ISSN 2347 - 3983
Volume 7, No. 9 September 2019
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter13792019.pdf
https://doi.org/10.30534/ijeter/2019/13792019

Large Scale Community Discovery of Complex Networks Based On Approximate Optimization

[1]P. Madhuravani, [2]Kesani Yamini, [3]T Nirmala
[1,2,3] Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, Telengana.

ABSTRACT
Study of structural features associated with complex networks has its utility. In the process a key role is played by community. Online networks are showing unprecedented growth in complexity. In this context, the traditional approach used to discovery community is not sufficient to deal with large networks. Thus it leads to the problem of inability to discovery communities efficiently in the large scale networks. To overcome this problem, a Parallel Community Discovery (PCD) algorithm is proposed based on the concept of approximate optimization. The algorithm has two parts. They are known as mountain model and landslide algorithm. The former is meant for providing approximate selection of nodes that are used for representation in graph theory. It shows users and the data posted by them in the network. The landslide algorithm on the other hand updates the modularity from time to time incrementally based on the aforementioned technique named approximate optimization. Thus the proposed methodology helps in discovering communities in parallel. The process of clustering on the modularity plays important role in mountain model as it employs weight associated with each edge in the given network. The communities also exhibit relationship among them. However, they are simplified by the landslide algorithm thus helping in extraction of community structural features from large networks.

Key words: Distributed computing, community discovery, graph theory, approximate optimization

1. INTRODUCTION

In the contemporary era, there are increased examples of complex networks in the real world. Best example is the social networks besides other such networks like client transaction networks and any sort of distribution reference based networks. There is complex connectivity among nodes and there are large number of networks involved and the whole network becomes much more complex [1]. Community structures can be extracted from such networks to have useful applications. For instance, it is possible to identify group of nodes that can be combined to form a network and there might be number of connected groups. There are so many edges in the group and with high density probably. Between group edges may exhibit lower density as explored in [2]. In such scenario, it is important to know the purpose and significance of structural features that can be extracted from the complex networks. The communities, information related to broadcasting, reference to points of interest and many other crucial features can be extracted from the same. In the process it is highly important to discovery communities to realize various applications [3]. Previous investigations on this research area focused on the small networks with possible and simple structure. The rationale behind this is the difficulties in computational resources for managing data and analyzing it.

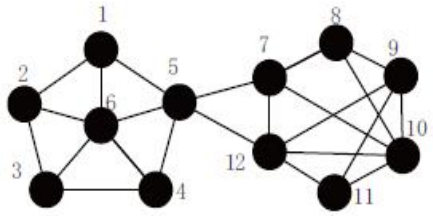

Figure 1: Sample network

To overcome the drawbacks of existing systems, the proposed research is motivated by different observations. Social networks became very large and complex and exhibit millions of users. For instance, 13.5 billion of users are found active in a month when Face book is considered [4]. With such growth of networks, it is important to devise algorithms in order to extract communities with high scalability in the presence of complex relationships and rapid growth of data from time to time. As the networks grow complex, there might be many hidden communities that may come across. Therefore, it is important to analyze, correlate and predict behavior of users to help different real time applications like advertising, marketing and so on. Thus, there is significance for the study of internal structural features of different communities discovered

from large networks. These structural features are used in many applications when efficient algorithms are employed. A sample network is shown in Figure 1. However, there are very large numbers of structural features that can be extracted. The remainder of the paper is structured as follows. Section 2 provides review of literature on the research topic. The Section 3 on the other hand provides the proposed methodology. Section 4 presents experimental results while section 5 concludes the paper and provides scope for future work.

2. RELATED WORK

This section provides the prior works on the extraction of communities from complex networks.

2.1 Powers-Law Distribution of the World Wide Web

World Wide Web (WWW) is an example for large network with different structural features involved. Erdos-Renyi (ER) theory has revealed that random networks have different scaling properties in the systems associated with WWW. However, there is an inconsistent theory that is lacking empirical observations with respect to extracting features from the WWW [4]. With respect to connectivity a vertex which more connections are said higher in connectivity rate according to [1] . Thus it is observed that a network grows with complexity and structural differences. The older vertices grow in more connections when compared with younger ones. This will lead to a phenomenon known as rich get richer.

2.2 Finding and Evaluating Community Structure in Networks

Identifying and evaluating community structures in a large network is very important. There are many connected sub groups when a complex network is considered. There are algorithms that focused on two important and distinctive features. First, they consider pruning edges from network in order to split into multiple communities. The edges pruned are used to establish connectivity among different connected networks [5]. A measure is also available for finding the strength of community structure discovered by algorithms. This objective metric is useful to understand the dynamics of the structures. Thus the algorithms are evaluated to be highly effective or not with respect to discovering community structures from complex network. The networks may be real world or computer generated that can help in demonstrating proof of the concept [6].

Finding community structures [7], generic observations of networks [8], spectral clustering for community detection [9], effective community detection process [10], robust local community finding [11], identification of overlapping communities [12], triangle driven detection of communities [13], time series based clustering [14], search process for overlapping communities [15], approximate closest community [16], finding community structure [17], subspace based approach [18], detecting overlapping community using seed expansion [19] and network clustering and modularity [20] are other related researches found in the literature. From the literature it is understood that there is need for fast and parallel discovery of communities which is realized in this paper.

3. COMMUNITY STRUCTUER vs. FUNCTION PREDICTION

It is understood from the literature and empirical study that the improvements in community structures led to improvements in predictions. When related nodes are clustered in complex networks, it results in communities. This kind of approach for detecting communities is widely used in the real world. However, it is found to be computationally complex and difficult as well. With respect to protein function context, these issues are understood with empirical study. First, the conventional method is employed in order to generate communities. Second, a better method is used to identify communities and predict the objective functions. It is understood that community information has significance as it is used in multiple applications. The observations revealed that two distinct and scientific communities can exist. First, different cost functions and optimizations can provide community structures and related solutions. Second, it is possible to extract functional information associated with the nodes in the large network which is denoted as an interactive dataset.

4. ALGORITHM FOR DISCOVERING COMMUNITIES

When graphs are used to find community structure, it led to useful applications. Graph data analysis became easier. Moreover, graphs assume importance due to unprecedented growth of data that needs to be represented. Such data is being made available in WWW and also social networks. Using graph data is increased in the research and academia in the last many years. In this paper also, a simple and effective method is used in order to explore communities from the large networks. There is modified algorithm known as Sequential Louvain Algorithm used for community

detection. The proposed method is nothing but distributed memory parallel algorithm that focuses on the first iteration which is costly. It leads to the initial approach to have parallel processing. With the MPI setup involving as many as 128 processes running in parallel, the proposed method is evaluated.

5. IMPLEMENTATION

The proposed system implemented using the Java programming language. It has provision for data processing and visualization. It has data structures that can help in representing complex data. For instance, it could provide data structures to handle graph theory and implement the data structure. Such data structure is used in order to have efficient navigation among data and discover communities from the complex networks. Java application is used in order to show the effectiveness of the proposed system. Java is object oriented and it has plenty of good features. One of them is platform independent. Though it is developed in Windows platform, its byte code can run in different environments and platforms.

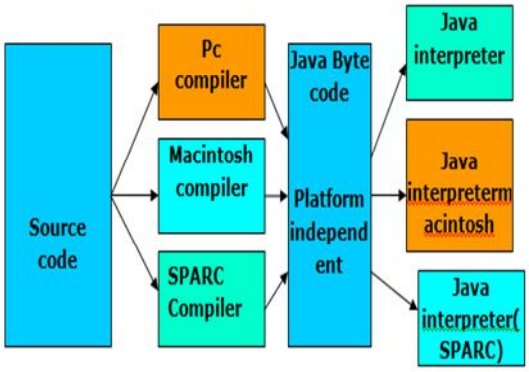

Figure 2: Java source code to execution process

As shown in Figure 2, Java runtime environment and compile time environment is shown. This knowhow is useful to work with Java applications with ease. When Java source code related to this work is given as input, the source code is subjected to compilation. It is done by the Java compiler provided. Once the compilation is done successfully, it results a new file called byte code saved with .class extension. Then at runtime Java Virtual Machine (JVM) takes care of execution of byte code in a platform independent manner. The code at runtime is interpreted to have machine code or native machine code and executed. There are many good features of Java platform. The Java is made simple with its syntax and object orientation resembling that of C++. It is object oriented to support modular development and promote reusability as much as possible. Java applications are robust as they can run in different platforms with reliability. Moreover, it is typed language and checking is made at compile time and also runtime. Java also gets rid of memory issues with its garbage collection process which takes place automatically.

6. MODULES INVOLVED IN THE SYSTEM

The proposed system implemented with different modules. They are known as admin module, mountain module, landslide strategy module and user module. More details of these modules are provided here.

6.1 Admin Module

This is the module associated with administrator user. This user can authorize other users in the system. This user can view front requests, different communities extracted in the network, check popularity of communities, find the best or parallel community, compare communities and perform activities related to landslide strategy module and mountain model.

6.2 Mountain Model

This model plays important role in the proposed system. The mountain model is part of this research and it is based on different concepts like graph theory, approximate optimization and modularity. It can be used to sort chain groups based on the weights of edges in the network. The community structures extracted from the complex network has its own features. Based on them, some chain groups may raise up like a mountain while other chain groups my fall down. Thus it is observed that a suitable number of such groups are found at the mountain top in order to generate new communities.

6.3 Landslide Strategy Module

It is another important module in the proposed application. There are many numbers of nodes and edges in the given network. Assuming that they do not change after community merging phenomenon, it is observed that there is equality between the sum of edges and the number of edges found in the new community. There are many edges found between the new community and other ones with respect to equality and merging of them in order to form different communities.

6.4 User Module

This module is related to user operations. It has certain sub modules like OSN construction and community

creation. With respect to OSN construction there are many steps. An OSN is built in order to have online social networking features in the application. This will help in creating or registration of new users. Thus credentials are given to users for further usage of the system. The after due authentication, users can send messages either publicly or privately besides building various options and posting and sharing with other. The users can also perform search of other user profiles and posts. Users are also involved in accepting friend requests and sending friend requests to other users. Thus OSN is built with different users. Then it can help in creating communities.

7. EXPERIMENTAL RESULTS

With the aforementioned environment, experiments are made and the results are presented in this section. The empirical results revealed that the proposed method achieved the performance improvement when compared with the state of the art.

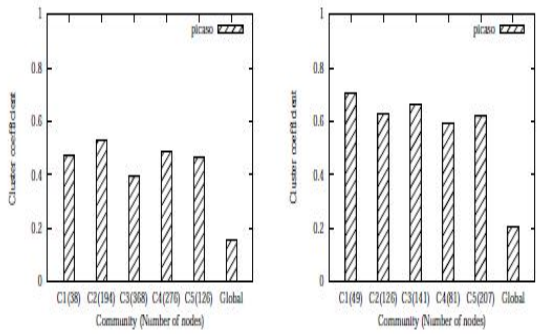

Figure 3: Experimental Results

From the results shown in Figure 3, it is understood that different communities are discovered as shown in horizontal axis. The vertical axis shows the cluster coefficients. Each community refers to number of nodes associated as cluster. Different clusters have different number of nodes and corresponding cluster coefficient. There is global cluster that has its importance with different cluster coefficient.

8. CONCLUSION AND FUTURE WORK

In this paper a novel community discovery algorithm is proposed. This algorithm is meant for dynamically discovering communities from large social networks. The algorithm has different provisions for ensuring it. Different innovative approaches are exploited by the algorithm. They are known as mountain model and landslide strategy algorithm besides approximate optimization technique. Apart from all these aspects, game theory is also involved. There is aggregation of different things in order to achieve the functionality required. Thus the proposed system is capable of discovering new communities. The results of experiments with Java application proved that the proposed method is useful as it was evaluated with complex networks. The results also reveal that the proposed method is better than state of the art. In future, the proposed method can be enhanced to work with more increased number of nodes with scalability and to give guarantee of performance. Another direction is to identify overlapping communities and take measures to handle them for the benefits of applications.

REFERENCES

1. Barabasi, R. Albert, H. Jeong, and G. Bianconi, “Power-law distribution of the world wide web,” Science, vol. 287, no. 5461, Art. No. 2115, 2000. https://doi.org/10.1126/science.287.5461.2115a
2. M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” Physical Review E, vol. 69, no. 2, Art. No. 026113, 2004. https://doi.org/10.1103/PhysRevE.69.026113
3. J. Lee, S. P. Gross, and J. Lee, “Improved network community structure improves function prediction,” Scientific Reports, vol. 3, no. 2, Art. No. 2197, 2013. https://doi.org/10.1038/srep02197
4. Wearesocial, “Gigital in 2016,” 2016. [Online]. Available:http://www.wearesocial.com
5. C. Wickramaarachchi, M. Frincuy, P. Small, and V. K. Prasannay, “Fast parallel algorithm for unfolding of communities in large graphs,” in Proceedings of 2014 IEEE High Performance ExtremeComputing Conference. IEEE, 2014, pp. 1–6. https://doi.org/10.1109/HPEC.2014.7040973
6. M. E. J. Newman, “Fast algorithm for detecting community structure in networks,” Physical Review E, vol. 69, Art. No. 066133,2004. https://doi.org/10.1103/PhysRevE.69.066133
7. Clauset, M. E. J. Newman, and C. Moore, “Finding community structure in very large networks,” Physical Review E, vol. 70, no. 2, Art. No. 066111, 2004. https://doi.org/10.1103/PhysRevE.70.066111
8. J. Qiu, J. Peng, and Y. Zhai, “Network community detection based on spectral clustering,” in Proceedings of the 2014 InternationalConference on Machine Learning and Cybernetics. IEEE, 2014, pp.648–652.
9. Y. Ruan, D. Fuhry, and S. Parthasarathy, “Efficient community detection in large networks using content and links,” in Proceedings of the 22nd International Conference on World Wide Web. ACM, 2013, pp. 1089–1098.
10. Y. Wu, R. Jin, J. Li, and X. Zhang, “Robust local community detection: on free rider effect and its elimination,” Proceedings of VLDB Endowment, vol. 8, no. 7, pp. 798–809, 2015.
11. X. Zhang, H. You, W. Zhu, S. Qiao, J. Li, Z. Zhang, and X. Fan, “Overlapping community identification

approach in online social networks," Physical A, vol. 421, pp. 233–248, 2015.
12. Prat-P´erez, D. Dominguez-Sal, J.-M. Brunat, and J.-L. Larriba-Pey, "Put three and three together: triangle-driven community detection," ACM Transactions on Knowledge Discovery from Data, vol. 10, no. 3, Art. No. 22, 2016. https://doi.org/10.1145/2775108
13. L. N. Ferreira and L. Zhao, "Time series clustering via community detection in networks," Information Sciences, vol. 326, pp. 227–242,2016. https://doi.org/10.1016/j.ins.2015.07.046
14. J. Shan, D. Shen, T. Nie, Y. Kou, and G. Yu, "Searching overlapping communities for group query," World Wide Web, vol. 19, no. 6, pp.1179–1202, 2016.
15. X. Huang, L. V. S. Lakshmanan, J. X. Yu, and H. Cheng, "Approximate closest community search in networks," Proceedings of the VLDB Endowment, vol. 9, no. 4, pp. 276–287, 2015. https://doi.org/10.14778/2856318.2856323
16. X. Li, M. K. Ng, and Y. Ye, "Multi Comm: finding community structure in multidimensional networks," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 4, pp. 929–941, 2014. https://doi.org/10.1109/TKDE.2013.48
17. Mahmood and M. Small, "Subspace based network community detection using sparse linear coding," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 801–812, 2016.
18. J. Whang, D. Gleich, and I. Dhillon, "Overlapping community detection using neighborhood-inflated seed expansion," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 5, pp.1272–1284, 2016. https://doi.org/10.1109/TKDE.2016.2518687
19. Praveen Kumar Kollu1, R. Satya Prasad, intrusion Detection System Using Recurrent Neural Networks and Attention Mechanism, International Journal of Emerging Trends in Engineering Research, Volume 7, No. 8 August 2019 https://doi.org/10.30534/ijeter/2019/12782019
20. T. N. Dinh, X. Li, and M. T. Thai, "Network clustering via maximizing modularity: Approximation algorithms and theoretical limits," in Proceedings of the 2015 IEEE International Conference on Data Mining. IEEE, 2015, pp. 101–110.