

# User Influence Analysis by Ripple Network on micro-blog Social Networks

Shengyuan Huang and Buqing Nie  
Shanghai Jiao Tong University

## 1. INTRODUCTION

### 1.1 Background

Recently, the micro-blog plays more and more important role in our daily life, which has become a main approach to receiving information from the network. However, the number of users on social network like twitter, micro-blog has risen sharply. Meanwhile, the amount of information on social media is exploding everyday. There are more and more information of low quality, which are difficult for us to distinguish.

Moreover, The experiment tell us in each social media, there are some influential users who can predetermine the topics for discussion and the attitudes to certain events or persons to a large extent. And we can seen these influential users as good information sources.

Besides, many commercial agents such as many advertising company need to identify opinion leaders and maximize the dissemination of the advertise message by analyzing existing micro-blogs. Thus, they need to compute the user influence value through analyzing the micro-blog logs.

In addition, the user influence estimation could be applied for friends recommendation, information dissemination prediction, user behavior prediction, etc. Above all, the user influence estimation has become an important aspect of informational and analytic work with social networks, which is the main problem we want to solve in this project.

### 1.2 Motivation

In this project, we want to transfer Ripple Network[Wang et al. 2018] model from Recommend systems to the user influence estimation in the social network, expecting a good performance compared to the existing models.

For any user  $u$ , we can estimate the influence value of  $u$  in the given social network, then we can believe that user  $u$  is a high quality information source, and the blogs he/she posted are information of high quality. The probability that the other users retweet his/her blogs is higher than other users.

### 1.3 contributions

The main contributions of this project can be summarized in these points:

- (1) We transfer Ripple Network model from Recommend systems to the user influence estimation in the social network to estimate the user influence value.
- (2) We start an experiment to show the performance of the Ripple Network in the user influence estimation region.

## 2. RELATED WORK

The user influence is one of the hot topics in the social network because of its wide application. The followings are the state-of-the-art related works in the recent years.

### 2.1 Ripple Network

The Ripple Network is an end-to end framework that naturally incorporates the knowledge graph into recommender systems for knowledge-graph-aware recommendation. The key idea behind Ripple Network is preference propagation: For each user, Ripple Network treats his historical interests as a seed set in the KG, then extends the users interests iteratively along KG links to discover his hierarchical potential interests with respect to a candidate item.[Wang et al. 2018]

### 2.2 User influence estimation

**2.2.1 Rank-based model.** Rank-based model usually measures users influence by exploiting the link relationship between the users with an iterative algorithm. They are mainly based on the Page Rank algorithm, like PageRank based model TwitterRank[Weng et al. 2010] which measures posters influence by taking into account the link structure of followers/following of individual users and the topical similarity between these users. Li et al. [Li et al. 2011] proposed an iterative algorithm to calculate the influence of the individual in the social network by conformity. The conformity in his model is defined as the probability that an individual is willing to accept others opinions.

**2.2.2 Feature-based model.** Feature-based model measures users influence by exploiting the social features in specific media like the number of followers and posts in the context of twitter. In the context of twitter, Pal and Counts[Ou et al. 2016] proposed a model Twitter Authority that uses the count of original tweets, conversational tweets, and retweets of a tweeter as features to rank the authority of each tweeter in different topics. In the context of blog, Agarwal et al. [Agarwal et al. 2008] proposed a model to measure the influence of bloggers. They think of bloggers posted enough influential blogs as influential bloggers.

### 2.3 Network Embedding

Network embedding methods aim at learning low-dimensional latent representation of nodes in a network, which has become a popular method to manipulate the network data. The network embedding can be categorized into Random Walk based methods, Edge modeling based methods, and Matrix factorization based methods.

The random walk based methods implement the function by performing truncated random walks. The information network is transformed into a collection of vertex sequences, in which, the occurrence frequency of a vertex-context pair measures the structural distance between them. The most popular one is DEEPWALK[Perozzi et al. 2014].

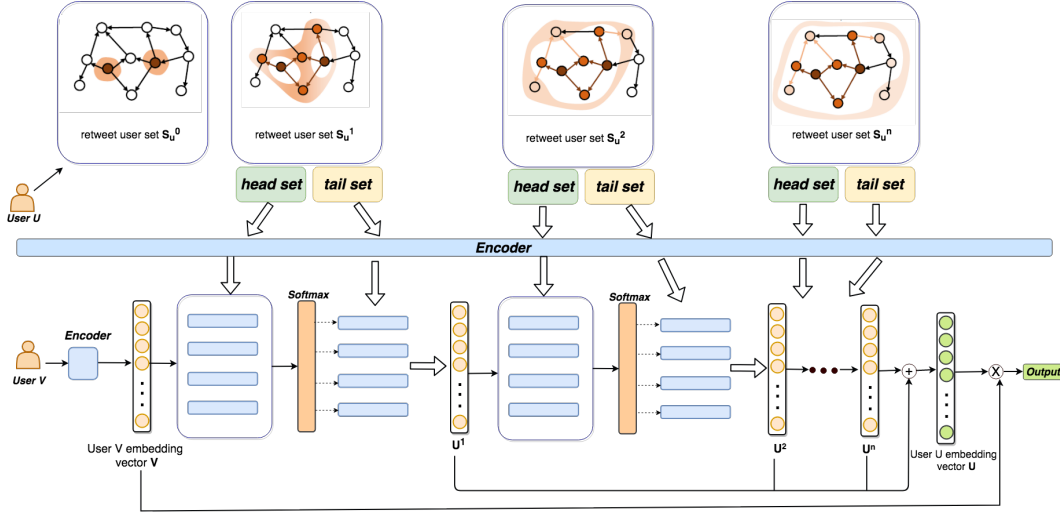


Fig. 1. The overall framework of our model. It takes a user  $u$  and a user  $v$  as input, and outputs the predicted user  $v$ 's influence to user  $u$ .

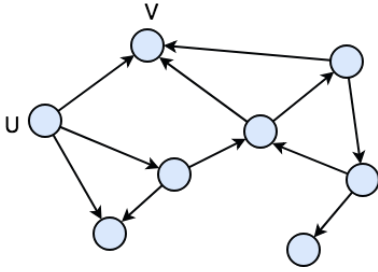


Fig. 2. One example of the social network graph

Edge modeling based methods directly learn vertex representations from vertex-vertex connections. One popular method is LINE[Tang et al. 2015].

Matrix factorization based methods represent the connections between network entities in the form of a matrix and use matrix factorization to obtain the embeddings. One popular method is HOPE[Ou et al. 2016].

### 3. PROPOSED METHODS

#### 3.1 Problem Formulation

- (1) *user-user influence matrix*: In our project, let  $U = \{u_1, u_2, \dots\}$  denotes the sets of users. The user-user influence matrix  $Y = \{y_{uv} | u, v \in U\}$  is defined according to users' action, where  $y_{uv} = 1$  if user  $u$  read and retweet user  $v$ 's weibo.
- (2) *Social network graph*:  $G = (V, E)$  notes a social network.  $V$  denotes users,  $E$  is the set of relationships between users, which consists of the information of following. That is, if user  $u$  follows user  $v$ , there is a directed edge from  $u$  to  $v$  in  $G$ . In other words, there are massive directed edge  $(u', v')$  in  $G$ , where  $u' \in U, v' \in U$  denote user  $u$  and the user  $v$ . The Fig2 is an example social network graph.

Given an influence matrix  $Y$  as well as a graph  $G$ , we aim to predict user  $v$ 's influence to user  $u$ . Our goal is to learn a prediction

function  $\hat{y}_{uv} = F(u, v; \theta)$ , where  $\hat{y}_{uv}$  denotes the user  $v$ 's influence to user  $u$ , and  $\theta$  denotes the model parameters of function  $F$ .

#### 3.2 the model of the problem

The framework of our model is illustrated in the Figure. It takes a user  $u$  and a user  $v$  as input, and outputs the predicted user  $v$ 's influence to user  $u$ . We define

$$S_u^k = \{v' | (u', v') \in G \text{ and } u' \in S_u^{k-1}\}, k = 1, 2, \dots, n.$$

$$S_u^0 = V_u = \{v | y_{uv} = 1\}$$

For the input user  $u$ , his following set of  $V_u$  is treated as seeds in  $G$ , then extended to form multiple sets  $S_u^k (k = 1, 2, \dots, n)$ , where

$$E_u^k = \{(u', v') | (u', v') \in G \text{ and } u' \in S_u^{k-1}\}, k = 1, 2, \dots, n.$$

These sets are used to interact with the user  $v$ 's embedding for obtaining the responses of user  $u$  with respect to user  $v$ , which are then combined to form the final user  $u$ 's embedding. Lastly, we use the embeddings of user  $u$  and user  $v$  together to compute the predicted user  $v$ 's influence to user  $u$   $\hat{y}_{uv}$ .

Also, as is shown in the Figure, the user  $v$  is associated with an embedding  $v \in R^d$  through a certain encoder (such as the one-hot encoder), where  $d$  is the dimension of embeddings. Given the embedding  $v$  and the 1-hop set  $E_u^1$  of user  $u$ , each tuple  $(u'_i, v'_i)$  in  $E_u^1$  is assigned a relevance probability by comparing  $v$  to the  $u'_i$  in this tuple:

$$p_i = \text{softmax}(v^T u'_i) = \frac{\exp(v^T u'_i)}{\sum_{(u', v') \in E_u^1} \exp(v^T u')}$$

, where  $u'_i \in R^d$  are the embeddings of  $u'_i$ . After obtaining the relevance probabilities, we take the sum of  $v'$  in  $E_u^1$  weighted by the corresponding relevance probabilities, and the vector  $U^1$  is returned:

$$U^1 = \sum_{(u'_i, v'_i) \in E_u^1} p_i v'_i$$

, where  $v'_i \in R^d$  is the embedding of  $v'_i$ . Repeating the procedure, we can get  $U^1, U^2, \dots, U^n$ . Then we can get the embedding of user  $u$

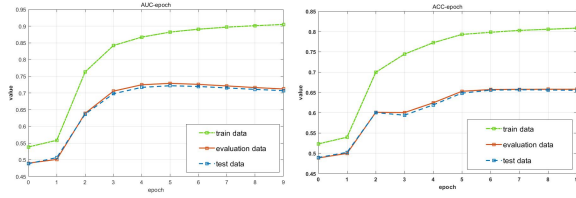


Fig. 3. The AUC and accuracy - epoch result

with respect to user  $v$ :

$$u = \alpha_1 H^1 + \alpha_2 H^2 + \dots + \alpha_n H^n$$

, where  $\alpha_i$ s are positive trainable parameters.

Finally, the embedding of  $u$  and the embedding of  $v$  are combined to output the predicted user  $v$ 's influence to user  $u$ :

$$\hat{y}_{uv} = \sigma(u^T v)$$

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

## 4. EXPERIMENT

To verify the performance of the Ripple Network in the social influence estimation, we conduct an user retweet behavior prediction experiment on the Sina micro-blog social network.

In the experiment, if the influence value  $v_{uv}$  of user  $u$  to user  $v$  less than 0.5, we assume that user  $v$  won't retweet the micro-blog of user  $u$ . Otherwise, he will retweet the micro-blog.

### 4.1 Dataset

The dataset used in our experiment is the Sina micro-blog public data on the *aminer.cn*, which is a quite big dataset, the following are the introduction of the dataset:

Dataset Source	Sina micro-blog
#Users	1,776,950
#Follow relationships	308,489,739
#Original micro-blogs	300,000
#Retweets	23,755,810

Because the dataset is too big for us to conduct the experiment, so the dataset used in our experiment is simplified.

### 4.2 Results

The Fig3 shows that the convergence is very quickly, however, the result of the test dataset is not satisfactory. The Fig4 is the influence value distribution figure on our dataset, the distribution is quite smooth, which is expected.

The Fig7 is the comparison between the Ripple Network and other existing models.

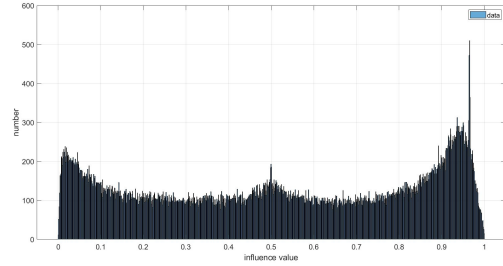


Fig. 4. The influence value distribution in our dataset

Fig. 5. Caption

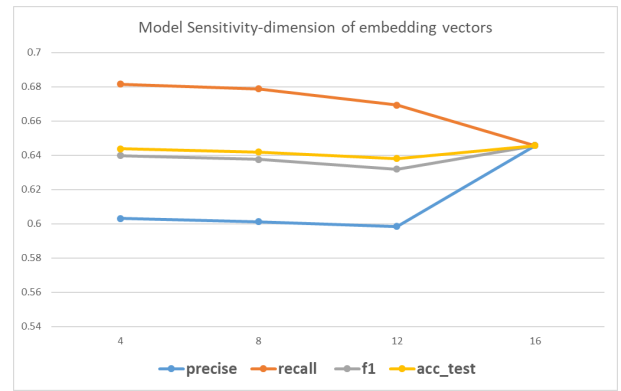


Fig. 6. Model sensitivity analysis

## 5. DISCUSSION

The Fig57 is the precision- $d$  figure,  $d$  is the dimension of embedding vectors, which is an important parameter in our model. The experiment shows that the change of  $d$  has just a little influence on the result. The Ripple Network is not sensitive to the parameter  $d$ .

Furthermore, the value of  $d$  has no influence to recall and f1 value fundamentally, however, the value of  $d$  has big influence of recall and precise. To get a higher precise, bid  $d$  is needed. Because the running time is too long to wait, so we only analyze the  $d = 4, 8, 12, 16$ , more experiments are needed.

## 6. CONCLUSION

In this project, we transfer the ripple network model from the recommender system to the social influence estimation region. The original knowledge graph is substitute to the social network, and the item vector is substitute to the user vector. The output of Ripple network is the influence value of user  $u$  to user  $v$ .

The experiment result is not satisfactory, which is expected because the improvement is not sufficient. The retweet behavior prediction experiment result show the promising potential of our model.

## 7. FUTURE WORK

Because of lack of time, some improvements of the ripple networks haven't be implemented. The encoder we used now is one-hot en-

Method	Precision	Recall	F1
PMF(2007)	0.484	0.434	0.458
LRC-BQ(2015)	0.518	0.677	0.587
MNMF(RP(2015)	0.674	0.715	0.694
HCFMF(2017)	0.787	0.805	0.796
RPSC-U	0.791	0.811	0.801
RPSC-M	0.785	0.815	0.799
<b>RIPPLE-N</b>	<b>0.630</b>	<b>0.673</b>	<b>0.650</b>

Fig. 7. Model sensitivity analysis

coder, which didn't include any user profile or user interest information. Here are the list of the not implemented improvements:

- (1) The user profile in the dataset isn't used, which is a big regret in this project. The user profile contain the location province, the number of followees, name, gender, the time when create, verified type etc. which can be used like the model of Featured based. Our plan is alternating the one-hot encoder with a more complex coder which can take advantage the user profile and output the user embedding vector containing the user profile information. Then the output value maybe better.
- (2) The user interest information can be used, too. It can be gotten from the user micro-blog content with the help of the natural language process. This information can be used by alternating the encoder, too.

#### ACKNOWLEDGMENTS

We are grateful to the following people for resources, discussions and suggestions: Prof. Jiang

We also thank the website who allowed us to collect pictures download dataset: aminer.cn

#### REFERENCES

- Nitin Agarwal, Huan Liu, Lei Tang, and Philip S Yu. 2008. Identifying the influential bloggers in a community. In *Proceedings of the 2008 international conference on web search and data mining*. ACM, 207–218.
- Hui Li, Sourav S Bhowmick, and Aixin Sun. 2011. Casino: towards conformity-aware social influence analysis in online social networks. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 1007–1012.
- Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. 2016. Asymmetric transitivity preserving graph embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1105–1114.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: On-line learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 701–710.
- Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1067–1077.
- Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. Ripple Network: Propagating User Preferences on the Knowledge Graph for Recommender Systems. *arXiv preprint arXiv:1803.03467* (2018).

Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. Twiterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 261–270.