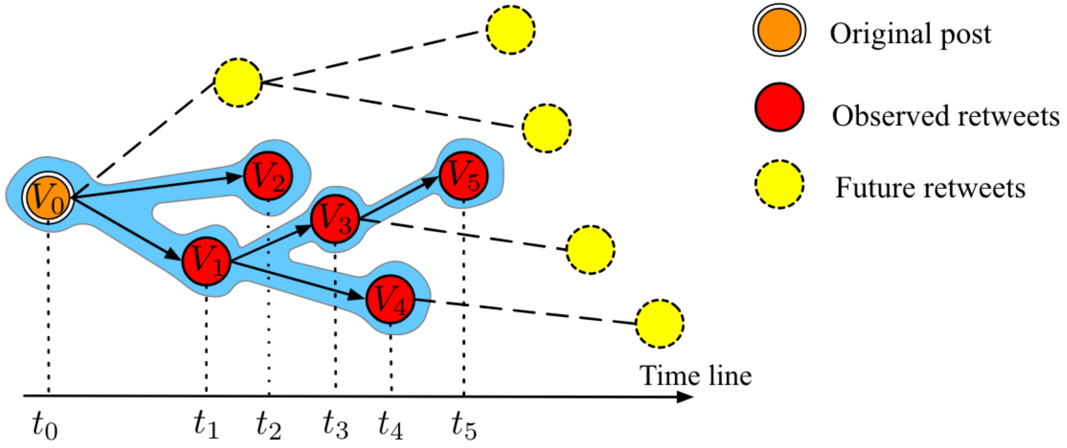
北航计算机学院 姬硕

2023年6月16日

问题概述

Node Popularity Prediction

- Forecast **how many users** would like to interact with a target item or online content in the future
- Online Social Network, Paper Citation Network…

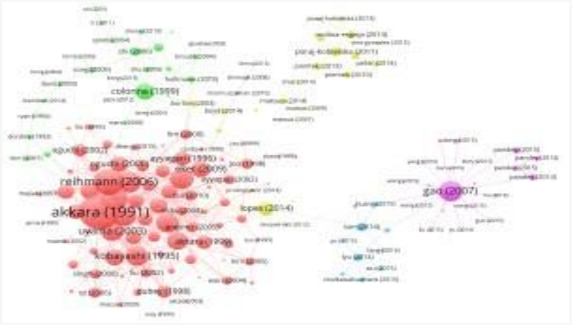


问题概述

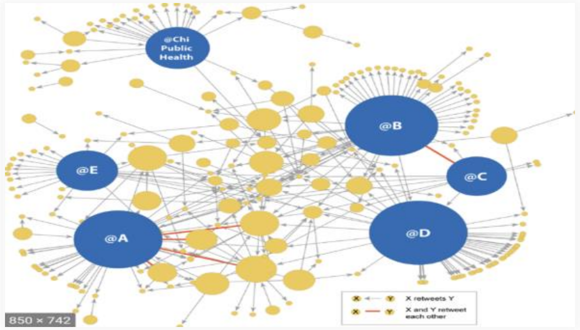
Node Popularity Prediction

- Help online shopping or social media platforms to **identify popular items or digital content**

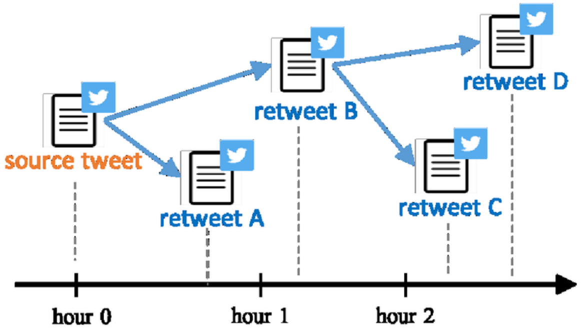
Citation of publication



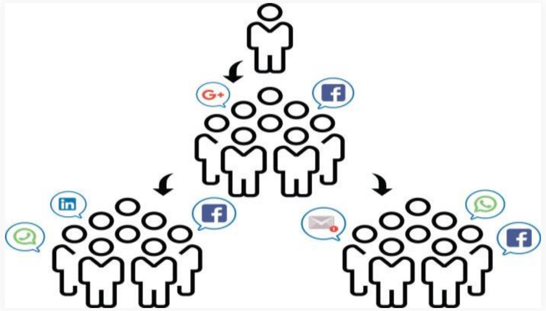
Forwarding of Weibo



Retweet of a tweet



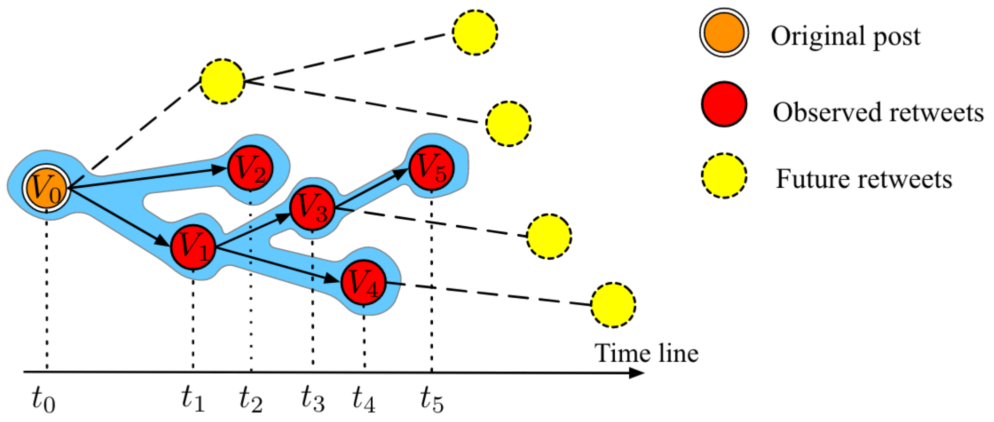
Viral marketing



问题概述

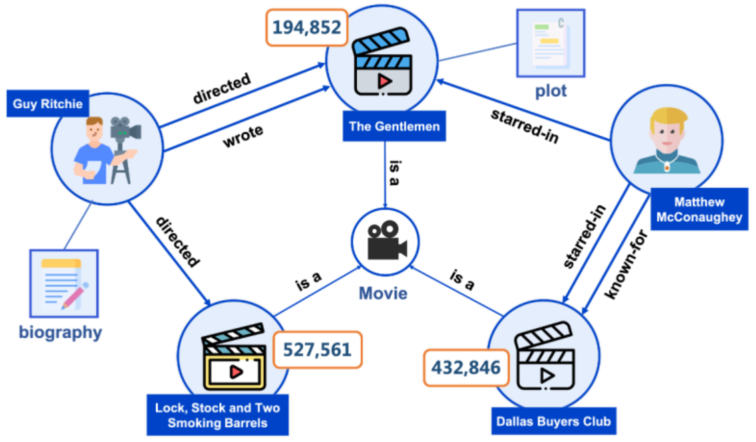
Node Popularity Prediction

- A task to forecast **how many users** would like to interact with a target item or online content in the future
- Online Social Network, Paper Citation Network...



Node Importance Estimation

- A task of inferring the **significance or the popularity** of a node in a graph according to the structure and attribute information
- Knowledge Graph



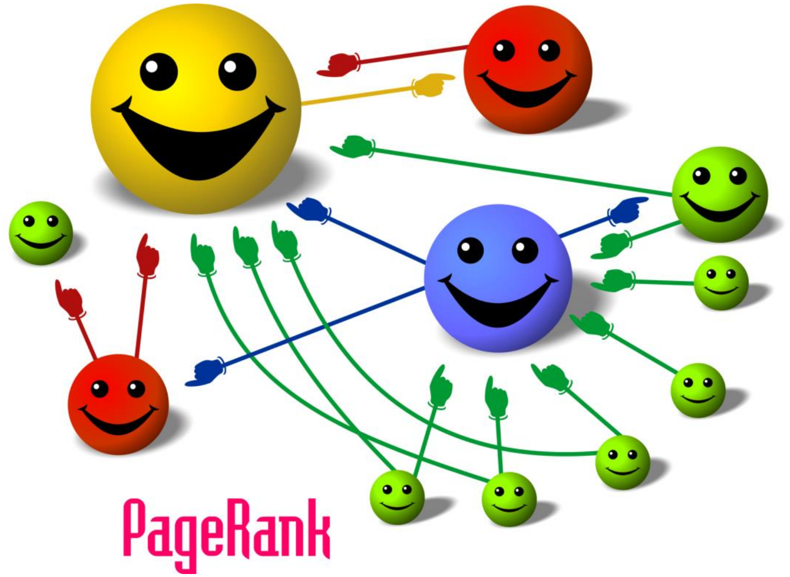
Node Importance Estimation

■ Hand-designed methods

Hand-designed methods

- Degree Centrality（度中心性）
- Betweenness Centrality（介数中心度）
- …

- prior assumptions
- no learnable parameters



Node Importance Estimation

■ Machine-learned methods

Hand-designed methods → Machine-learned methods

Node Features + Machine Learning Model

- 性能高度依赖于提取的节点特征
- 无法捕获图中丰富的信息

Node Importance Estimation

■ Graph Representation Learning Methods

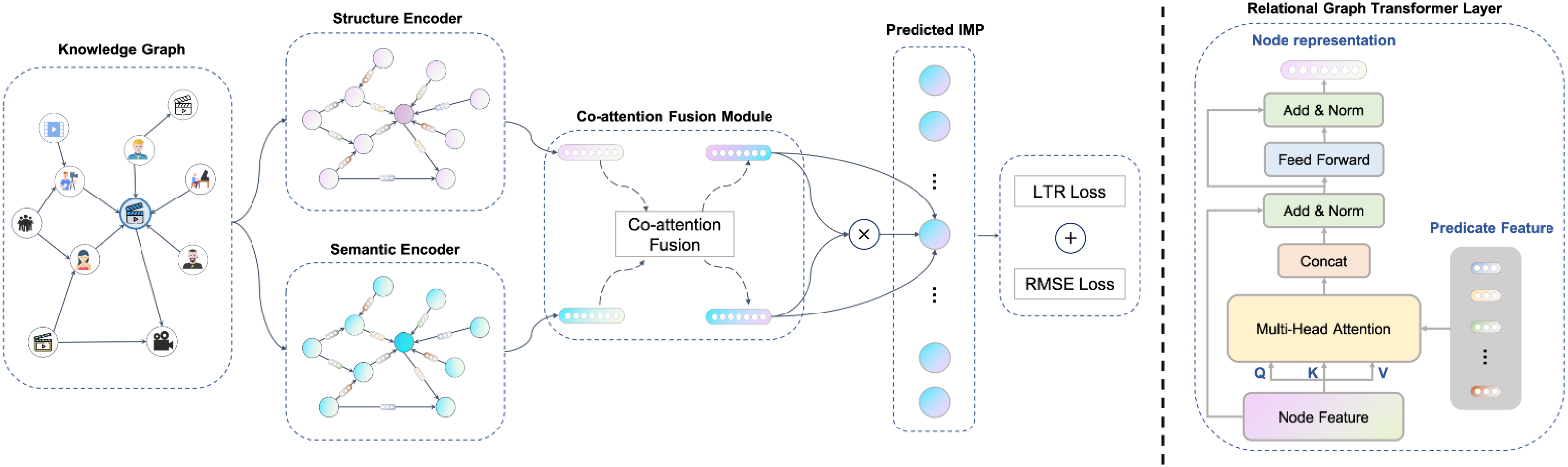
DEEP LEARNING！！！

Hand-designed Methods → Machine-learned Methods → Graph Representation Learning Methods

- 自动地学习节点的表示
- 提取图结构信息

Node Importance Estimation

Graph Representation Learning Methods



- Aggregate messages of neighbors
- Leverage abundant structural and semantic information

Huang H, Sun L, Du B, et al. Representation Learning on Knowledge Graphs for Node Importance Estimation. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 646-655.

Node Importance Estimation

Related Work

Hand-designed Methods → Machine-learned Methods → Graph Representation Learning Methods

手工设计 → 自动学习

统计信息 → 图结构表示

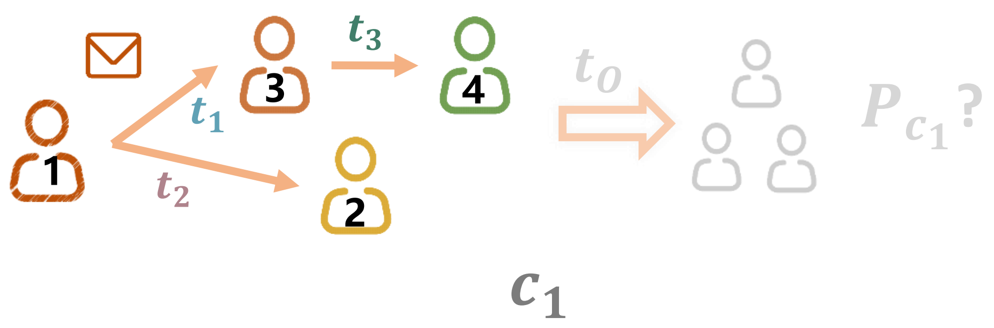
Node Popularity Prediction

Cascade

- Given a message m, the diffusion process of m on the network generates a cascade c.
- We use a chronological sequence $c(t) \; = \; (u_i^c, v_i^c, t_i^c)$ to represent the diffusion process of m before time t .

Node Popularity Prediction

- Given a cascade c begins at t_0^c, predict its incremental popularity from t_o^c to $t_o^c + t_p$.



Node Popularity Prediction

■ Feature-based Model

Feature-based Model

- Temporal Feature — Observation Time, Publication Time,…
- Structural Feature — Edge Density, Depth, Node degree…
- Additional Feature — User Information…

Hand-designed Feature + Machine Learning Model

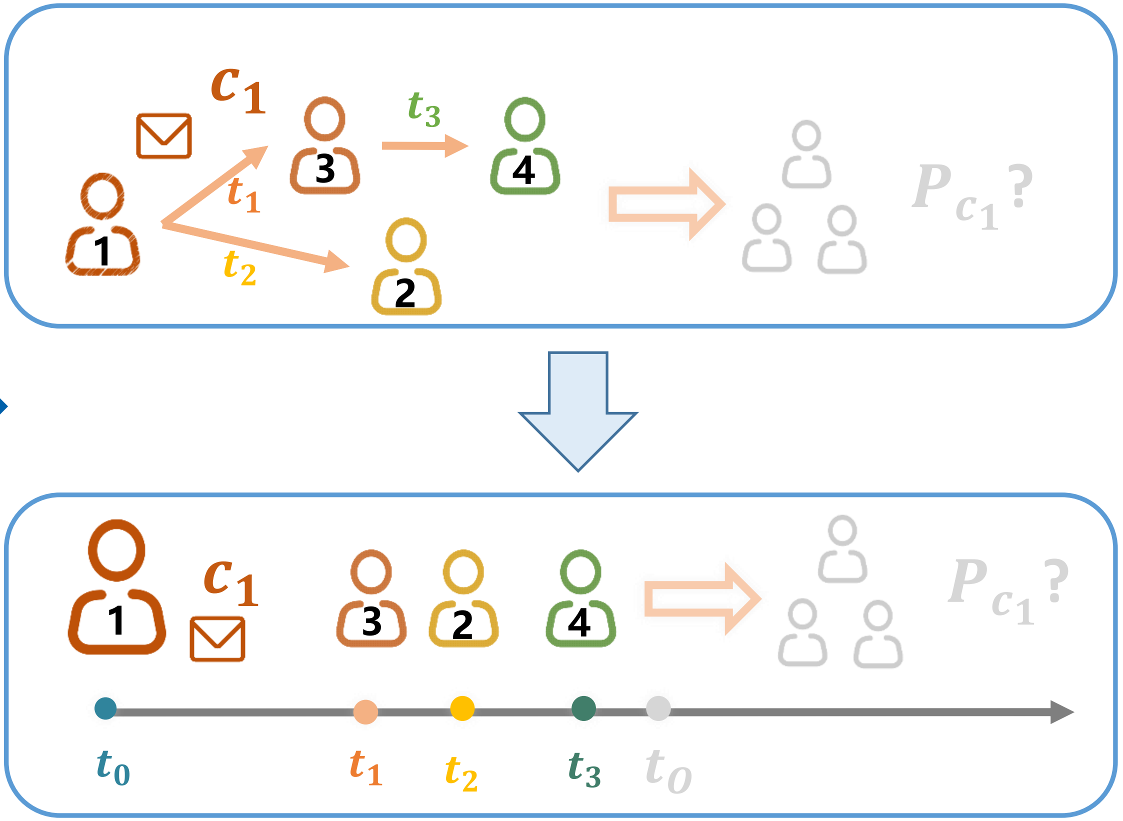
- 性能高度依赖于提取的节点特征
- 特征提取过程中丢失大量信息

Node Popularity Prediction

■ Sequential Model

DEEP LEARNING！！！

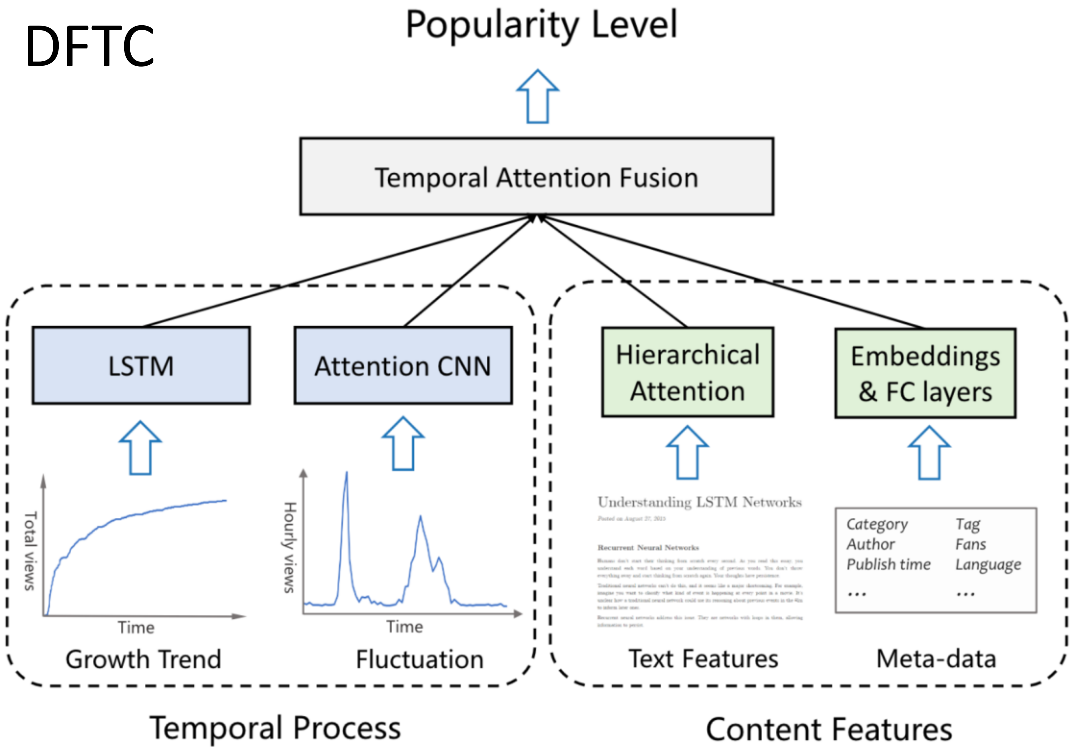
Feature-based → Sequential Model



Node Popularity Prediction

■ Sequential Model

- DFTC

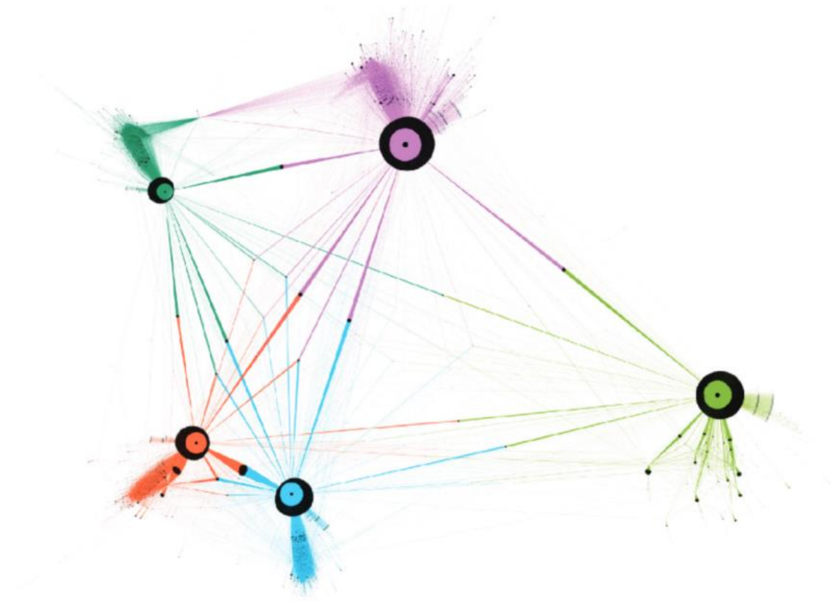


- Adopt recurrent neural network for modeling the **long term growth trend** and convolutional neural network for capturing **short term fluctuations**

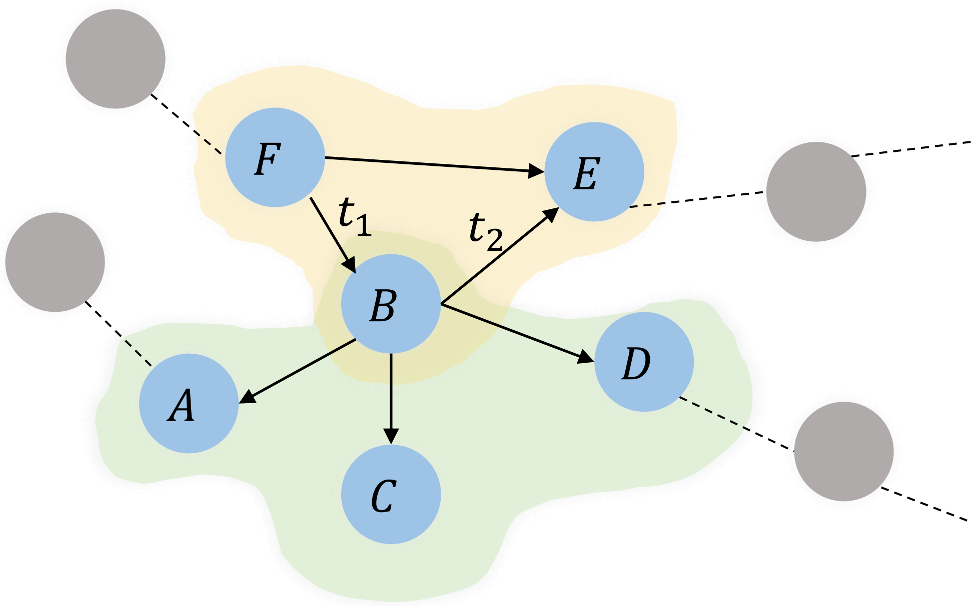
Chen X, Zhou F, Zhang K, et al. Information diffusion prediction via recurrent cascades convolution. 2019 IEEE 35th international conference on data engineering (ICDE). IEEE, 2019: 770-781.

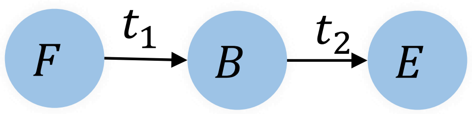
Node Popularity Prediction

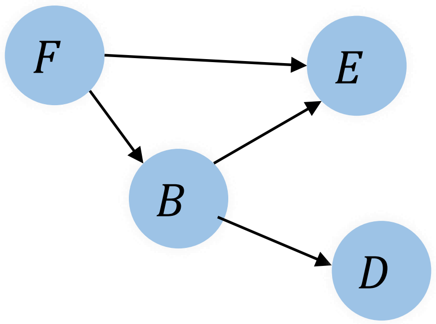
■ Sequential Model



Correlation between cascades ✗







Temporal Feature ✓

Structural Feature ✗

Node Importance Estimation

Related Work

Hand-designed Methods → Machine-learned Methods → Graph Representation Learning Methods

手工设计 → 自动学习

统计信息 → 图结构表示

Node Popularity Prediction

■ Graph Representation Learning Model

Feature-based Model

Sequential Model

Graph Representation Learning Model

!!!

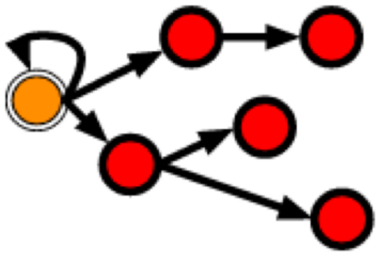
Temporal Feature ✓

Structural Feature ✓

Structural Feature

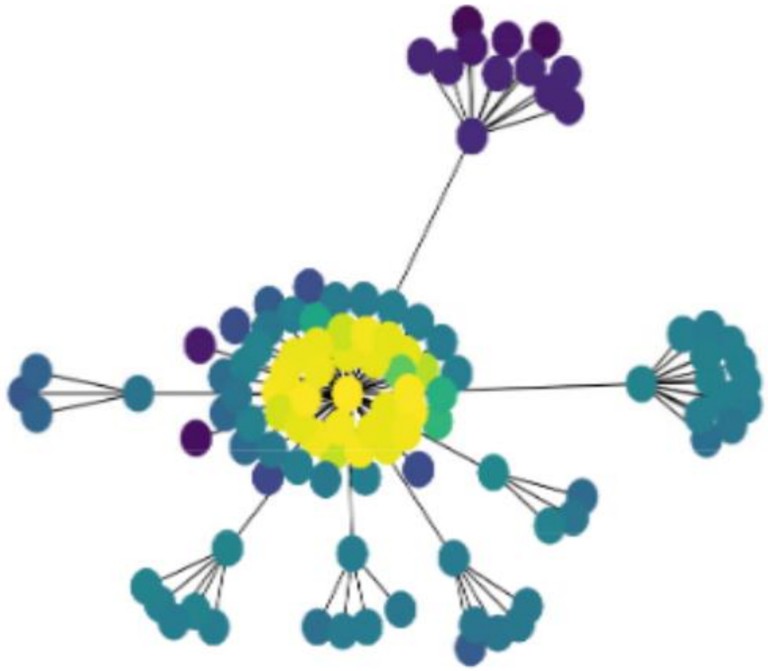
Cascade Graph

- Given an information item I_i and the corresponding cascade c_i, a **cascade graph** is defined as $\mathcal{G}_g = (\mathcal{V}_c, \mathcal{E}_c)$, where nodes $\mathcal{V}_c$ are **all participants** of cascade c_i, and matrix $\mathcal{E}_c$ contains a set of edges representing all the **relationships** between $\mathcal{V}_c$ in a cascade.



Global Graph

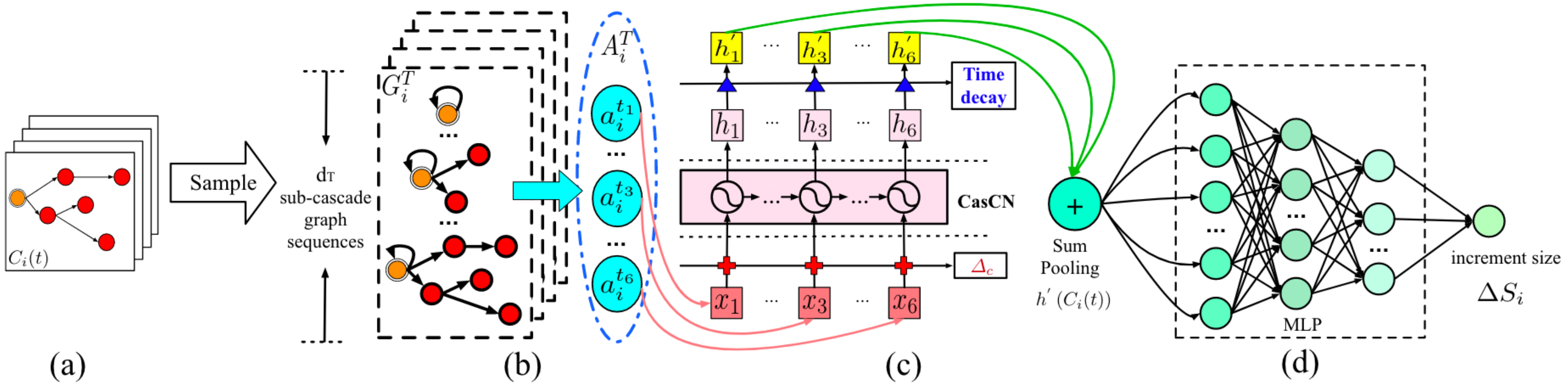
- A global graph $\mathcal{G}_g = (\mathcal{V}_g, \mathcal{E}_g)$ is **a collection of $\mathcal{V}_g$ nodes and a set of $\mathcal{E}_g \subseteq \mathcal{V}_g \times \mathcal{V}_g$ edges**.



Node Popularity Prediction

Chen X, Zhou F, Zhang K, et al. Information diffusion prediction via recurrent cascades convolution. IEEE 35th international conference on data engineering (ICDE). 2019.

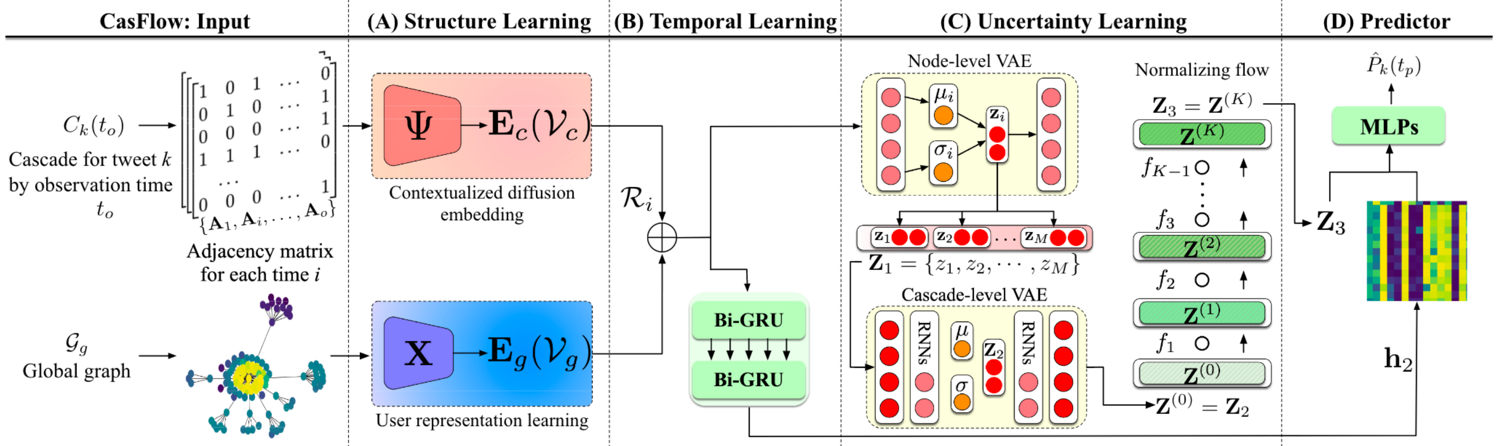
■ CasCN



• Learn the structural and temporal patterns via the combination of classical LSTM and GCN

■ CasFlow

Xovee Xu, Fan Zhou, Kunpeng Zhang, Siyuan Liu, and Goce Trajcevski. 2021. CasFlow: Exploring Hierarchical Structures and Propagation Uncertainty for Cascade Prediction. IEEE Transactions on Knowledge and Data Engineering (2021)



• Graph embedding technique and sparse matrix factorization

• Jointly models cascades from both a micro (user) and a macro (overall cascade estimating) level

Motivation

Cascade Graph

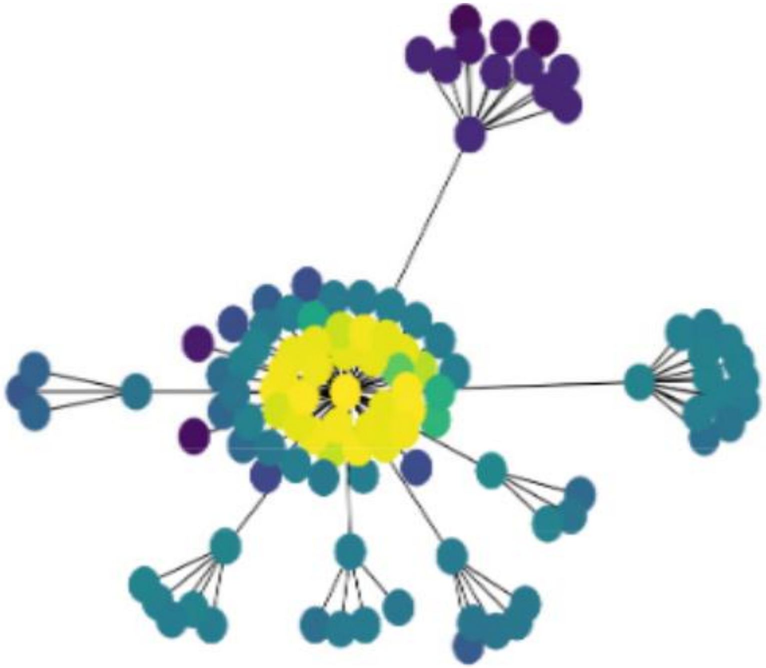
- A cascade graph is constructed based on its **participants and their interactions**.

Local Spread of Information

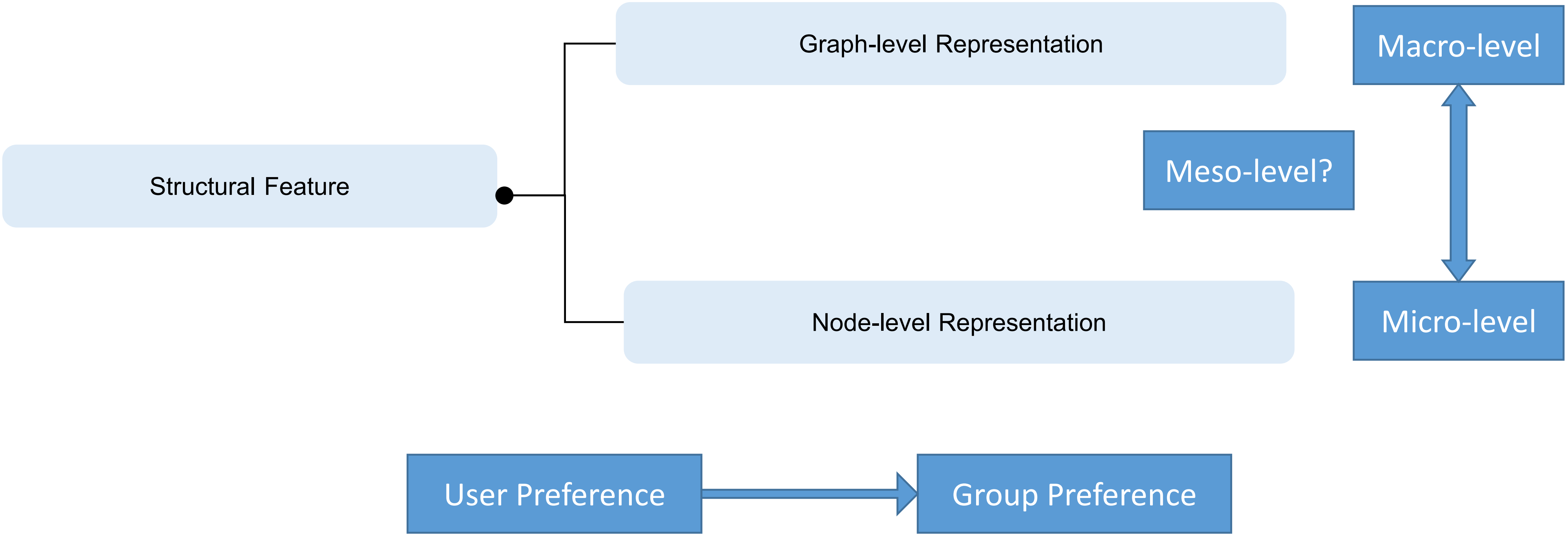
Global Graph

- A global graph is **a collection of all nodes and a set of their edges**.

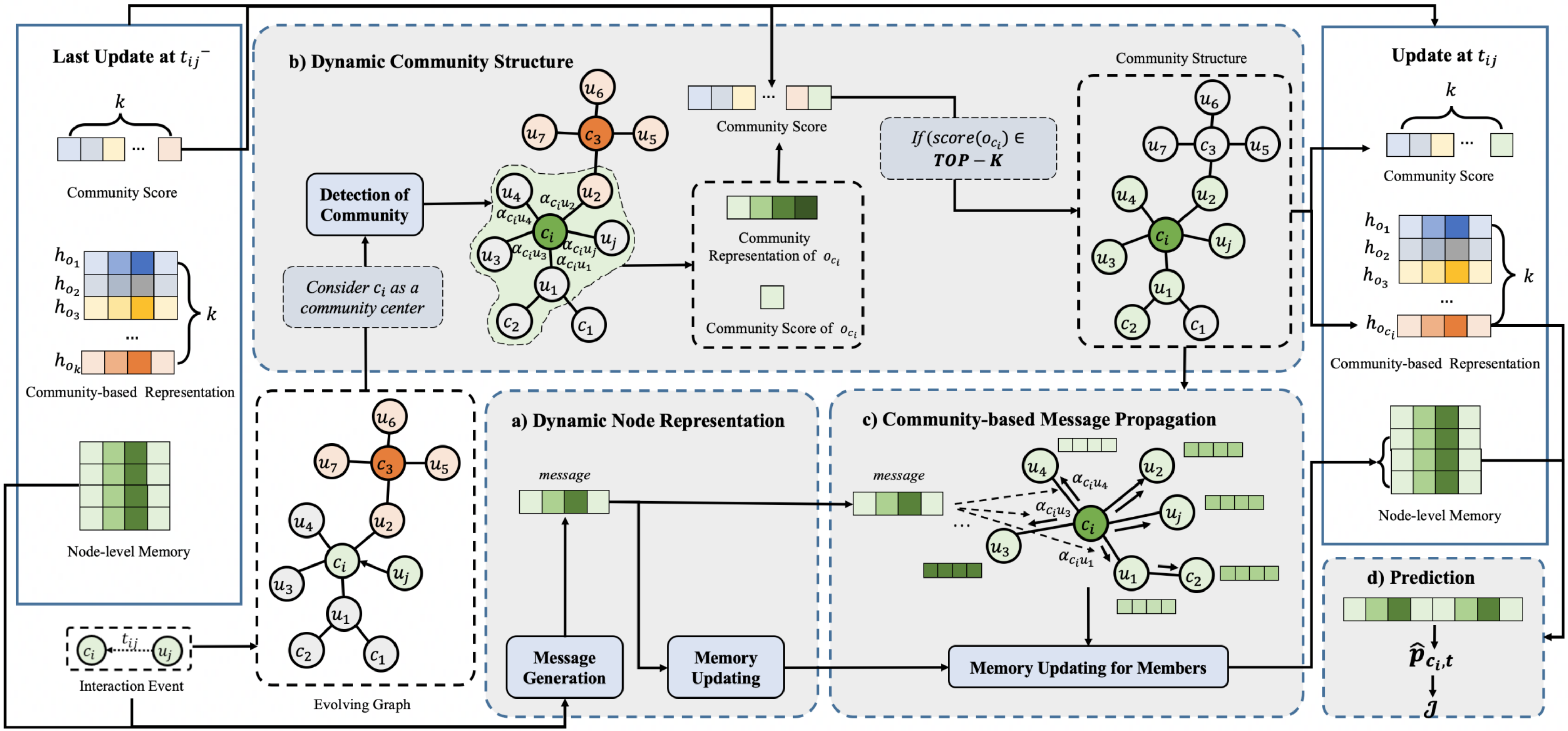
Global Information



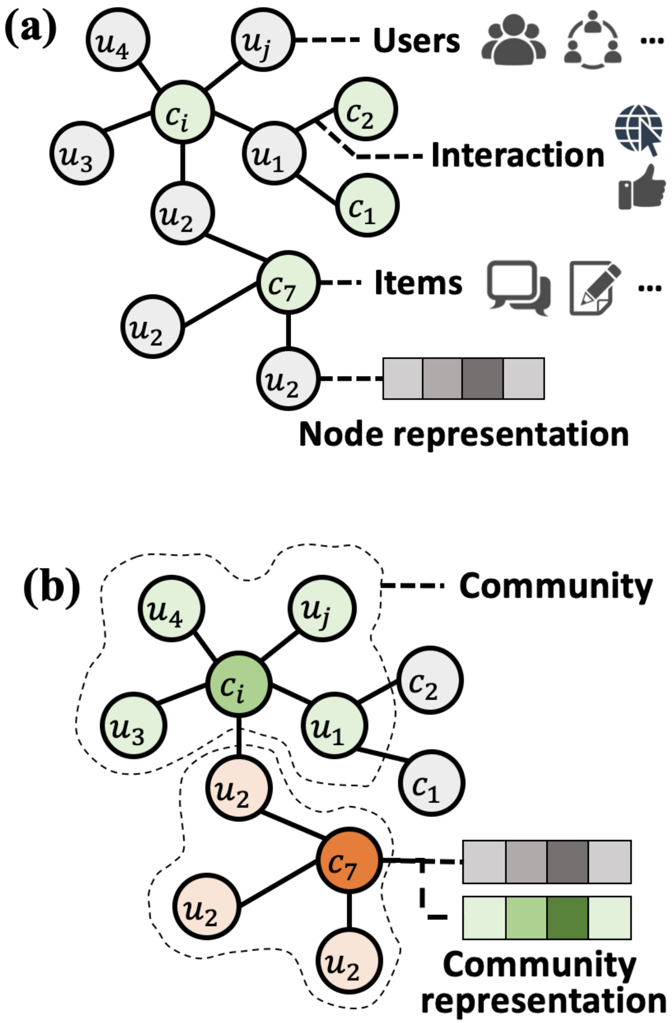
Motivation


Graph-level Representation
Macro-level
Structural Feature
Meso-level?
Node-level Representation
Micro-level
User Preference
Group Preference

Community-based Dynamic Graph Learning for Popularity Prediction

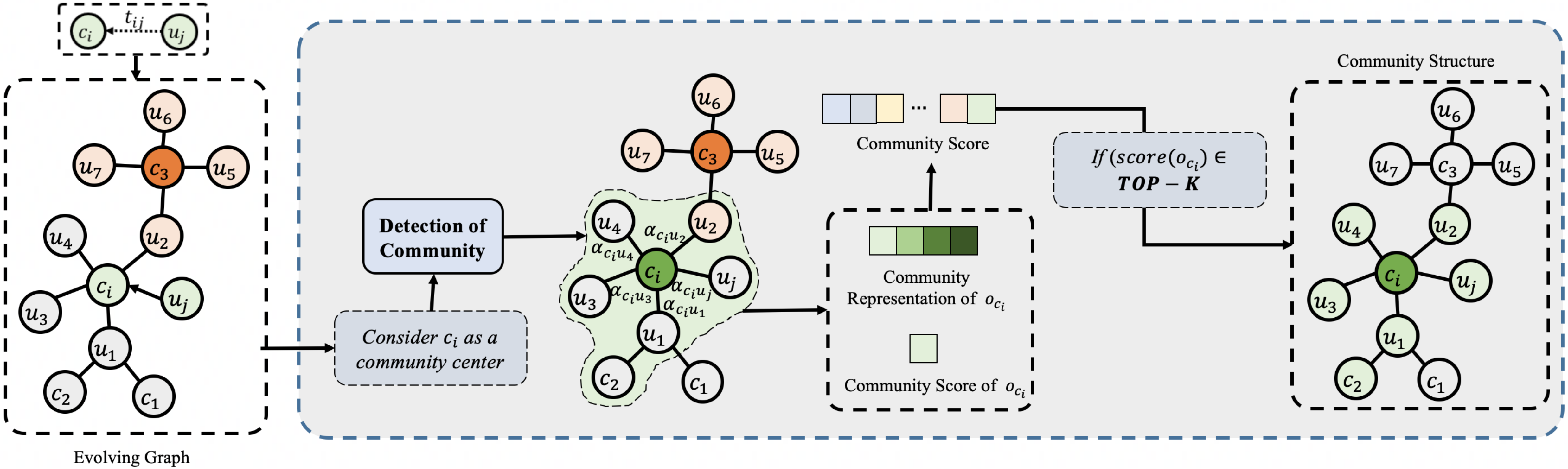


Community



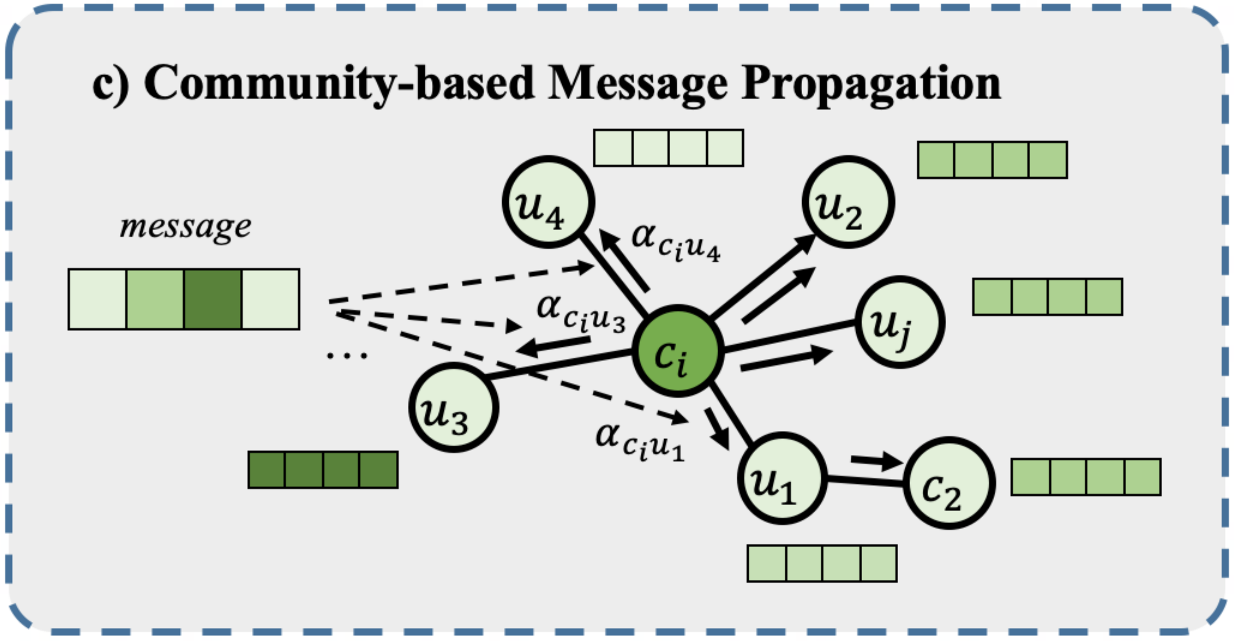
- People tends to form groups based on **similar interests, backgrounds and hobbies**, etc. Moreover, graphs in real world often contain a number of communities with **star-like structures** which consist of a popular central node connected to many peripheral nodes.
- We model network as a user-item bi-graph, and communities are formed based on **group preference**. Each community is represented by **an influential user or typical item** which act as a **community center**. People with similar interests, and items with similar topic are assigned to the community.

Community Detection



- learning both temporal group preference and high-order structural information

Community-based Message Propagation



- Propagate messages to the whole community to **spread the effect of event**
- Since users in same community have similar interests and more likely to interact with similar items

Experiment

Datasets

Table 2: Statistics of datasets.

Datasets	#Users	#Items	#Interactions
Twitter	18297	11544	3312348
AMiner	25490	8562	2104091
APS	23989	11159	1944509

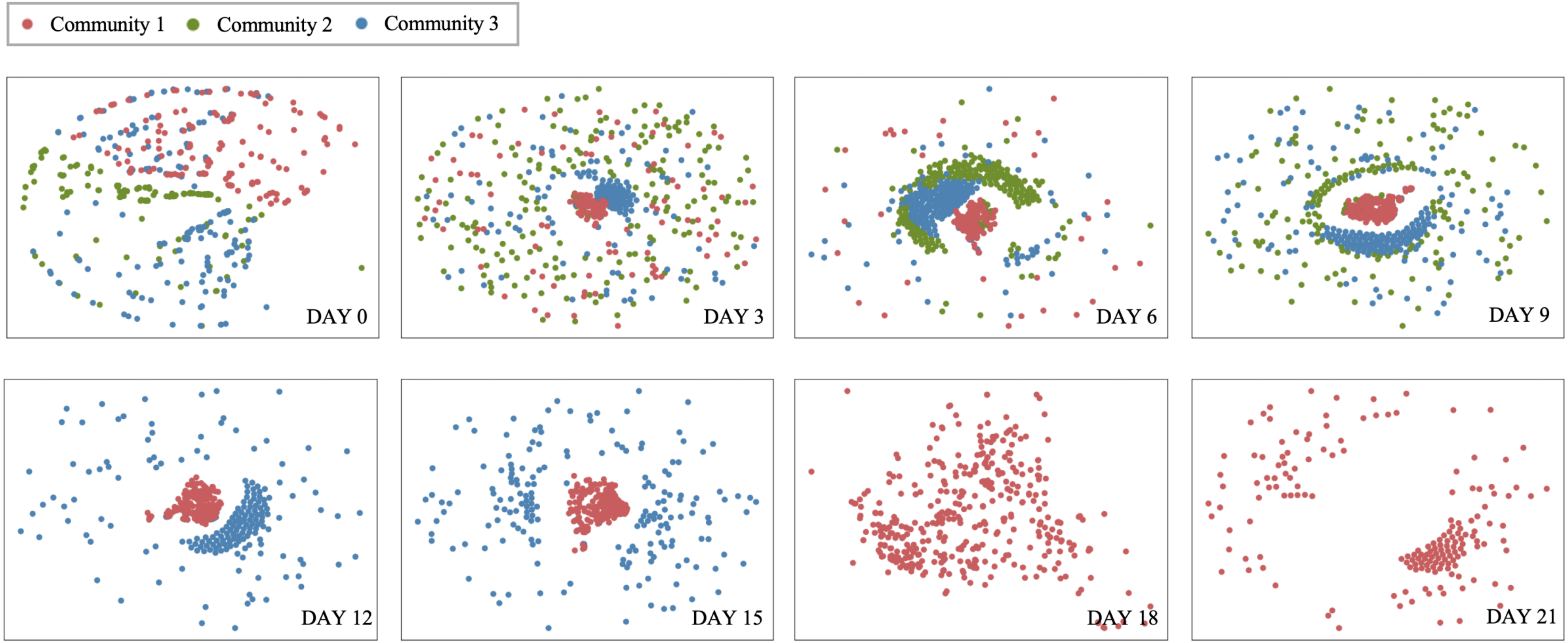
Results

Table 3: Performance of different methods.

Model	Twitter				AMiner				APS			
	MSLE	MALE	MAPE	PCC	MSLE	MALE	MAPE	PCC	MSLE	MALE	MAPE	PCC
DT	4.4247	1.6282	0.4831	0.4416	2.1208	1.1182	0.3133	0.7554	2.5917	1.2643	0.3669	0.6382
SVR	3.8866	1.6257	0.5237	0.4828	2.6203	1.2680	0.4016	0.6601	2.3908	1.2203	0.3836	0.5910
LSTM	4.2281	1.6653	0.5309	0.5056	2.3736	1.2591	0.4074	0.8215	2.2889	1.2108	0.3926	0.7298
DFTC	2.3737	1.1200	0.3611	0.6387	1.3159	0.8901	0.2903	0.8690	1.6744	1.0079	0.3256	0.7760
LightGCN	4.3951	1.4277	0.3560	0.5965	2.7766	1.2604	0.2991	0.7784	2.3054	1.1885	0.3114	0.7546
CasFlow	2.4185	1.1732	**0.3355**	0.5980	1.4883	0.9401	0.2672	0.7908	1.7894	1.0557	0.3101	0.6903
TGN	2.2789	1.1550	0.3470	0.6377	1.1302	0.8330	0.2455	0.8475	1.2763	0.8846	0.2779	0.7985
CDGP	**2.0182**	**1.0742**	0.3370	**0.7028**	**0.9574**	**0.7677**	**0.2317**	**0.8727**	**1.2256**	**0.8683**	**0.2680**	**0.8004**

Experiment

Results



Future Work

Feature-based Model

Sequential Model

Graph Representation Learning Model

Temporal Feature ✓

Structural Feature ✓

Hierarchical Information

Clustering + Graph Representation Learning

感谢聆听

节点流行度预测

北航计算机学院 姬硕

2023年6月16日