Utilizes the Community Detection for Increase Trust using Multiplex Networks

Rahimeh Habibi[1]; Ali Haroun Abadi[2]

Abstract — **Today, e-commerce has occupied a large volume of economic exchanges. It is known as one of the most effective business practices. Predicted trust which means trusting an anonymous user is important in online communities. In this paper, the trust was predicted by combining two methods of multiplex network and community detection. In modeling the network in terms of a multiplex network, the relationships between users were different in each layer and each user had a rank in each layer. Then, the ratings of two layers including the weight of each layer were aggregated and four effective features of the Trust were achieved. Then, the network was divided into overlapping groups via community detection' algorithms, each group representative was considered as the community centers and other features were extracted through similar comments. At the end, 48J decision tree algorithm was used to advance the work. The proposed method was assessed on Epinions data set and accuracy of trust was 96%.**

Index Terms — **Trust, Community Detection, Multiplex Network, social Network.**

I. INTRODUCTION

TODAY social networks with hundreds of millions of members, are considered as an integral part of life. This growth has created substantial research opportunities on these networks. Connecting with friends and relatives, easy access to the opinions of others, entertainment and using news sites for easy understanding of the world events, are the main components of a social network. Despite the popularity and growth of the social media, the trust can be used as a tool to support decision-making [1]. Nowadays, the social commerce is moving towards e-commerce and a lot of shopping are done on the Website. In non-electronic interactions, the user and the Seller are placed face to face when buying while in the e-commerce, the Seller and the user cannot see each other and the only mean of communication is a computer network. So, the mutual trust is very important here. Among the advantages of e-commerce are removing the limitations of space and time, globalization of trade and easy knowledge of the users of others comments. Assessing anonymous users' trust before interacting directly with them and using their input are important and so, solving the problem of trust in social networks is important. Due to the large number of comments on goods and presence of users with different interests and ideas on a social network, it is necessary to predict users' trust to each other because there is no opportunity to study all comments and, even if there is the opportunity, any comment is not useful for each user. So, the goal of predicting trust can be to provide the comments of trusted users to each other and hiding the comments of users who do not trust each other. Each social

1- Student of ACRCE Khozestan (ar.habibi69@gmail.com)
2- Depatment of Computer, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

network is modelled with a G (V, E) graph where the V is the users and the E is the edges that exist between users. If edges show the trust relationships between the users, trust network can be classified in three groups: a binary trust network, a signed trust network and a value trust network [2]. In online websites, each user is known with an ID and his/her comments. In article [3], the homophily theory shows the effects of users' ideas on the goods and their trust to each other. This theory argues that the probability of trust between users who have the same rating on a common commodity is higher. Similarly, the users who trust each other, are more similar in rating the goods than those who do not trust each other. In the following paper is organized as follows: In Section 2 introduce the basic concepts used in this paper. Section 3 related work is reviewed. Section 4 described the proposed method is fully and Section 5 shows the results of experiments conducted to evaluate and finally conclude the paper in section 6.

II. PROCEDURE FOR PAPER SUBMISSION

In this section, we introduce the basic concepts used in this paper.

1. Multiplex Network

Given that most of the work done so far show all the network with one concept and the edges are all indicative of one kind of relationship, using a multiplex network allows us to provide a context for expressing multi-layered networks with different relations in each layer. Each layer determines one of the major concepts in the world of reality. Here, a two-layer multiplex network is used where one layer shows the trust relationships and the other layer shows the similarity relationships.

2. Community Detection

Another important research issue related to the social networks is the recognition of structure of societies. Community detection in a network lead to understanding the overall structure of a network, distribution on the platforms and network activities. Also, the social structure of a network is a way that affects the information transmitted in the network and also affects the behavior of individuals. In general, there are two types of research in community detection: Non-

overlapping and overlapping communities. In identifying overlapping communities, one can be a member in a number of communities, but in identifying Non-overlapping communities, all of them should be completely separate from each other [3].

III. RELATED WORK

Javari and Jalali[4], presented their work in two phases. In the first phase, they clustered the network so the clustering is all the balanced. This means that the inner edges of clusters were positive and the edges between the clusters were negative and finally in the second phase, to determine and predict trust, collaborative filtering method was executed. Mahani and Analoui [6], used reputation and optimism to determine trust and eventually used regression for learning and classification. Beigi and Jalali [5], used a semi-supervised method to solve the problem of predicting trust. This community detection based approach used the same time relations of trust and similarities between the comments of the user to solve this challenge. This method uses the values of trust of members of the Associations to predict trust between other members. In this paper, two algorithms of Game theory [11] and SLM [12] were used to identify the communities. Zheng [13], stated that social science studies have shown that people are looking for recommendations from their friends in the real world. They accept the recommendations based on their trust to their friends which is influenced by interpersonal features. Here, predicted reliability is a process that estimates the relationship between two persons in a field, when they are not directly interacting with each other [14]. In real life, trust is influenced by many features, including the tendency to trust, distribution of trust, similarity rating of trust and distribution similarity of trust [16.15].

IV. PROPOSED METHOD

In this paper, we modeled a two-layer multiplex network with different relations for each layer. A rank is defined for each node in each layer. After aggregating the rate of each node and assigning a rank to each node in the network, reputation and optimism are achieved for users according to these ranks. In the next section of this article, through the community identified algorithm, the

network was divided into overlapping groups and another significant feature in predicting trust is obtained based on the similarity of users' comments. (In general, the work stages can be seen in Fig. 1). Where U1 and U2 are user 1 and user 2, respectively.
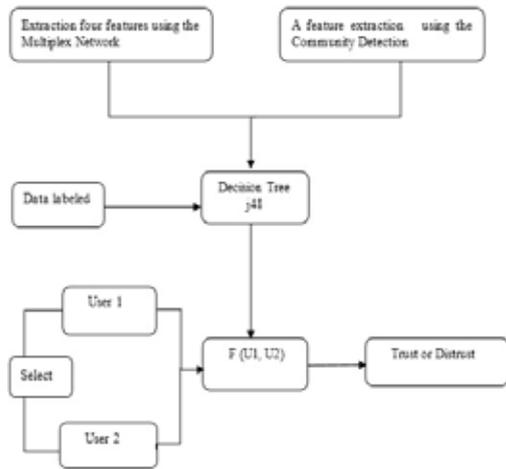

Fig. 1. The general trend of the proposed method.

1. Multiplex Networks and Ranking

Since in reality, people can have several relations with each other, network modeling in the form of a multiplex network can create such a situation. The following multiplex network is a subset of the multi-layer network in Fig. 2 so that each layer nodes are the same and edges (connections between nodes) in each layer is treated differently [17].
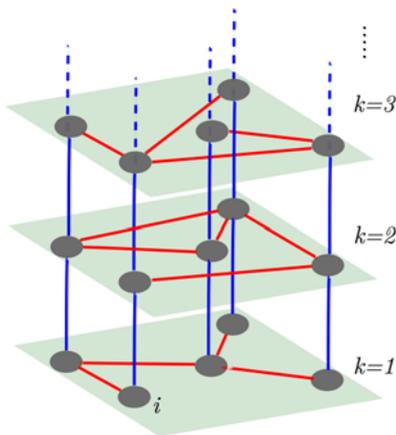

Fig. 2. Multiplex Network

A rate is defined for each node in this network

and, finally, by taking into account the weight of each layer which shows the effect of that layer in the network, a single rating is determined for each node.

$$R_i = \sum_{k=1}^{k=n} R_{ik} * \omega_i, \quad \sum_{k=1}^{k=n} \omega_i = 1 \quad (1)$$

Here, according the paper [6], a two-layer multiplex network is used. The first layer represents a layer of trust, the edges are signed in this layer and the relationships show the trust between users (+ means there is trust between users and - means distrust). The second layer is the similarity layer. Undirected edges, are asymptomatic and are weighted. Between any two vertices which jointly voted for a commodity, an edge is created and edge weight shows cosine-based similarity between users in Fig. 3.

$$\text{Sim}_{i,j} = \frac{\sum_{s \in S_{i,j}} r_{i,s} r_{j,s}}{\sqrt{\sum_{s \in S_{i,j}} r_{i,s}^2} \sqrt{\sum_{s \in S_{i,j}} r_{j,s}^2}} \quad (2)$$

By equation (2) cosine-based similarity [6] of users is obtained for similarity layer. Here, like S_i, j is a list of goods that two users of i, j voted for them and r_i, s is the i user rate to the product s and r_j, s is j user rate to the product s. Sim_i, j is the cosine similarity between two users of i, j.
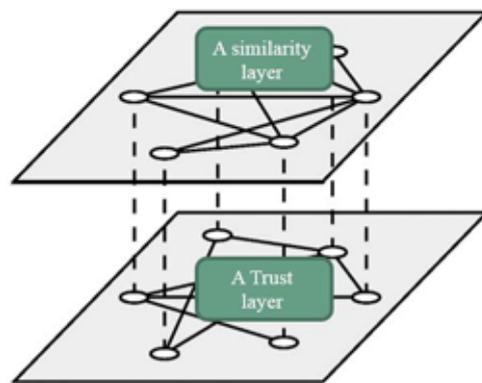

Fig. 3. A two-layer multiplex network

After modelling the network in two layers, we should obtain the users' ranks in each layer by Degree Raking algorithm. In the trust layer, each user's rate is dependent on the trusts and distrusts.

$$Pr_i = \frac{\left| IN_i^{(+)} \right| - \left| IN_i^{(-)} \right|}{\left| IN_i^{(+)} \right| + \left| IN_i^{(-)} \right|} \qquad (3)$$

IN_i^+ is the number of persons who trust user i.

IN_i^- is the number of persons who do not trust user i.

The rate of each user in the similarity layer is dependent on its resemblance to other users [7].

$$dc_i(t) = \frac{\sum_{e \in E_i} weight(e)}{|V| - 1} \qquad (4)$$

$\sum_{e \in E_i} weight(e)$ is the weight of edges that are connected to the node i.

| V | is the total number of nodes in the graph Since each node in the multiplex network must have a rating, a function aggregating the node rank in two layers is defined according to the following equation (5) [6]:

$$R_i = \omega_i r_i^1 + (1 - \omega_i) r_i^2 \qquad (5)$$

In this equation, r_i^1 is the rank of node in the trust layer and r_i^2 is the rank of layer in the similarity layer. ω_i is weight for trust layer and is obtained from the formula (6). In fact, ω_i is the cumulative distribution of edges of the input node i [6].

$$F(x_i) = \sum_{xk \leq xi} Pr(X = x_k) \qquad (6)$$

Finally, at the end of this section, two features are extracted per user. One of the characteristics is reputation and is obtained from the following equation (7). The witness of this feature is that how a user is trusted by others. The longer a user is trusted by the others and have less distrust in the user, the reputation increases.

$$Rep_i = \frac{\sum_{j \in IN_i^{(+)}} R_j - \sum_{j \in IN_i^{(-)}} R_j}{\sum_{j \in IN_i^{(+)}} R_j + \sum_{j \in IN_i^{(-)}} R_j} \qquad (7)$$

Another characteristic is optimism. The witness for this feature is that how this user trusts. As the user trusts others and do not have distrust to them, he is more optimistic.

$$Opt_i = \frac{\sum_{j \in out_i^{(+)}} R_j - \sum_{j \in out_i^{(-)}} R_j}{\sum_{j \in out_i^{(+)}} R_j + \sum_{j \in out_i^{(-)}} R_j} \qquad (8)$$

In this section, to predict trust between two users of i, j, four features of Repi, Opti, Repj, Optj were extracted.

2. Community Detection and the SLPA Algorithm

In this section, according to the paper [7], identified communities are used to split similar users into several groups. In fact, community detection means finding groups with similar interests. Various algorithms (Cfinder, Markov Cluster, Antcloney, etc.) have been used to communities detection up to now. Here, SLPA [8] which is a developed algorithm of the LPA [9] and acts based on the label is used. This algorithm consists of three steps. First, the memory of each node is labeled with the node ID. All nodes are randomly arranged in the next stage. Respectively, one node is selected from the first. It asks about the label from its neighbors. Neighbors send their labels and the node adds it to his memory based on the frequency. Finally, the memory of each node is checked and nodes with similar labels are located in a similar community. The GANXiSw is used for this and the users who voted for four common goods are grouped. After grouping of users, representative of each community is achieved through intermediate methods as the center of the community.

The intermediate approach: as the rating is greater, it shows that the node is in the direction of more communication paths. In fact, according to the shortest path between all nodes in the network, the node is selected as the center which has been mostly seen in the shortest paths [10].

$$B_C(v) = \sum_{(u,w)\epsilon V\times V, u\neq w, u\neq v, w\neq v} \frac{\sigma_{uw}(v)}{\sigma_{uw}} \quad (9)$$

Here, σ_{uw} is the number of short paths from the node u to node w and $\sigma_{uw}(v)$ is the number of short paths between u, w that passes from v. The representative selected by the above method is considered as a central member of the community and is used to calculate the probability of final trust:

$$P_{ij} = \max_{c_i \epsilon c_{si}, c_j \epsilon c_{sj}} Avg\{Sim_{ici}, Sim_{jcj}, Sim_{cicj}\} \quad (10)$$

Sim_{ici} is the similarity between i user comments and any representatives of the associations which belongs to them and Sim_{cicj} is the similarity between the comments of representatives of associations that the i, j vertices belong to them.

In this section, the fifth feature effective in the diagnosis of trust and distrust between the two users is obtained and after determining these features in the top sections, machine learning method is used to find a model based on the training data.

3. Decision Tree j48

Due to the large amounts of data, gaining knowledge and models on the properties of the data, is possible via machine learning methods. In the classification algorithm, the initial data set is divided into two parts as tutorial dataset and experimental dataset. Here, the J48 decision tree is used as a learning algorithm. This algorithm is the improved version of ID3 which is able to classify with a continuous range and considers a threshold for all kinds of selectable modes, measures the amount of applicability of information and the threshold with the most useful information is considered as the division index making of that node.

V. IMPLEMENTATION AND EVALUATION OF THE PROPOSED METHOD

The data set used for the comparison of predicted trust method was Epinions (Table1), which is related to product- review and any user can write a Opinion and rate the goods and establish a trust network with like-minded users. This set contains two parts. In the first part, the trust relationships between the users are shown by positive edges (trust relations) and negative edges (relationships of distrust) and the other part include users' ratings.

TABLE I
EPINIONS DATASET

Epinions			
Users	Relations	Trust Relations	Distrust Relations
131,828	841,372	715,167	126,205

Evaluation of a classification model is done based on training and experimental samples. In this study, Fold 10 method was used. In this way, the data are divided into 10 categories of roughly the same size where 9 categories are used to train the model, and the other is used to test the model. In this way, for 10 times, the model is tested and trained and the advantage of this method is that eventually all the existing samples participate in both training and testing processes. Finally, to get the final accuracy, the average of these 10 repetitions are used. Different Scenarios for categories according to input data sets for categorization based on the TP, FP, TN, FN are given in the (Table п).

TABLE II
SCENARIOS FOR CATEGORIES

Type of Record		Predicated Records	
Actual Records	Kind of Categories	Category -	Category +
	Category -	TN	FP
	Category +	FN	TP
	Category +	FN	TP

- TN: the number of people who are not trusted and properly sorted;

TP: number of people who are trusted and properly sorted;

FP: the number of people who have been diagnosed as trusted while not trusted in the dataset;

FN: distrusted number of people who have been diagnosed as trusted in data set;

Accuracy is the most important criteria for determining the efficiency of a standard clustering algorithm. This criterion shows that what percentage of the total data set is correctly classified.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (11)$$

According to Fig. 4, it is shown that the proposed method using decision tree j48 worked better than other methods and increased the precision of trust.
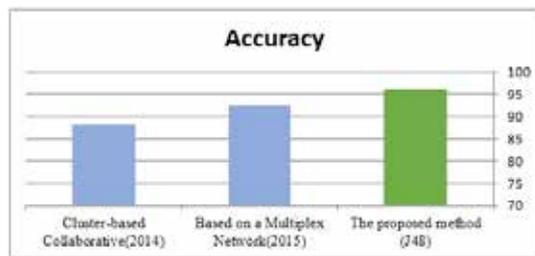

Fig. 4. Comparison of proposed method with the earlier work

CONCLUSION

In this paper, the combination of a multiplex network and identified communities were used. In this respect, five characteristics were obtained in effectively diagnosing trust. By comparing the proposed method with other works, it was found that the learning methods have high accuracy in this field. The decision Tree increased trust accuracy by 96 percent. It should be pointed out at the end the work, in addition to the prediction of relationships of trust, distrust was also determined between the users. To improve this method, we can define other characteristics for trust. Other algorithms of community detection can be used and to find the members of the Association, functions such as special vector and the highest degree can be used. Also, other classification algorithms can be used in machine learning.

REFERENCES

[1] Wang Y., Wang X., Tang J., Zuo W., Cai G. 2015: Modeling Status Theory in Trust Prediction. The AAAI Conference on Artificial Intelligence (AAAI).

[2] Chakraborty P., Karform, S. 2012: Designing trust propagation algorithms based on simple multiplicative strategy for Social Networks. Procedia Technology, Vol. 6, pp.534-539.

[3] Tang, J., Gao, H., Hu, X., Liu, H. 2013: Exploiting homophily effect for trust prediction. In   WSDM., ACM, pp. 53-62.

[4] Javari, A., Jalili, M. 2014: Cluster-Basad Collaborative Filtering for Sign Prediction on Intelligent Systems and Technologi., (TIST), vol. 5,no. 2,p.24.

[5] Beigi, Gh., Jalili, M., Alvari, H., Sukthankar, G. 2014: Leveraging Community Detection for Accurate Trust Prediction. in Proceedings of the sixth ASE international conference on Social Computing, Stanford, CA, USA.

[6] Torkzadeh Mahani, R, Analoui, M. 2015: Trust Prediction in Multiplex Networ. International conference on Knowledge-Based Engineering and Innovation (KBEI), pp263-268.

[7] Kasprzak, R. 2012: Diffusion in networks. Jornal of Telecommunications and Information Technology, pp.99-106.

[8] Jierui, Xie, Boleslaw, K. Szymanski and Xiaoming Liu, 2011" SLPA: Uncovering Overlapping Communities in Social Networks via A Speaker-listener Interaction Dynamic Process", IEEE ICDM workshop on DMCCI.

[9] Raghavan, U. N., Albert, R., Kumara, S. 2007: Near linear time algorithm to detect community structures in large-scale networks. Physical Review. E, vol. 76, p. 036106.

[10] Combe, D., Largeron, C,. Egyed-Zsigmond, E., Géry, M. 2010: A comparative study of social network analysis tools. International Workshop on Web Intelligence and Virtual Enterprises.

[11] Alvari, H., Hashemi, S., Hamzeh, A. 2011: Detecting overlapping communities in social networks by game theory and structural equivalence concept. Artificial Intelligence and Computational Intelligence, pp. 620–630.

[12] Waltman, L., van Eck, N. J. 2013: A smart local moving algorithm for large-scale modularitybased community detection. CoRR, vol. abs/1308.6604.

[13] Zheng, X. 2015: Trust Prediction in Online Social Networks. A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computing Faculty of Science and Engineering Macquarie University.

[14] Zheng, X., Wang, Y., Orgun, M. A., Zhong, Y., Liu. G. 2014: Trust prediction with propagation and similarity regularization. In 28th AAAI Conference on Artificial Intelligence, pages 237–243, Quebec City, Quebec, Canada.

[15] Zhang, H., Wang, Y., Zhang, X. 2013: The approaches to contextual transaction trust computation in e-commerce environments. Security and Communication Networks.

[16] Yao, Y., Tong, H., Yan, X., Xu, F., Lu J. 2013: Multi-aspect + transitivity + bias: An integral trust inference model. IEEE Transactions on Knowledge and Data Engineering.

[17] Leskovec, J., Huttenlocher D., Kleinberg, J. 2010:Predicting positive and negative links in online social networks. in In Predictings of the 19th international conference on World Wide Web.