

Popularity Prediction on Social Platforms with Coupled Graph Neural Networks

Qi Cao^{1,2}, Huawei Shen^{1,2}, Jinhua Gao¹, Bingzheng Wei³, Xueqi Cheng¹

{caoqi,shenhuawei,gaojinhua,cxq}@ict.ac.cn,coltonwei@tencent.com

¹ CAS Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

² University of Chinese Academy of Sciences, Beijing, China

³ WeChat, Tencent Inc

ABSTRACT

Predicting the popularity of online content on social platforms is an important task for both researchers and practitioners. Previous methods mainly leverage demographics, temporal and structural patterns of early adopters for popularity prediction. However, most existing methods are less effective to precisely capture the *cascading effect* in information diffusion, in which early adopters try to activate potential users along the underlying network. In this paper, we consider the problem of *network-aware popularity prediction*, leveraging both early adopters and social networks for popularity prediction. We propose to capture the cascading effect explicitly, modeling the activation state of a target user given the activation state and influence of his/her neighbors. To achieve this goal, we propose a novel method, namely CoupledGNN, which uses two coupled graph neural networks to capture the interplay between node activation states and the spread of influence. By stacking graph neural network layers, our proposed method naturally captures the cascading effect along the network in a successive manner. Experiments conducted on both synthetic and real-world Sina Weibo datasets demonstrate that our method significantly outperforms the state-of-the-art methods for popularity prediction.

CCS CONCEPTS

• **Human-centered computing** → *Social networks; Social media.*

KEYWORDS

Popularity Prediction, Graph Neural Networks, Cascading Effect, Network-aware

ACM Reference Format:

Qi Cao, Huawei Shen, Jinhua Gao, Bingzheng Wei, Xueqi Cheng. 2020. Popularity Prediction on Social Platforms with Coupled Graph Neural Networks. In *The Thirteenth ACM International Conference on Web Search and Data Mining (WSDM '20)*, February 3–7, 2020, Houston, TX, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3336191.3371834>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WSDM '20, February 3–7, 2020, Houston, TX, USA

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-6822-3/20/02...\$15.00

<https://doi.org/10.1145/3336191.3371834>

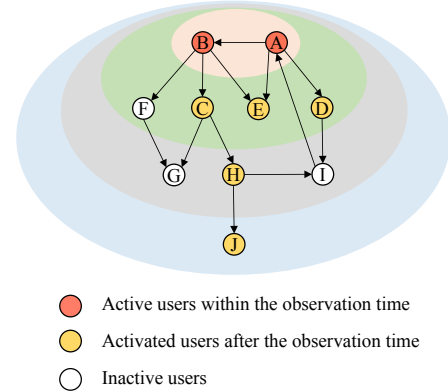


Figure 1: Example of cascading effect in information diffusion. The light orange circle represents the subgraph of early adopters. The green, gray, and blue circles represent the neighborhoods of early adopters that are reachable within one-hop, two-hops, and three-hops respectively.

1 INTRODUCTION

With the booming of social media platforms, e.g., Twitter, Facebook, Sina Weibo, Instagram, etc, the production and dissemination of user-generated online content, which we refer to as a piece of *information*, becomes extremely convenient and common in our life. Every day, there are tens of millions of information generated on these platforms [1]. With such a vast amount of information, predicting the *popularity* of pieces of information is valuable for us to discover the hot information in advance and to help people out of the dilemma of information explosion. However, due to the openness of social platforms and the cascading effect of information diffusion, it's very challenging to accurately predict the popularity of online content.

In the past decade, a series of efforts have been devoted to the popularity prediction problem in social networks, considering this problem either as a regression [2, 16, 46] or classification task [17, 33]. Generally speaking, popularity prediction aims to predict future popularity when observing early adopters at a specific observation time (see the light orange circle in Figure 1 as an example). Various hand-crafted features of early adopters are extracted to predict the future popularity, e.g., demographic features in user profile, user activity [6], user degree [14], density of the subgraph of early adopters [10, 45], as well as substructure [30] and community [4, 33]. With the success of representation learning methods,

end-to-end deep representation learning methods are also proposed to automatically learn the representation of the subgraph of early adopters [16]. In addition, to further improve the prediction performance, temporal information [21, 27, 32] of the early adopters and the content information [17, 44] are further utilized. The methods mentioned so far mainly focus on the characteristics of early adopters or the subgraph of early adopters, ignoring the cascading effect (shown in Figure 1) in information diffusions which is one of the keys to accurately predict future popularity of online content in social platforms.

To further characterize the cascading effect, researchers have also made some attempts. They adopt some statistics, such as the average number of fans of users, to approximate the impact of cascading effect in each generation [19, 40, 46]. However, since they only adopt simple statistics and regardless of the explicit network structure governing the cascading effect, they are less effective for popularity prediction.

In this paper, we focus on the *network-aware popularity prediction* problem, leveraging both early adopters and network structure for predicting the popularity of online content on social platforms. To effectively capture the crucial cascading effect, we devote to applying graph neural networks to successively characterize the activation state of each user. Specifically, the activation of a target user is intrinsically governed by two key components, i.e., the state of neighbors and the spread of influence, along social networks. As a result, we propose to model the iterative interplay between node states and the spread of influence by two coupled graph neural networks. One graph neural network models the spread of interpersonal influence, gated by the activation state of users. The other graph neural network models the update of the activation state of each user via interpersonal influence from their neighbors. With the iterative aggregation mechanism of the neighborhood in graph neural networks, the cascading effect along the network structure is naturally characterized. Note that, other information like temporal or content, if available, can be further included in the prediction model by representation fusion flexibly.

We verify the effectiveness of our proposed coupled graph neural networks on both the synthetic data and real-world data in Sina Weibo. Experimental results demonstrate that our proposed method significantly outperforms all the state-of-the-art methods. For convenience of the reproduction of the results, we have made the source code publicly available¹.

2 RELATED WORK

In this section, we briefly review the research on the popularity prediction, traditional diffusion models, the development and application of graph neural networks.

2.1 Popularity Prediction

Popularity prediction aims to predict the future popularity of online content when observing early adopters within the observation time. Due to the openness of social platforms and the cascade phenomenon of online content, future popularity results in huge variance and is challenging to predict. The predictability of particular types of information has been proved to some extent, e.g.,

tweet/microblogs [6, 18], images [44], videos [23], recipes [24], and academic papers [27].

Generally speaking, existing methods for popularity prediction mainly focus on four types of information, i.e., content, temporal information, early adopters and network structure. For content information, hierarchical attention networks [17] or user-guided hierarchical attention mechanisms [44] are proposed to characterize the content features. For temporal information, heuristical temporal features [21], time series models including recurrent neural network [34] and temporal convolutional network [26], or point process method including reinforced Poisson processes [27] and Hawkes process [19, 23, 38, 46], are proposed to devote to capture the underlying laws or patterns behind the temporal information.

As for early adopters and network structure, which is also the focus of this paper, both feature-based methods and representation learning methods are proposed. The designed effective features in the former one including node degree [14, 46], the number of nodes in the frontier graph [11], cascade density [10, 45], substructures [30], community [4, 33] and so on. Unfortunately, the performance of such methods heavily depends on the quality of the hand-crafted features, which are generally extracted heuristically. To avoid the above heuristic feature extraction process, attempts of end-to-end deep representation learning fashion are proposed to automatically learn the impact representation of early adopters by cascading effect [1, 16, 32].

However, the above methods are less effective to capture the cascading effect in information diffusion, since they neglect the explicit interactions between users along the underlying social network. In contrast, the method proposed in this paper effectively capture such cascading effect along the network structure by coupled graph neural networks.

2.2 Diffusion Models

Modeling how information diffuse is of outstanding interest over the past decades. There are two classic diffusion models in this category, i.e., Independent Cascades (IC) model [7] and Linear Threshold (LT) model [9]. The diffusion process of these models is both iteratively carried on a synchronous way along discrete time steps from initial adopters. The synchronicity assumption is further relaxed by proposing asynchronous continuous-time extensions [8, 10]. Such diffusion models can well capture the cascading effect along network by iteratively modeling the specific activation process. However, they generally need an extremely high number of Monte-Carlo simulations to estimate the final influence spread, i.e., the popularity to be predicted. Such a prediction process is time-consuming and limits its applicability to real scenarios.

The difference between our proposed method and works of this line is that we do not model the specific diffusion process, but utilize the graph neural networks to directly model the influence of cascading effect by neighborhood aggregation, which is more efficient and flexible.

2.3 Graph Neural Networks

Inspired by the huge success of neural networks in Euclidean space, recently there has been a surge of interest in graph neural network approaches for representation learning of graphs [12, 31, 36, 37].

¹<https://github.com/CaoQi92/CoupledGNN>.

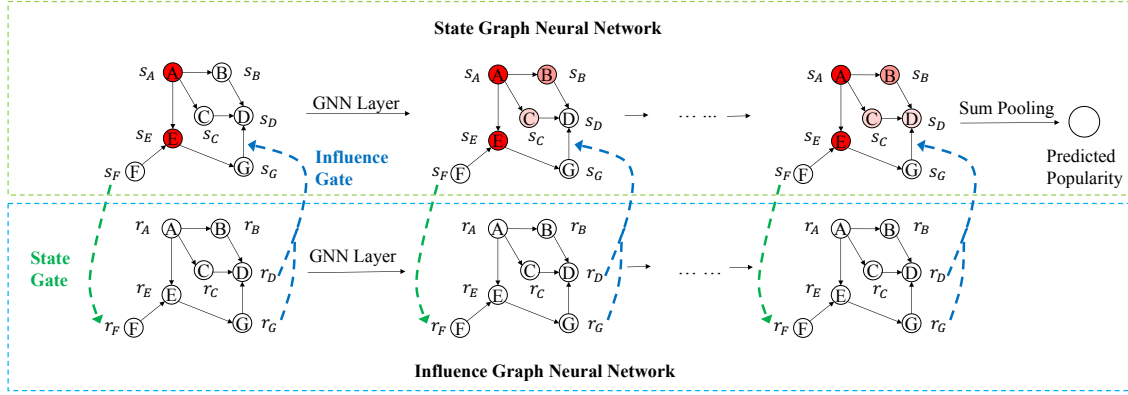


Figure 2: The framework of coupled graph neural networks for popularity prediction. s_* and r_* are the activation state and influence representation of user $*$ respectively.

Graph neural networks (GNNs) broadly follow a recursive neighborhood aggregation fashion, where each node updates its representation by aggregating the representation of its neighborhood. After K iterations of aggregation, the updated representation of each node captures both the structural and representation information within the node's K -hop neighborhood [37].

GNNs have been successfully applied to a lot of non-Euclidean domain problems, e.g., semi-supervised learning on graph [13, 35], social influence prediction [22], correlated temporal sequence modeling [25]. Among the above, the application to social influence prediction, i.e., DeepInf [22], is the most related one with our work. However, since DeepInf more focuses on the prediction of the micro action status of a user on a fixed-sized local network, rather than the macro popularity prediction on the global diffusion network studied in this paper, it performs not well for future popularity.

In this paper, we devote to utilizing GNNs to characterize the cascading effect in popularity prediction. To better adapt to the scenario of information diffusion, we design a novel model coupled graph neural networks to solve the popularity prediction problem.

3 PRELIMINARIES

This section gives the formal definition of the popularity prediction problem studied in this paper and the general framework of GNNs.

3.1 Problem Definition

Supposing that we have M pieces of information, the observed cascade of information m is recorded as the set of early adopters within the observation time window T , i.e., $C_T^m = \{u_1, u_2, \dots, u_{n_T^m}\}$, where n_T^m is the total number of adopted or active users of information m within the observation time window T . For example, the observed cascade in Figure 1 is recorded as $C_T^m = \{A, B\}$. In addition to the observed cascades, given the underlying network which governing the information diffusion, e.g., the following relationships in Sina Weibo, we can formalize the popularity prediction problem studied in this paper as:

Network-aware Popularity Prediction. Given the observed cascades C_T^m and the underlying network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of all users, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of all relationships

between users, this problem aims to predict the final popularity of information m , i.e., n_∞^m .

Note that, the network-aware popularity prediction problem emphasizes the role of the network, i.e., there are interactions between early adopters and potential active users, or among the potential active users. It's precisely because of this characteristic, making the capture of the cascading effect along the network becomes the key to accurately predict the future popularity of online content.

3.2 General Framework of GNNs

GNNs is an effective framework for representation learning of graphs. As introduced in Section 2.3, many variants of graph neural networks have been proposed. They usually follow a neighborhood aggregation strategy, where the representation of a node is updated by recursively aggregating the representation of its neighboring nodes. Formally, the k -th layer of a graph neural network is generally formulated as in [37]:

$$a_v^{(k)} = \text{AGGREGATE} \left(\left\{ h_u^{(k)} : u \in \mathcal{N}(v) \right\} \right), \quad (1)$$

$$h_v^{(k+1)} = \text{COMBINE} \left(h_v^{(k)}, a_v^{(k)} \right), \quad (2)$$

where $h_v^{(k)}$ is the feature vector of node v at the k -th layer, $\mathcal{N}(v)$ is the set of nodes which appear in the neighborhood of node v . The choice of the function $\text{AGGREGATE}(\cdot)$ and $\text{COMBINE}(\cdot)$ in GNNs is crucial.

The representation of the entire graph $h_{\mathcal{G}}$ is obtained by a READ-OUT function:

$$h_{\mathcal{G}} = \text{READOUT} \left(\left\{ h_v^{(k+1)} : v \in \mathcal{G} \right\} \right). \quad (3)$$

$\text{READOUT}(\cdot)$ can be a simple function such as summation or a more sophisticated graph-level pooling function [39, 43].

4 METHODS

In this section, we introduce the proposed coupled graph neural network (CoupledGNN) for network-aware popularity prediction. We design the CoupledGNN model to capture the *cascading effect* widely observed in information diffusion over social networks.

4.1 Framework of CoupledGNN

The cascading effect indicates that the activation of one user will trigger its neighbors in a successive manner, forming an information cascade over social networks. For a target user, whether he/she could be activated is intrinsically governed by two key components, i.e., *the state of neighbors* and *the spread of influence*, direct or indirect, over social networks. In this sense, the cascading effect is intrinsically the iterative interplay between node states and the spread of influence. Previous methods, e.g., independent cascade model, assumes a fixed-yet-unknown interpersonal influence and probes the interplay manifested as the cascading effect over the social network via Monte-Carlo simulation.

In this paper, we propose to use two coupled graph neural networks to naturally capture the cascading effect, or more specifically, the interplay of node states and the spread of influence. One graph neural network, namely state graph neural network, is used to model the activation state of nodes. The other graph neural network, namely influence graph neural network, is used to model the spread of influence over social networks. The two graph neural networks are coupled through two gating mechanisms. The framework of our coupled graph neural networks is shown in Figure 2.

4.2 State Graph Neural Network

The state graph neural network is to model the activation of each user during the cascading effect. Specifically, for a target user $v \notin C_T^m$, since he/she is usually influenced by the active users in the neighborhood $\mathcal{N}(v)$, we apply a graph neural network to model the activation state of each user (shown in Figure 3). Each user is associated with a one-dimensional value s_v , indicating the activation state of user v . Besides, since the interpersonal influence between the pair of users is generally various, we model such heterogeneous influence weight by an influence gating mechanism, i.e.,

$$\text{InfluGate}(\mathbf{r}_u^{(k)}, \mathbf{r}_v^{(k)}) = \beta^{(k)} [\mathbf{W}^{(k)} \mathbf{r}_u^{(k)} \parallel \mathbf{W}^{(k)} \mathbf{r}_v^{(k)}], \quad (4)$$

where $\mathbf{r}_u^{(k)} \in \mathbb{R}^{h^{(k)}}$ is the influence representation of user u at the k -th layer, $\mathbf{W}^{(k)} \in \mathbb{R}^{h^{(k+1)} \times h^{(k)}}$ is a weight matrix to transform the influence representation from dimension $h^{(k)}$ to $h^{(k+1)}$, $\beta^{(k)} \in \mathbb{R}^{2h^{(k+1)}}$ is a weight vector. Note that, Equation 4 is just one instance of the $\text{InfluGate}(\cdot)$ function. We can also choose other types of functions which can reflect the influence gating between the pair of users.

After obtaining the heterogeneous influence weight by influence gating function, the aggregation of the expected influence that the target user v receives from his/her neighborhood is:

$$a_v^{(k)} = \sum_{u \in \mathcal{N}(v)} \text{InfluGate}(\mathbf{r}_u^{(k)}, \mathbf{r}_v^{(k)}) s_u^{(k)} + p_v, \quad (5)$$

where $\mathcal{N}(v)$ is the neighborhood of user v , $\text{InfluGate}(\cdot) s_u^{(k)}$ is the expected influence considering the activation state $s_u^{(k)}$ of neighbor u , $p_v \in \mathbb{R}$ is a self activation parameter to reflect the probability that user v may be activated by ways out of following relationships, e.g., offline communication or browsing the hot list in the front pages. Note that, Equation 5 is actually a specific design of the AGGREGATE function as mentioned in Section 3.2.

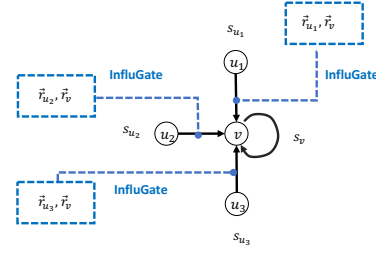


Figure 3: Mechanisms of state graph neural network.

Then the $\text{COMBINE}(\cdot)$ function used to update the activation state of user v is defined as the weighted sum of the neighborhood aggregation and the activation state of user v itself:

$$s_v^{(k+1)} = \begin{cases} 1, & v \in C_T^m \\ \sigma(\mu_s^{(k)} s_v^{(k)} + \mu_a^{(k)} a_v^{(k)}), & v \notin C_T^m \end{cases}, \quad (6)$$

where $\mu_s^{(k)}, \mu_a^{(k)} \in \mathbb{R}$ are weight parameters, σ is a nonlinear activation function. The initial activation state of user v is defined as

$$s_v^{(0)} = \begin{cases} 1, & v \in C_T^m \\ 0, & v \notin C_T^m \end{cases}. \quad (7)$$

4.3 Influence Graph Neural Network

The influence graph neural network is to model the diffusion of interpersonal influence in the social network. Specifically, each user v is associated with an influence representation \mathbf{r}_v . Then the influence representation of active users further diffuses to other users along with network structure, implemented by neighborhood aggregation of graph neural networks and a state gating mechanism. The entire mechanism of influence graph neural network is shown in Figure 4. Specifically, the neighborhood aggregation is defined as:

$$\mathbf{b}_v^{(k)} = \sum_{u \in \mathcal{N}(v)} \text{StateGate}(s_u^{(k)}) \alpha_{uv}^{(k)} \mathbf{W}^{(k)} \mathbf{r}_u^{(k)}, \quad (8)$$

where $\mathbf{r}_u^{(k)} \in \mathbb{R}^{h^{(k)}}$ is the influence representation of user u at (k) -th layer, $\mathbf{W}^{(k)} \in \mathbb{R}^{h^{(k+1)} \times h^{(k)}}$ is a weight matrix to transform the influence representation from dimension $h^{(k)}$ to $h^{(k+1)}$. $\text{StateGate}(\cdot)$ is the state gating mechanism, implemented by a 3-layer MLP in this paper to reflect the nonlinear effect of state. $\alpha_{uv}^{(k)}$ is the attention weight from user u to user v , where we adopt the formulation used in [31], i.e.,

$$e_{uv}^{(k)} = \gamma^{(k)} [\mathbf{W}^{(k)} \mathbf{r}_u^{(k)} \parallel \mathbf{W}^{(k)} \mathbf{r}_v^{(k)}], \quad (9)$$

$$\alpha_{uv}^{(k)} = \text{softmax}(e_{uv}^{(k)}) = \frac{\exp(e_{uv}^{(k)})}{\sum_{z \in \mathcal{N}(v)} \exp(e_{zv}^{(k)})}, \quad (10)$$

where $\gamma^{(k)} \in \mathbb{R}^{2h^{(k)}}$ is a weight vector.

Then the influence representation of user v at $(k+1)$ -th layer is updated by

$$\mathbf{r}_v^{(k+1)} = \sigma(\zeta_r^{(k)} \mathbf{W}^{(k)} \mathbf{r}_v^{(k)} + \zeta_b^{(k)} \mathbf{b}_v^{(k)}), \quad (11)$$

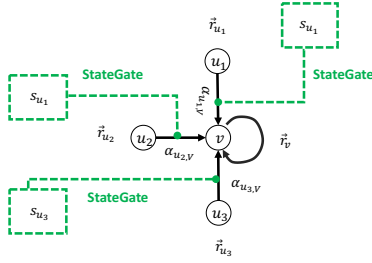


Figure 4: Mechanisms of influence graph neural network.

where $\zeta_r^{(k)}, \zeta_b^{(k)} \in \mathbb{R}$ are weight parameters, σ is a nonlinear activation function. The initial influence representation $\mathbf{r}_v^{(0)}$ of user v used in this paper consists of two parts: node embeddings and node features. We will discuss each part in detail in the section of implementation details.

4.4 Output Layer

After K layers of graph neural networks for both activation state and influence representation, the output activation probability of each user in the network is $s_v^{(K)} \in [0, 1]$, i.e., the output of the last layer in the state graph neural network. The popularity to be predicted is then obtained by a sum pooling mechanism over all users in the network, i.e.,

$$\hat{n}_\infty^m = \sum_{u \in \mathcal{V}} s_u^{(K)} \quad (12)$$

As for the loss function to be optimized, we consider the mean relative square error (MRSE) loss [2, 29], which is robust to outliers as well as smooth and differentiable:

$$L_{\text{MRSE}} = \frac{1}{M} \sum_{m=1}^M \left(\frac{\hat{n}_\infty^m - n_\infty^m}{n_\infty^m} \right)^2, \quad (13)$$

where M is the total number pieces of information, n_∞^m is the true final popularity of information m .

To avoid over-fitting and accelerate the process of convergence, we also add a L2 and user-level cross entropy to the objective function as regularization:

$$L = L_{\text{MRSE}} + L_{\text{Reg}}, \quad (14)$$

where $L_{\text{Reg}} = \eta \sum_{p \in \mathcal{P}} \|p\|_2 + \lambda L_{\text{user}}$, \mathcal{P} is the set of parameters, η and λ are hyper-parameters. L_{user} is the user-level cross entropy, i.e., $L_{\text{user}} = \frac{1}{M} \sum_{m=1}^M \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \left(s_v^\infty \log s_v^{(K)} + (1 - s_v^\infty) \log (1 - s_v^{(K)}) \right)$, s_v^∞ is the true final activation state of each user.

4.5 Computational Complexity

For the state graph neural network, the computational complexity including the influence gating mechanism at k -th layer, i.e., $O(|\mathcal{V}|h^{(k-1)}h^{(k)} + |\mathcal{E}|h^{(k)})$, and the updation of activation state, i.e., $O(|\mathcal{V}| + |\mathcal{E}|)$. For the influence graph neural network, the computational complexity at k -th layer is $O(|\mathcal{V}| + |\mathcal{V}|h^{(k-1)}h^{(k)} + |\mathcal{E}|h^{(k)})$. Sum up, the computational complexity of coupled graph neural network is $O(p|\mathcal{V}| + q|\mathcal{E}|)$, where p, q are small constant associated with the hidden dimension $h^{(k)}$ at each layer. It's worth noting that

the above computational complexity is based on the computation of the whole network. To make it more efficient, we can also address several mini-batch with R samples, which makes the algorithm independent of the graph size and achieve $O(R)$ complexity [25].

5 EXPERIMENTAL SETUP

We compare our CoupledGNN with several state-of-the-art methods on different data sets under various evaluation metrics. The detailed experimental settings are introduced in this section.

5.1 Data Sets

To thoroughly evaluate the performance of our methods, we conduct experiments on both the synthetic data set and a real-world data set from Sina Weibo.

5.1.1 Synthetic Data Set. The synthetic network is constructed by Kronecker generator [15], which can generate networks that have common structural properties of real networks, i.e., heavy tails for both in- and out-degree distributions, small diameters. The parameter matrix is set to be $[0.9; 0.5; 0.5; 0.1]$ and we retain the largest connected component as the final network, containing 1,086 nodes and 4,038 edges.

As for information cascades, we first sample the seed set of each cascade. The size of the seed set is sampled according to the power-law distribution with parameter 2.5, i.e., $p(n) \propto n^{-2.5}$ [5], and the node in each seed set is uniformly sampled. With a given seed set, the commonly used IC model [7] is applied to generate the diffusion data, where the activation probability from node u to node v is set to be $1/d_v$ and d_v is the in-degree of node v . The observation time window T is set to be 2 time steps in this scenario, i.e., we observe the diffusion process at time step $t = 0$ and $t = 1$.

In total, there are 108,600 information cascades are generated and the cascades with less than 3 active users are filtered out. Finally, 27,218 information cascades are taken as our data. We randomly sample 80% of the data as our train set, 10% as the validation set and 10% as the test set.

5.1.2 Sina Weibo Data Set. For real-world data, let's turn our attention to the information cascades on Sina Weibo, one of the most popular social platform in China. The Sina Weibo data set used in this paper is from [41, 42] and publicly available online². The network in this data set is the following network, reflecting the following relationships between users. Note that, such following network is quite related to the retweet information cascades since the posted messages by user B will appear in user A's feed when user A follows user B in Sina Weibo. The following network contains 1.78 million users and 308 million following relationships in total. 300 thousand popular microblog retweet information cascades of these users are included. To analysis the retweets behavior of a specific group of users with corresponding the information cascades, we construct a subset of users and messages on the user-microblog bipartite graph. Specifically, we start with a randomly chosen user and then obtain all the messages with coverage $\geq \eta$. The coverage is defined as the number of users in the chosen set normalized by the number of total users in the message. The users appeared in the obtained messages are then added into the chosen

²<https://www.aminer.cn/Influencelocality>.

user set. We repeated the above steps and obtain 23,732 users with corresponding 149,53 information cascades. The largest component of the following network between these users is regarded as the final network, containing 23,681 users and 1,802,146 edges. Information cascades with less than 5 active users are filtered out and the remaining 3,228 pieces of information are taken as our data. As for the observation time window, we set the observation time window $T = 1$ hour, 2 hours and 3 hours respectively.

5.2 Baselines

Since this paper focuses on the network-aware popularity prediction without temporal information, we mainly consider methods that utilize early adopters and network structure as our baselines. Existing methods for this problem are mainly classified into two categories: feature-based methods and deep representation learning methods. We choose the state-of-the-art method in each category as our baselines. Besides, we also include the representative attempt of capturing the cascading effect in Popularity prediction.

5.2.1 Feature-based. We extract all the effective hand-crafted features that can be easily generalized across data sets [3, 6, 16, 28]. The extracted features are conducted on three types of graphs: the global graph \mathcal{G} , the cascade graph g_c , and the frontier graph g_f . Specifically, the cascade graph g_c contains all early adopters and the corresponding edges between these users. The frontier graph g_f contains all users in the one-hop neighborhood of early adopters and the edges between these neighboring users. As for features, we extract the mean and 90th percentile of the degrees of users [14], the number of leaf nodes, edge density [10, 45] in g_c ; the number of substructures [28, 30], including nodes, edges and triangles, and the number of communities and the corresponding coverage of the partition of these communities [4, 33] in both g_c and g_f . In addition, since the node identity is quite important for popularity prediction [16], here we also include the global node ids in \mathcal{G} as the structure feature. Once the cascade is represented as a bag of features, we feed them into a linear regression model with L2 regularization.

5.2.2 DeepCas [16]. DeepCas is the state-of-the-art deep representation learning method for network-aware popularity prediction, which learns the representation of cascade graphs in an end-to-end manner. Specifically, it represents the cascade graph as a collection of sequences by random walks, and then utilizes the embeddings of nodes and recurrent neural networks to obtain the representation of each sequence. Attention mechanisms are further applied to assemble the representation of the cascade graph from sequences.

5.2.3 SEISMIC [46]. SEISMIC is a representative method for attempts of capturing the cascading effect. It is an implementation of Hawkes self-exciting point process and estimates or approximates the impact of cascading effect in each generation by the average number of fans of users.

5.3 Implementation Details.

For all baselines and our CoupledGNN model, the hyper-parameters are tuned to obtain the best results on the validation set. The L2-coefficient is chosen from $\{10^{-8}, 10^{-7}, \dots, 0.01, 0.1\}$. For feature-based method, since the features of node ids are high-dimensional

and sparse, we set a learning rate alone for the parameters of these features and choose from $\Phi_1 = \{10^{-5}, 5 \times 10^{-5}, 10^{-4}, \dots, 0.01\}$. For the parameters of other features, we choose the learning rate from $\Phi_2 = \{0.0005, 0.001, 0.005, 0.01\}$. Similarly, for DeepCas, the learning rate for user embeddings are chosen from Φ_1 , while the learning rate for other parameters are chosen from Φ_2 . The user embeddings for DeepCas is initialized by DeepWalk [20] which will be further optimized during the training process, while the user embeddings for our CoupledGNN are also obtained by DeepWalk but without further fine-tuning. The dimension of the embeddings is all set to be 32. The hidden units of RNN in DeepCas is set to be 32, and the units of the first dense layer and the second dense layer in the output part are 32 and 16 respectively. As for SEISMIC, we adopt the setting of parameters used in [46], i.e., setting the constant period s_0 to 5 minutes and power-law decay parameters $\theta = 0.242$. Besides, we choose mean degree n^* from $\{1, 3, 5, 10, 20, 50, 100\}$ to minimize the mRSE of validation set. For our CoupledGNN model, similar to baselines, the learning rate for self activation parameters of all users are chosen from Φ_1 , and the learning rate for other parameters are chosen from Φ_2 . The coefficient λ in the loss function, which balances the weight of the regularization of user-level cross entropy, is set to be 0.5 in our experiments. The number of GNN layers K is chosen from $\{2, 3, 4\}$, and each layer contains the same number of hidden units as input. Following [22], the vertex features for our CoupledGNN contains coreness, pagerank, hub score, authority score, eigenvector centrality, and clustering coefficient.

5.4 Evaluation Metrics

We adopt several different evaluation metrics to comprehensively demonstrate the performance of each method.

5.4.1 Mean Relative Square Error (MRSE) [2, 29]. We take the mean relative square error loss also as our evaluation metric for popularity prediction.

5.4.2 Median Relative Square Error (mRSE). Since SEISMIC is sensitive to outlier error, we also use median RSE as an evaluation metric, which is defined as the 50th percentile of the distribution of RSE over test data.

5.4.3 Mean Absolute Percentage Error (MAPE) [27, 34]. This metric measures the average deviation between the predicted and true popularity. The formulation is

$$MAPE = \frac{1}{M} \sum_{m=1}^M \frac{|\hat{n}_{\infty}^m - n_{\infty}^m|}{n_{\infty}^m}. \quad (15)$$

5.4.4 Wrong Percentage Error (WroPerc). The wrong prediction error is defined as the percentage of online contents that are incorrectly predicted for a given error tolerance ϵ :

$$WroPerc = \frac{1}{M} \sum_{m=1}^M \mathbb{I} \left[\frac{|\hat{n}_{\infty}^m - n_{\infty}^m|}{n_{\infty}^m} \geq \epsilon \right]. \quad (16)$$

We set the threshold $\epsilon = 0.5$ in this paper.

Note that, among all these three evaluation metrics, the smaller the value is, indicating the better the performance of the corresponding method.

Table 1: Popularity prediction in Sina Weibo

Observation Time	1 hour				2 hours				3 hours			
Evaluation Metric	MRSE	mRSE	MAPE	WroPerc	MRSE	mRSE	MAPE	WroPerc	MRSE	mRSE	MAPE	WroPerc
SEISMIC	-	0.2112	-	48.63%	-	0.1347	-	34.59%	-	0.0823	-	27.15%
Feature-based	0.2106	0.1254	0.3749	35.17%	0.1796	0.1041	0.3557	28.86%	0.1581	0.0804	0.3147	18.97%
DeepCas	0.2077	0.0930	0.3633	30.00%	0.1650	0.0670	0.3134	20.55%	0.1365	0.0361	0.2813	17.24%
CoupledGNN	0.1816	0.0946	0.3515	25.68%	0.1397	0.0519	0.2989	17.81%	0.1120	0.0333	0.2611	13.01%

Table 2: Popularity prediction in synthetic dataSet

Evaluation Metric	MRSE	mRSE	MAPE	WroPerc
SEISMIC	-	0.2025	-	47.92%
Feature-based	0.1225	0.0452	0.2718	16.90%
DeepCas	0.1199	0.0361	0.2657	16.82%
CoupledGNN	0.1101	0.0339	0.2517	14.71%

6 EXPERIMENTAL RESULTS

In this section, we first compare our CoupledGNN with baselines on the target task: popularity prediction. Besides, the superior of the coupled structure in CoupledGNN over single-GNN is also demonstrated. Finally, the effect of hyper-parameters or experimental settings is analyzed comprehensively.

6.1 Overall Performance

The experimental results for popularity prediction on both the synthetic data and real-world data in Sina Weibo are shown in Table 2 and Table 1 respectively.

For SEISMIC, due to it predicts infinite popularity for some pieces of information, we only use mRSE and WroPerc as the evaluation metrics for a fair comparison. From the experimental results, we can see that SEISMIC performs not well on both data sets. Since it only estimates the impact of the cascading effect in each generation by the average number of fans, it's easy to deviate from complex and real situations, thus having limited predictive power. As for DeepCas, the deep representation learning methods, it does perform better than the feature-based methods. This result indicates that it's effective to automatically learning the representation of the cascade graph through an end-to-end manner rather than heuristically design hand-crafted features with prior knowledge.

As for our CoupledGNN model, it outperforms all the baselines on both synthetic and real-world datasets, achieving more than 10% improvement over DeepCas in Sina Weibo under the MRSE. These results demonstrate that it's effective to utilize graph neural networks to capture the cascading effect along network structure and to predict the popularity of online content on social platforms. In other words, considering the interactions between early adopters and the potential active users, as well as the interactions among potential active users over network structure is useful to further improve the prediction performance for future popularity.

As for the observation time in Sina Weibo (Figure 1), we can see that the longer the observation time is, the smaller the errors are (MRSE, MAPE, WrongPerc). This is applicable to all the methods.

Table 3: Compare CoupledGNN with Single-GNN

Observation Time	1 hour		
Evaluation Metric	MRSE	MAPE	WroPerc
Single-GCN	0.1964	0.3707	29.11%
Single-GAT	0.1999	0.3754	30.82%
CoupledGNN	0.1816	0.3515	25.68%
Observation Time	2 hours		
Evaluation Metric	MRSE	MAPE	WroPerc
Single-GCN	0.1595	0.3201	22.26%
Single-GAT	0.1569	0.3199	20.55%
CoupledGNN	0.1397	0.2989	17.81%
Observation Time	3 hours		
Evaluation Metric	MRSE	MAPE	WroPerc
Single-GCN	0.1230	0.2653	16.10%
Single-GAT	0.1222	0.2655	16.10%
CoupledGNN	0.1120	0.2611	13.01%

The reason is that the longer the observation time is, the more information is available, making the prediction easier.

6.2 Compare CoupledGNN with Single-GNN

To further demonstrate the advantages of our CoupledGNN structure, we simplify our method with two versions: Single-GCN and Single-GAT. In both these two simplified versions, we concatenate the activation state and influence representation as one vector associated with each user. Then a single graph neural network, i.e., the commonly used graph convolution network (GCN) [13] or graph attention neural networks (GAT) [31] are applied to iteratively update the vector associated with each user. The final popularity is obtained similarly as CoupledGNN, i.e., applying a sum pooling mechanism over all users after transforming the vector of each user at the last layer into one-dimensional value. Other hyper-parameters and the implementation details are the same as CoupledGNN.

The experimental results are shown in Table 3. Single-GCN and Single-GAT perform almost similarly, indicating that when modeling the future popularity, the normalized Laplacian matrix used in GCN is already a good reflection of the correlation between pair of users with a linking edge. The attention mechanism adopted in GAT won't significantly improve the performance further. As for our CoupledGNN, it significantly improves the prediction performance

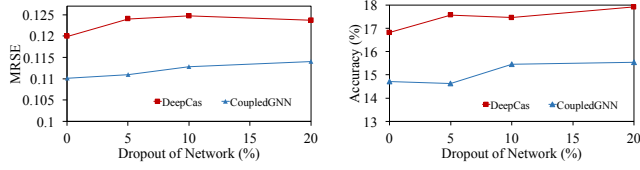


Figure 5: The influence of a partial lack of network.

under all the evaluation metrics compared with the Single-GNN methods. These results demonstrate that the activation state and influence representation play different roles in the modeling of future popularity. Instead of mixing them up together, it's effective to model the activation state and influence representation by two graph neural networks respectively and then couple them by gating mechanisms.

6.3 Parameter Analysis

We further analyze the effect of the coefficient λ , the number of layers in CoupledGNN in this subsection. The influence of a partial lack of network is also analyzed.

6.3.1 The coefficient λ in loss function. We vary the coefficient λ from 0.0, 0.5, 1.0, 10.0 to 20.0, and the corresponding mean relative square loss is 0.1109, 0.1101, 0.1111, 0.1111, 0.1141. In other words, the MRSE loss is first reduced with the increasing of λ , indicating that adding the user-level cross entropy loss is beneficial for macro popularity prediction. However, with the continuous increase of λ , the model pays too much attention to the user-level prediction while ignoring the macro popularity prediction task, thus resulting in less effective prediction performance.

6.3.2 The influence of a partial lack of network. Considering that our method is based on the given underlying network, we further construct an experiment with the partial lack of network to better demonstrate the applicability and generality of our methods. Specifically, we randomly dropout a certain percentage of edges in the social network under the premise of network connectivity. Then we train and test the prediction model based on such an incomplete network. Note that, not only our CoupledGNN are influenced by such dropout of network edges, but also the baseline methods. Here, we compare our methods with the strong baseline, i.e., DeepCas, while varying the dropout of the network from 0%, 5%, 10% to 20%. Figure 5 shows that both the performance of our CoupledGNN and DeepCas will be slightly degraded by the dropout of the network. But by comparison, our methods always significantly perform better than DeepCas. In conclusion, our CoupledGNN model is still suitable for predicting the future popularity of online content even when a small part of the network structure is lacking.

6.3.3 The number of layers in CoupledGNN. While we apply our CoupledGNN to capture the cascading effect along the network, it's interesting to further find out whether the number of layers K in our CoupledGNN can correspond to the scope of the cascading effect. We take the Sina Weibo as a showing case. On the one hand, as mentioned in Section 5.2, the number of layers K is chosen from 2, 3, 4. When the observation time window is 1 hour, we obtain the best performance on the validation set with $K = 3$. On the other

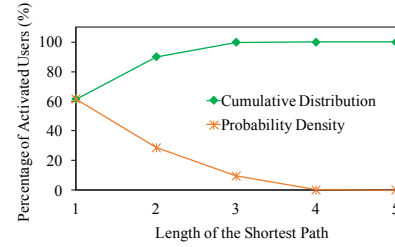


Figure 6: The distribution of the shortest path length between activated users after the observation time and the set of early adopters within observation time.

hand, Figure 6 shows the distribution of the length of the shortest path from the set of **early adopters** within observation time to the **activated users** after the observation time. Such distribution reflects the range of cascading effect caused by the early adopters. We can see that almost all the activated users, i.e., 99.76%, can be covered within three-hops in the neighborhood of early adopters. This means that the optimal number of layers obtained by our methods can exactly match the scope of the cascading effect in this case, which provides guidance for setting the hyper-parameter K in other situations, as well as further supports the effectiveness of our proposed method.

7 CONCLUSION

In this paper, we focus on the problem of network-aware popularity prediction of online content on social platforms. How to capture the cascading effect is one of the keys to accurately predict future popularity and tackle this problem. Inspired by the success of graph neural networks on various non-Euclidean domains, we propose CoupledGNN to characterize the critical cascading effect along the network structure. We devote to modeling the two crucial components in the cascading effect, i.e., the iterative interplay between node activation states and the spread of influence, by two coupled graph neural networks respectively. Specifically, one graph neural network models the interpersonal influence, gated by the activation state of users. The other graph neural network models the activation state of users via interpersonal influence from their neighbors. The iterative update mechanism of neighborhood aggregation in GNNs effectively captures such a cascading effect in popularity prediction along the underlying network. The experiments conducted on both synthetic and real-world data validate the effectiveness of our proposed method for popularity prediction. As for future work, we will devote to modeling the cascading effect along the network when further given the specific adoption time of early adopters.

ACKNOWLEDGMENTS

This work is funded by the National Natural Science Foundation of China under grant numbers 61425016, 61433014, 91746301, and 61472400. This work is supported by Beijing Academy of Artificial Intelligence (BAAI). Huawei Shen is also funded by K.C. Wong Education Foundation and the Youth Innovation Promotion Association of the Chinese Academy of Sciences.

REFERENCES

- [1] Qi Cao, Huawei Shen, Keting Cen, Wentao Ouyang, and Xueqi Cheng. 2017. DeepHawkes: Bridging the Gap Between Prediction and Understanding of Information Cascades. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. 1149–1158.
- [2] Qi Cao, Huawei Shen, Hao Gao, Jinhua Gao, and Xueqi Cheng. 2017. Predicting the Popularity of Online Content with Group-specific Models. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17)*. 765–766.
- [3] Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can Cascades Be Predicted?. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. 925–936.
- [4] Aaron Clauset, M. E. J. Newman, and Christopher Moore. 2004. Finding community structure in very large networks. *Phys. Rev. E* 70 (Dec 2004), 066111. Issue 6.
- [5] Nan Du, Yingyu Liang, Maria-Florina Balcan, and Le Song. 2014. Influence Function Learning in Information Diffusion Networks. In *Proceedings of the 31th International Conference on Machine Learning (ICML '14)*. II–2016–II–2024.
- [6] Xiaofeng Gao, Zhenhao Cao, Sha Li, Bin Yao, Guihai Chen, and Shaojie Tang. 2019. Taxonomy and Evaluation for Microblog Popularity Prediction. *ACM Trans. Knowl. Discov. Data* 13, 2, Article 15 (March 2019), 40 pages.
- [7] Jacob Goldenberg, Barak Libai, and Eitan Muller. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters* 12, 3 (2001), 211–223.
- [8] Manuel Gomez-Rodriguez, David Baldazzi, and Bernhard Schölkopf. 2011. Uncovering the Temporal Dynamics of Diffusion Networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML '11)*. 561–568.
- [9] Mark Granovetter. 1978. Threshold Models of Collective Behavior. *Amer. J. Sociology* 83, 6 (1978), 1420–1443.
- [10] Adrien Guille and Hakim Hacid. 2012. A predictive model for the temporal dynamics of information diffusion in online social networks. In *Proceedings of the 21st international conference on World Wide Web (WWW '12)*. 1145–1152.
- [11] Ruocheng Guo and Paulo Shakarian. 2016. A Comparison of Methods for Cascade Prediction. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '16)*. 591–598.
- [12] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*. 1025–1035.
- [13] Thomas N Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 6th International Conference on Learning Representations*.
- [14] Kristina Lerman and Aram Galstyan. 2008. Analysis of Social Voting Patterns on Digg. In *Proceedings of the First Workshop on Online Social Networks (WOSN '08)*. 7–12.
- [15] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. 2010. Kronecker Graphs: An Approach to Modeling Networks. *J. Mach. Learn. Res.* 11 (March 2010), 985–1042.
- [16] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. 2017. DeepCas: An End-to-end Predictor of Information Cascades. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. 577–586.
- [17] Dongliang Liao, Jin Xu, Gongfu Li, Weijie Huang, Weiqing Liu, and Jing Li. 2019. Popularity Prediction on Online Articles with Deep Fusion of Temporal Process and Content Features. In *Thirty-Three AAAI Conference on Artificial Intelligence (AAAI '19)*. 200–207.
- [18] Travis Martin, Jake M. Hofman, Amit Sharma, Ashton Anderson, and Duncan J. Watts. 2016. Exploring Limits to Prediction in Complex Social Systems. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. 683–694.
- [19] Swapnil Mishra, Marian-Andrei Rizoio, and Lexing Xie. 2016. Feature Driven and Point Process Approaches for Popularity Prediction. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16)*. 1069–1078.
- [20] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. 701–710.
- [21] Henrique Pinto, Jussara M. Almeida, and Marcos A. Gonçalves. 2013. Using Early View Patterns to Predict the Popularity of Youtube Videos. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM '13)*. 365–374.
- [22] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. 2018. DeepInf: Social Influence Prediction with Deep Learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '18)*. 2110–2119.
- [23] Marian-Andrei Rizoio, Lexing Xie, Scott Sanner, Manuel Cebrian, Honglin Yu, and Pascal Van Hentenryck. 2017. Expecting to Be HIP: Hawkes Intensity Processes for Social Media Popularity. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. 735–744.
- [24] Satoshi Sanjo and Marie Katsurai. 2017. Recipe Popularity Prediction with Deep Visual-Semantic Fusion. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. 2279–2282.
- [25] Jin Shang and Mingxuan Sun. 2019. Geometric Hawkes Processes with Graph Convolutional Recurrent Neural Networks. In *Thirty-Three AAAI Conference on Artificial Intelligence (AAAI '19)*.
- [26] Jiangli Shao, Huawei Shen, Qi Cao, and Xueqi Cheng. 2019. Temporal Convolutional Networks for Popularity Prediction of Messages on Social Medias. In *China Conference on Information Retrieval*. Springer, 135–147.
- [27] Huawei Shen, Dashun Wang, Chaoming Song, and Albert-László Barabási. 2014. Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI '14)*. 291–297.
- [28] Benjamin Shulman, Amit Sharma, and Dan Cosley. 2016. Predictability of popularity: Gaps between prediction and understanding. In *Tenth International AAAI Conference on Web and Social Media (ICWSM '16)*. 348–357.
- [29] Alexandru Tatar, Marcelo Dias de Amorim, Serge Fdida, and Panayotis Antoniadis. 2014. A survey on predicting the popularity of web content. *Journal of Internet Services and Applications* 5, 1 (13 Aug 2014), 8.
- [30] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109, 16 (2012), 5962–5966.
- [31] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph attention networks. In *Proceedings of the 7th International Conference on Learning Representations*.
- [32] Jia Wang, Vincent W Zheng, Zemin Liu, and Kevin Chen-Chuan Chang. 2017. Topological recurrent neural network for diffusion prediction. In *2017 IEEE International Conference on Data Mining (ICDM)*. 475–484.
- [33] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. 2014. Predicting successful memes using network and community structure. In *Eighth international AAAI conference on weblogs and social media (ICWSM '14)*. 535–544.
- [34] Qitian Wu, Chaoqi Yang, Hengrui Zhang, Xiaofeng Gao, Paul Weng, and Guihai Chen. 2018. Adversarial Training Model Unifying Feature Driven and Point Process Perspectives for Event Popularity Prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. 517–526.
- [35] Bingbing Xu, Huawei Shen, Qi Cao, Keting Cen, and Xueqi Cheng. 2019. Graph Convolutional Networks Using Heat Kernel for Semi-supervised Learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI '19)*. 1928–1934.
- [36] Bingbing Xu, Huawei Shen, Qi Cao, Yunqi Qiu, and Xueqi Cheng. 2019. Graph Wavelet Neural Network. In *Proceedings of the 8th International Conference on Learning Representations*.
- [37] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *Proceedings of the 8th International Conference on Learning Representations*.
- [38] Junchi Yan, Xin Liu, Liangliang Shi, Changsheng Li, and Hongyuan Zha. 2018. Improving Maximum Likelihood Estimation of Temporal Point Process via Discriminative and Adversarial Learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI '18)*. 2948–2954.
- [39] Rex Ying, Jiaxuan You, Christopher Morris, Xiang Ren, William L. Hamilton, and Jure Leskovec. 2018. Hierarchical Graph Representation Learning with Differentiable Pooling. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS '18)*. 4805–4815.
- [40] Linyun Yu, Peng Cui, Fei Wang, Chaoming Song, and Shiqiang Yang. 2015. From micro to macro: Uncovering and predicting information cascading process with behavioral dynamics. In *2015 IEEE International Conference on Data Mining (ICDM '15)*. 559–568.
- [41] Jing Zhang, Biao Liu, Jie Tang, Ting Chen, and Juanzi Li. 2013. Social Influence Locality for Modeling Retweeting Behaviors. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI '13)*. 2761–2767.
- [42] Jing Zhang, Jie Tang, Juanzi Li, Yang Liu, and Chunxiao Xing. 2015. Who Influenced You? Predicting Retweet via Social Influence Locality. *ACM Trans. Knowl. Discov. Data* 9, 3, Article 25 (April 2015), 26 pages.
- [43] Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. 2018. An end-to-end deep learning architecture for graph classification. In *Thirty-Second AAAI Conference on Artificial Intelligence (AAAI '18)*.
- [44] Wei Zhang, Wen Wang, Jun Wang, and Hongyuan Zha. 2018. User-guided Hierarchical Attention Network for Multi-modal Social Image Popularity Prediction. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. 1277–1286.
- [45] Xiaoming Zhang, Zhoujun Li, Wenhan Chao, and Jiali Xia. 2014. Popularity Prediction of Burst Event in Microblogging. In *Web-Age Information Management*. Springer International Publishing, 484–487.
- [46] Qingyuan Zhao, Murat A. Erdogdu, Hera Y. He, Anand Rajaraman, and Jure Leskovec. 2015. SEISMIC: A Self-Exciting Point Process Model for Predicting Tweet Popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*. 1513–1522.