Community Discovery Algorithm Based on Parallel Recommendation in Cloud Computing

Jian-li Zhai(✉) and Fang Meng

Huali College Guangdong University of Technology, Guangzhou 511325, China
zhaijianli2033@163.com

Abstract. In the cloud computing environment, traditional social network community discovery algorithms have low accuracy in social network community discovery, leading to information waste, community overlap and low scalability, and unable to achieve ideal computing results. Therefore, a social network based on parallel recommendation is proposed. Network community discovery algorithm. By mining the candidate trusted user set, the number and composition of the community are obtained, and the communication units are divided into overlapping communities and non-overlapping communities according to the different numbers of communities belonging to the nodes in the network. Combining the mining of candidate trusted user sets and community division, social networking is realized Network community discovers and calculates. Experiments show that the algorithm improves the accuracy and stability of social network community discovery, and has good application value.

Keywords: Social networks · Recommendation system · Overlapping communities

1 Introduction

Cloud computing is an addition, use, and interaction model of the Internet-based services, often involving the use of the Internet to provide dynamic, easily scalable and often virtualized resources. Cloud is the network, a metaphor for the Internet. In the past, clouds were often used to represent telecom networks [1], and later to represent the abstraction of the Internet and underlying infrastructure. Cloud computing consists of a range of resources that can be dynamically upgraded and virtualized, shared by all cloud computing users and easily accessed over the network, without the need to master cloud computing technology. It only need to rent cloud computing resources according to the needs of individuals or groups. With the rapid development of Internet technology and the gradual popularization of mobile terminals and mobile Internet, more and more people have become a member of social networks. The continuous development of social networks has led to explosive growth in the size of social networks.

With the development of social networks, various forms of social media [2], such as facebook, twitter, Sina Weibo, WeChat, and so on, have realized peer-to-peer

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 348, pp. 192–201, 2021.
https://doi.org/10.1007/978-3-030-67874-6_18

information interaction among network users, and users can record everything in their lives anytime and anywhere. Including what you see and hear in your work and what you think about hot topics. Users are not only recipients of information, but also publishers and communicators of information, so as to gain more opportunities to be recognized by others. The change of the information transmission mode greatly reduces the communication cost among the network individuals, it is also possible to spontaneously form different communities by participating in line-on-line activities in a specific area with those who own common attributes. Real social network is a multi-dimensional network, users have emotional and preference attributes, there are various types of contacts between users. If the multiple information can be used to detect the community structure in the social network correctly, then the other members can be inferred from the known members.

In recent years, social network community discovery algorithms based on parallel recommendation have been put forward [3]. However, most of these methods only focus on the information of single dimension in the network. The parallel recommendation algorithm is faced with a large amount of data and high complexity. Since most of the traditional social network community discovery algorithms can only be applied to small-scale networks or experimentally generated networks, when the number of users in the network is large. Using the traditional parallel recommendation algorithm is limited by the complexity of the hardware and the algorithm itself. Therefore, it is very difficult to deal with such a large amount of data efficiently, which seriously restricts the social network community discovery algorithm for parallel recommendation in large-scale social networks. Based on this, the design of social network community discovery algorithm based on parallel recommendation in cloud computing is proposed. Because of the large amount of data in cloud computing environment, the stability is low, the information is wasted, the overlapping community is not high and the scalability is not high when traditional classification is used. To this end, a social network community discovery algorithm based on parallel recommendation is proposed and designed. The validity of the method is verified in the simulation platform [4]. The results show that LER algorithm can improve the computational accuracy of social network community discovery, reduce the waste of resources, and improve scalability.

2 Design of Social Network Community Discovery Algorithm

Early community discovery algorithms are implemented by hierarchical clustering, but traditional hierarchical clustering methods tend to ignore the members who are less connected to the community. In order to avoid the disadvantages of the traditional method, the LER algorithm uses a new way to detect the community.

2.1 Mining Candidate Trusted User Sets

At present, trust relationship is the most widely used socialized relationship in social network community discovery algorithm. In real life, people ask friends for advice, users rely on the items recommended by their friends, and target users add explicit trust relationships on their own on social platforms. The target user's evaluation of the item depends to some extent on the trust user's rating of the item. This feature is used to mine candidate user sets. Usually, when users add trust relationships on social platforms [5], they are influenced by the user's communication circle and the user's own character and habit. Trust information that can be collected is often sparse. Therefore, only directly using explicit trust relationships is a very limited set of families to improve the accuracy of recommendations and alleviate the problem of sparse trust information.

Figure 1 shows that trust relationship is represented by graph, there are many connections between users, there will be a lot of common ground, there will be different trust relations, with strong expansibility [6].

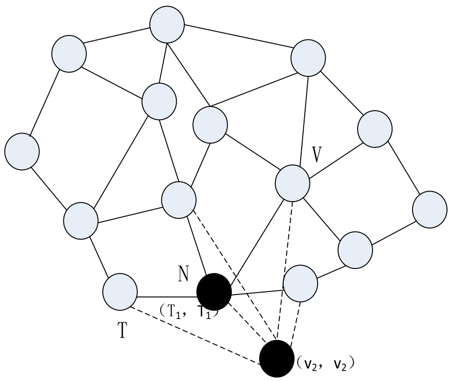


Fig. 1. Trust relationship Association Graph

For this trust relationship characteristics, cluster technology is used to calculate. The social friends in the same trust community as the target user are extended to the candidate trust user set of the target user. For a user set of n users $U=\{u_1, u_2, \ldots, u_n\}$. The item set for m projects is $I=\{i_1, i_2, \ldots, i_m\}$, and the project category set of l categories $C=\{c_1, c_2, \ldots, c_l\}$. We use indicator variables $T_{\mathrm{u,\,v}} = 1$ or 0 indicates whether there is an explicit trust relationship between user u and user v. Use symbol $N_{\mathrm{v}}^{\mathrm{ci}}$ represents the number of scores for items of user v in item classification ci. when the user u explicitly trusts the user v, and when user v has scored the item under a project category, it can be thought that user u initial trust user v under item classification ci. In this way, the initial trust relationship under each category is obtained. In the next step, the initial trust relationship is used to mine trust communities through community discovery methods under each project category. The initial trust conditions are as follows:

$$T_{\mathrm{u,\ v}} = 1 \text{且} N_{\mathrm{v}}^{\mathrm{ci}} > 0 \tag{1}$$

After calculating the initial trust condition, when user v_1 initially trusts v_2, there is a directed edge from v_1 to v_2. The algorithms are as follows:

First step: Specify a unique tag value for all user nodes, such as the user tag initially for the user id.

$$\forall_{\mathrm{u}} \in V : I_u = U_{id} \tag{2}$$

Of which, $\forall_{\mathrm{u}}$ represents the result of tag value; I representing the simplified coefficient of the network model j, this calculation does not do the directional analysis; k_j represents the training features representing the j convolution layer; U_{id} represents the commonality of the i trusted user.

Step two: to adjust the label values of all user nodes, each user needs to traverse and count the current tab values of all his neighbors, and take the label value with the largest number of occurrences as the new label for the user. When multiple label values appear in the largest number of times, a label value is randomly selected as the user's new label.

$$\forall_{\mathrm{u}} \in V, 1_u = angmsc_k \left| N^K \right| \tag{3}$$

Of which, k represents the label, N^K represents a collection of u-trusted users with a tag value of k, 1_u represents the value of the user tag, this calculation does not do directional analysis.

The number of iterations required to divide a community to a stable state is generally five, when more than 90% of the nodes in the network have been divided into the correct communities. After mining the trust user set, there are some overlapping communities, and there are some differences in the community division [7]. Therefore, after mining the candidate trust user set, the community division is carried out.

2.2 Community Division

Through the community discovery, we can dig out the number and composition of the community from the network. It can help us to discover the deep laws in the network, analyze the realistic significance of the community division representation in the network, and thus help solve the practical problems. According to the different number of communities belonging to the nodes in the network, the communities are roughly divided into overlapping communities and non-overlapping communities [8]. In overlapping communities, each node can belong to multiple communities. Non-overlapping communities only allow each node to belong to one community, and there are no nodes that exist across multiple communities. This paper defines the community based on the idea of edge partition, and puts forward the HCL algorithm. In the HCL algorithm, the overlapping community merge and partition error correction are incorporated into the two steps. In each iteration, the similarity between edges is calculated, and the similarity between edges is calculated in each iteration. All the edges are divided into specific communities, and the combination is applied to community discovery of large-scale networks. The edge similarity algorithm is formulated as follows:

$$S(eik, ejk) = \frac{|n + (\mathrm{i}) \cup \mathrm{n} + (\mathrm{j})|}{|n + (\mathrm{i}) \cup \mathrm{n} + (\mathrm{j})|} = \frac{|\mathrm{n} + (\mathrm{i}) \cap \mathrm{n} + (\mathrm{j})|}{|\mathrm{n} + (\mathrm{i})| + |\mathrm{n} + (\mathrm{j})| - |\mathrm{n} + (\mathrm{i}) \cap \mathrm{n} + (\mathrm{j})|} \quad (4)$$

Of which, $n + (\mathrm{i})$ represents a collection of neighbor nodes for node S. And ejk and jk are S sides, this calculation does not do directional analysis.

The basic process of the HCL algorithm is: through the hierarchical clustering algorithm, the edges of the network are clustered into a community, because each node may have multiple connected edges, and these edges are divided into different communities, and each node may be divided into multiple different communities. The overlapping community structure is obtained. HCL algorithm can calculate the similarity between edges [9], and the larger the similarity is, the better the edge can be merged into a community.

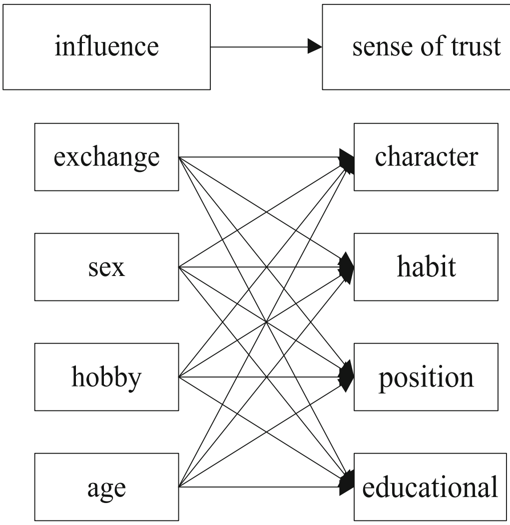


Fig. 2. Community Partition proc

Figure 2 mainly demonstrates the process of community division by analyzing the age distribution, sex, education level, interests and interests of community members, this paper studies the interaction between individuals in the community and the evolution of community structure. Through the analysis of the community structure, the flawlessness, robustness and stability of the entire community network are studied, the impact of key nodes in the community on information dissemination is analyzed, and the stability of the community structure is analyzed through the analysis of the community. The structure completes the design of the social network community discovery algorithm.

The HCL algorithm improves the accuracy and processing speed of community partition to a certain extent, which is of great significance to the design of social network community discovery algorithm.

2.3 Implement Social Network Community Discovery Computin

Combined with mining candidate trust user set and community partition, the implementation of social network community discovery calculation is realized. Next, the implementation process of each step will be described in detail in this paper.

According to the influence degree of attribute features on the classification results, the features are classified. In classification, a large weight is given to certain features of nodes with greater impact on community discovery, and a smaller weight is given to certain types of features of nodes with less impact on community discovery. Based on this, the LER algorithm is proposed for calculation and generation. The formulas are as follows:

$$X_i = \min c\left(\sum_{m=1}^{n} mxc_j\right) + k\mu_j \tag{5}$$

Of which, n represents the number of nodes in the simulated network; m represents the number of edges in the simulated network; k represents the average node degree of each node; μ_j represents the proportion of the edges of each node connected to a node that is not in the same community as the sum of all the edges of the node, the larger the value, the less obvious the community structure of the simulated network is; min$c$ represents the number of nodes owned by the smallest community in the generated simulation network; mxc_j represents the maximum number of nodes owned by the community in the generated simulated network.

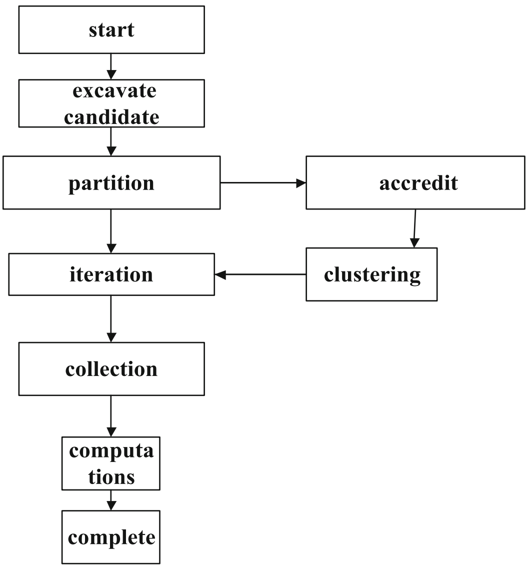


Fig. 3. LER Algorithm flow-process diagram

Figure 3 shows the LER algorithm to discover the social network computing process for the community. This calculation method can reduce the degree of community ambiguity, calculate in the case of large community size, and improve the precision of community calculation. This is because the LER algorithm uses multiple iterations to calculate the spanning tree to distinguish the community structure in the network, and the cohesion within the community is expressed by the number of spanning trees. This makes the real community size of the LER algorithm smaller in the network [10–12] and facilitates the generation of more spanning trees that cover the nodes within the entire community. Therefore, it is easier to detect the community structure in the network effectively. It is shown that the LER algorithm can improve the accuracy of the calculation and can carry out the effective calculation under the condition of big data. It can effectively highlight the topological structure of the network, thereby effectively mining the community structure in the network, and realize community discovery calculations based on parallel recommendations in cloud computing [13–15].

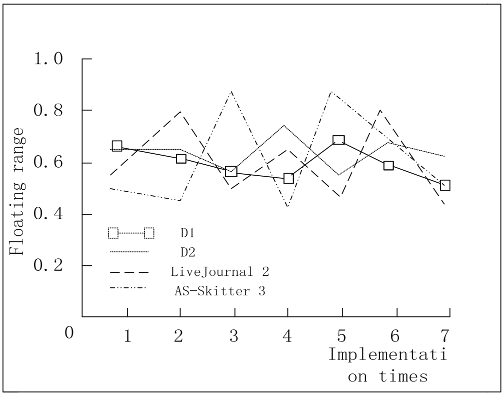
The LER algorithm also has a huge storage effect. Generally, the data is a graph with more edges than points, so the graph data is stored by the way of point segmentation. It can avoid too much redundancy in the stored procedure of edges, and the interaction between nodes and their neighbors only needs to satisfy the exchange law and union law. This method can effectively reduce the network transmission and storage overhead. The underlying implementation process is to store the edges in each node, and when data interaction occurs, it can be transmitted by broadcasting the nodes between each machine. The cost is that multiple redundant backups are needed for each node's attributes, and there is data synchronization overhead when the node update operation is needed.

3 Experimental Demonstration and Analysis

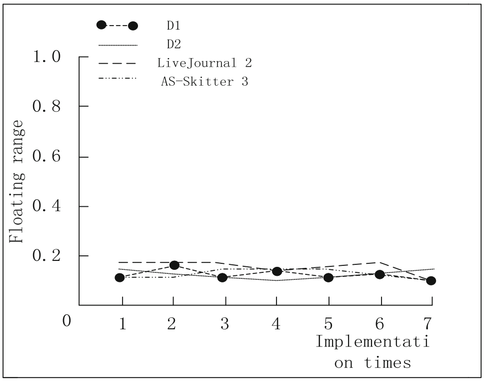
In order to verify the validity of the LER algorithm in this paper, a simulation experiment is designed. In order to ensure the rigor of the experiment, the traditional community discovery algorithm is compared with this algorithm.

3.1 Stability Testing

In order to test the stability of LER algorithm, the traditional algorithm is compared with LER algorithm. Four different data sets, LiveJournal 2, AS-Skitter 3, D1 and D2, were used to test the framework. In four datasets, LiveJournal is an online dating blog network of friends; ASkitter is a network that describes the topology of Internet; data sets D1 and D2 are simulated networks generated by LFR. The algorithm in this paper and the traditional algorithm are used to test the four test data sets, and the test results are shown in Fig. 4.



(a) Floating range of traditional algorithm



(b) The floating range of the algorithm in this paper

Fig. 4. Stability comparison

According to the comparative analysis of Fig. 4, when the computing time of the LER algorithm is less affected by the change of the number, the stability is high and the number of tasks is large, the running time of the algorithm is relatively small, and the proportion of the total running time of the calculation is smaller when the number of tasks is more than the number of tasks. The amplitude of the fluctuation produced by the algorithm in this paper is small and basically fluctuates about 0 amplitude. But in the traditional image classification process, the fluctuation is large, the highest is about ± 5 dbs. When the number of algorithms increases and the total running time of the algorithm decreases, the traditional algorithm floats greatly and its stability is low.

Through the above analysis, we can basically determine the validity of this algorithm. When designing community discovery algorithms in cloud computing environment, the data can be classified and calculated accurately and stably, which is convenient for further management and analysis, and the stability is very high.

4 Conclusion

In recent years, with the rapid development of social networks, there are new requirements for community recommendation algorithms based on parallel recommendation. Among them, community discovery and parallel recommendation have become the focus of academic research. Because of the scale of today's social networks, the relationships within a single network are also very complex, and there is a relationship between multiple networks because of the same account number.

Therefore, this paper proposes a community discovery algorithm based on parallel recommendation, combined with the mining of candidate trusted user sets and the division of overlapping communities and non-overlapping communities, to realize the discovery calculation of network communities. By analyzing the community structure, the accuracy of information recommendation can be improved, and it is of great significance to study the parallel recommendation algorithm recommended by social network communities.

Acknowledgments. T his work is supported by Second batch of school-level scientific research projects of Huali College of Guangdong University of Technology (project number: HLKY-2018-ZK-07)

References

1. Xu, W., Lin, B., Lin, S., et al.: Research on a method of social network community discovery based on user interaction behavior and similarity. Inf. Network Secur. **33**(7), 77–83 (2015)
2. Lu, Y., Cao, B., Rego, C., et al.: A Tabu search based clustering algorithm and its parallel implementation on Spark. Appl. Soft Comput. **63**, 97–109 (2018)
3. Longyuan, Zhang, J.: Parallelization of community discovery algorithms based on local expansion. Mod. Comput. **21**(11), 41–45 (2016)
4. Wang, S., Xu, G., Shi, S.: Label propagation overlapping community discovery algorithm based on large groups. Electron. Measur. Technol. **23**(2), 45–49 (2016)
5. Liu S, Liu D, Srivastava G, et al. Overview and methods of correlation filter algorithms in object tracking. Complex Intell. Syst. 1–23 (2020). doi:10.1007/s40747-020-00161-4
6. Wang, S., Xu, G., Shi, S.: Maxima-based label propagation overlapping community discovery algorithm. Electronic Measur. Technol. (2): 45–49 (2018)
7. Wahab, O.A., Bentahar, J., Otrok, H., et al.: Towards trustworthy multi-cloud services communities: a trust-based hedonic coalitional game. IEEE Trans. Serv. Comput. **PP**(99), 184–201 (2018)
8. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)
9. Zhao, R., Wu, Y., Chen, X.: Community discovery and visualization algorithm for large-scale social networks. J. Comput.-Aided Des. Graphics **29**(2), 328–336 (2017)
10. Wen, R.: Research and implementation of personalized recommendation system based on social network. Zhejiang University of Technology **9**(3), 28–36 (2015)
11. Wang, J.: Research on social network community discovery method based on local expansion. Yanshan University **9**(3), 28–36 (2015)

12. Zhao Runqian, WuYu., Xin, C.: Large-scale social network community discovery and visualization algorithm. J. Comput. Aided Des. Graphics **29**(2), 328–336 (2017)
13. Liu, S., Wang, S.: Trajectory community discovery and recommendation by multi-source diffusion modeling. IEEE Trans. Knowl. Data Eng. **29**(4), 898–911 (2017)
14. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. Mob. Networks Appl. **24**(1), 5–17 (2019)
15. Shuai, L., Gelan, Y.: Advanced Hybrid Information Processing Springer International Publishing, USA, 1-594