



Overlapping Community Detection using Label Sharing Approach

Dr. Kulvinder Singh¹, Dr. Sanjeev Dhawan², Vinay³

^{1,2} Faculty of Computer Science & Engineering, University Institute of Engineering and Technology,
Kurukshetra University, Kurukshetra-136119, Haryana, INDIA

³ Research Scholar, University Institute of Engineering and Technology,
Kurukshetra University, Kurukshetra-136119, Haryana, INDIA

Abstract: Community detection is a crucial task in social network analysis. Social interactions exist at intervals some social context and communities are a basic kind of social contexts. Communities are intuitively characterized as unusually densely connected subsets of a social network. This notation becomes a lot of problematic if the social interactions modification over time. Community mining is one of the major directions in social network analysis. Most of the existing methods on community mining assume that there is only one kind of relation in the network, and moreover, the mining results are independent of the users' needs or preferences. The interplay between social interacts and social contexts are crucial to know the evolution of networks. Thus, it's necessary to each notice communities and tracks their changes. According to the past work related to community detection there were so many algorithms or methods which were successful in detection of communities. In this paper, we develop a new method (label sharing method) with the help of which we will detect communities. In this method, we made a separate table which has the list of communities present in the network and the number of members belonged to these communities.

Keywords: community, community detection, label sharing approach, social networks.

I. Introduction

Social networks are playing a very important role in the present conditions as everyone is running behind success by creating relationships with each other. Social networks provide a base to the users to communicate with the people of their interest with the help of communities. Community is a term which has no proper definition but for ease we can define community as it is a group of people with common features or interests. These interests can be a favorite cricket team, a college, a country or follower of any politician. Users belonging to the same community have strong understanding or relationship as compare to those users which are present in different communities. We have studied a lot about social networks and communities like social networks consists of communities and the communities are the group of users of same interest but the main problem is how to find out that in which community the users present or who are the users present in a specific community. Many researchers have worked on it and they gave their idea about community detection by means of many algorithms or methods. Some of the popular methods were as follows: label propagation algorithm, clique based method for community detection. Some methods which were successful in the detection of community have some disadvantages also. Drawbacks which we understand or faced are like: some methods failed in detection of overlapping communities and which gain success in finding of overlapping communities take so much time. Before discussing our method we would like to define overlapping communities. Overlapping communities consist of a group of people which is present in more than one community. We will try to detect these overlapping communities by a new approach named as label sharing approach. In this approach we made a separate table which maintains the record of the number of communities present in a social network and also the list of the members belongs to each community. The whole working of this approach we will discuss in the working section.

II. Related Work

Ding Xiao *et al.* [2] advised that consistently analyse the matter of mining hidden communities on heterogeneous social networks supported the observation that completely different have relations and different importance with reference to a definite question, the author propose a replacement methodology for learning an optimum linear combination of those relations which may best meet the user's expectation. With the obtained relation, higher performance may be achieved for community mining. The authors approach to social network analysis and community mining represents a significant shift in methodology from the standard one a shift from single-network, user-independent analysis to multi-network, user-dependant and query-based analysis.

Experimental results on Iris information set and DBLP information set demonstrate the effectiveness of our methodology. Guilan Hu *et al.* [3]. The authors advised that private info mining will resolve the hidden relationship and characteristics of the target individuals which may be used for active post operation. The initial options that were enclosed within the personal info data typically have high dimension and redundancy which regularly drags down data processing potency. A feature optimisation methodology is planned here to resolve the matter. The strategy with the aim of knowledge spatiality deduction relies on effective association of rough pure mathematics with PCA approach. The experimental results demonstrate that the hybrid feature optimisation methodology is effective in up classification accuracy. L. Sørensen [4] advised that social networks are the very best growing internet application in terms of users. Totally different surveys show that users are most involved with their privacy in reference to web-based social networks. Anyhow users compete within the variety of friends they'll attach to their own profile. This suggests that the trust relations user area unit mistreatment to determine friends within the internet applications becomes considerably totally different from the trust relations utilized in face-to-face conferences. This paper compared and mentioned a number of the prevailing self-management mechanisms in trust in 3 of the foremost used web based mostly social networking applications and suggests totally different aspects for handling trust from a user-centric style perspective. Steve Gregory [5] planned an algorithmic program for locating overlapping community structure in terribly giant networks. The algorithmic program relies on the label propagation technique of Raghavan, Albert, and Kumara, however is ready to notice communities that overlap. The author's main contribution was to increase the label and propagation step to incorporate info regarding over one community: every vertex will currently belong to up to v communities, wherever v is that the parameter of the algorithmic program. The planned algorithmic program conjointly handles weighted and bipartite networks. It conjointly in no time and may method terribly giant and dense networks in an exceedingly short time. During this work every vertex can't be updated severally that is that the main downside of this analysis.

U Kang *et al.* [7] introduced GIM-V a vital primitive that Pegasus uses for its algorithms to investigate structures of huge graphs. The author conjointly introduced HEigen an oversized scale Eigen thinker that is additionally a district of Pegasus. each GIM-V and HEigen were extremely optimized achieving linear proportion on the quantity of machines and edges and providing nine.2x and 76x quicker performance than their naïve counterparts severally. The authors analyzed terribly giant world graphs with billions of nodes and edges mistreatment Pegasus. Lovro Subelj *et al.* [8] planned a quick methodology for detection of communities in giant advanced networks i.e. Label propagation methodology. The author gave a sophisticated label propagation algorithmic program that creates a hybrid technique for community formation, namely, defensive preservation and offensive enlargement of communities. 2 ways were collaborated in an exceedingly ranked manner, to recursively extract the core of the network. The algorithmic program was evaluated on 2 categories of benchmark networks with planted partition and on nearly twenty five real-world networks starting from networks with tens of nodes to networks with many tens of ample edges. It absolutely was compared to this progressive community detection algorithms and superior to any or all previous label propagation algorithms, with comparable time complexness. S. Simranjit *et al.* [9] planned a system mistreatment neural network and diversified weights supported multilayer text extraction, frequency of communication for friend recommendation from friends of friends. The authors have taken under consideration varied factors which can assign score to his friends of friends and can suggest them a lot of expeditiously. The results weren't reliable and correct that could be a limitation to the present work. A genetic based approach to find communities in social networks has been planned by Brian poet *et al.* [10]. The algorithmic program optimized a straightforward however efficacious fitness operate ready to establish densely connected teams of nodes with thin connections between teams. This methodology was economical as a result of the variation operators were changed to require into thought solely the particular correlations among the nodes therefore reasonably reducing the analysis house of potential solutions. The most downside of this system was it doesn't notice part connected nodes within the network. S. Thayananthporkalanchiam *et al.* [11] represented the matter of community detection in advanced networks as a Multi Objective clustering problem and presented a biological process Multi Objective approach to uncover community structure. The planned algorithmic program optimized 2 objective functions that were ready to establish densely connected teams of nodes having thin lay connections. The strategy generated a collection of network divisions at totally different ranked levels during which solutions at deeper levels, consisting of the next variety of modules that were contained in solutions having a lower variety of communities. optimisation methodology outlined an Objective operate that allowed the division of a graph in sub graphs and take a look at to maximise this objective so as to get the simplest partitioning of the network.

III. Problem Formulated

The main problem which is encountered in community detection algorithms is how to find overlapping communities. Based on past work related to community detection problem we develop an approach named as label shared approach for the detection of overlapping communities.

IV. Working

The algorithm label shared approach works in way as described below:

1. First of all we create a random social network for any number of nodes which are entered by a user who wants to login.
2. After that user is asked for entering the threshold value. Here threshold value tells the contribution of the people present in a community to that community. If the contribution is more than the threshold value then that community is neglected.
3. After entering these two terms, a social network is generated and a random user who wants to check the suggestions from the different communities is asked to enter the email id and password.
4. After entering username and password, a list of friends is suggested from different communities.
5. If a person is suggested from more than one community it means that person is present in two or more communities. So, in this way we overlapping communities are detected.

The output produced from this approach will look like as shown in the figures below:

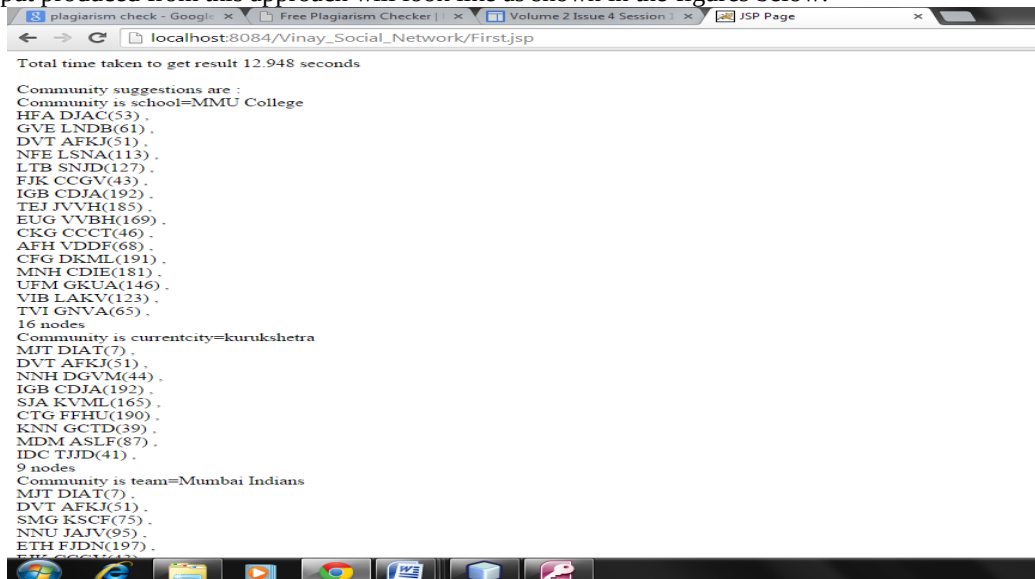


Fig. 1 list of suggestions

Message of overlapping community detection is shown below:

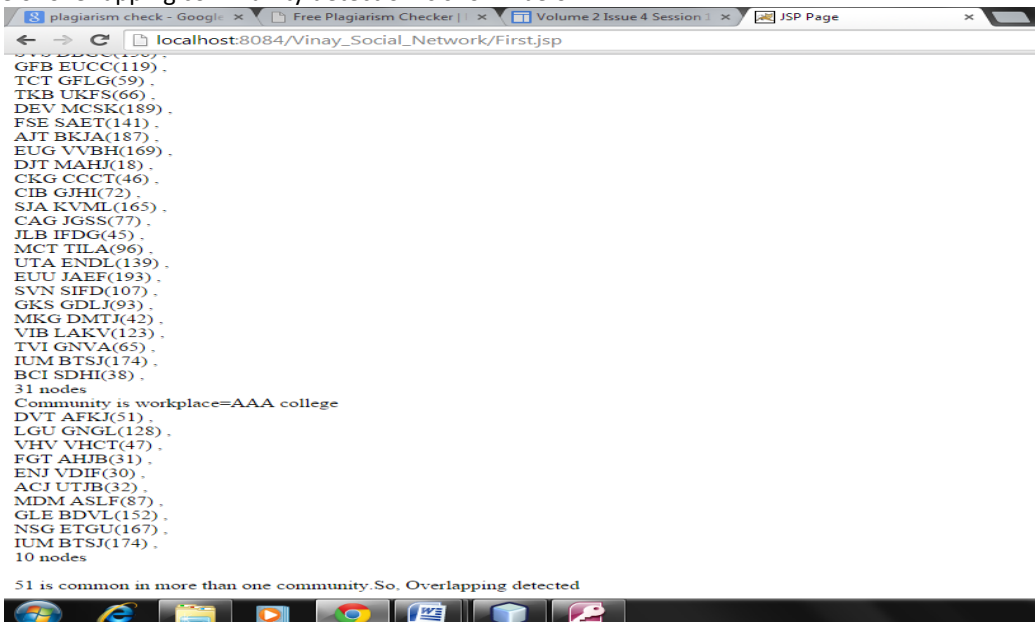


Fig. 2 overlapping message

V. Results

On behalf of working and number of nodes suggested we generate two results which are described as follows:

1. Firstly we fix the threshold value and the social network is generated for different number of nodes says 100, 200, 300 and so on. Suppose we fix a threshold value 12. The graph shown below on behalf of this shows that how the number of suggestions from each communities changes.

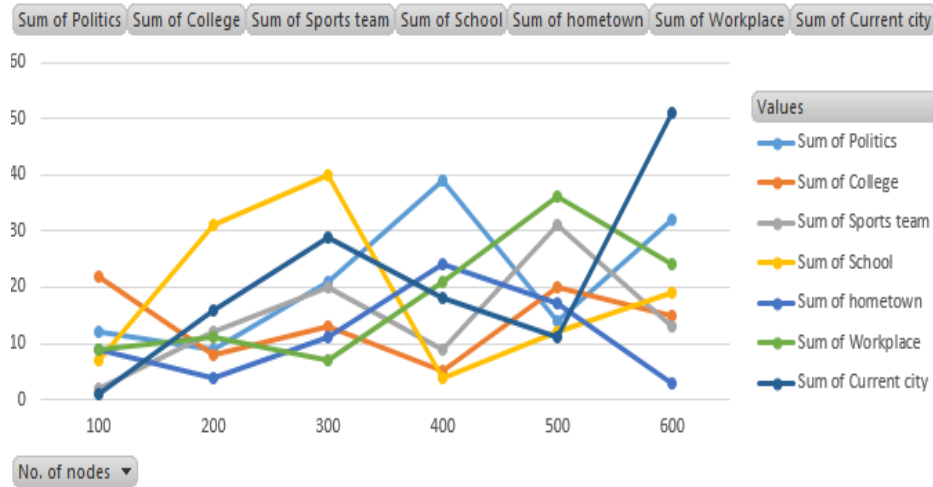


Fig.3 number of suggestions for each community varies for different number of nodes in a social network for a fixed threshold.

2. Second result based on a situation that we fix the nodes in a network and changes the threshold value. So the results produced are shown with the graph shown below in which we will notice how the number of suggestions changes for each community.

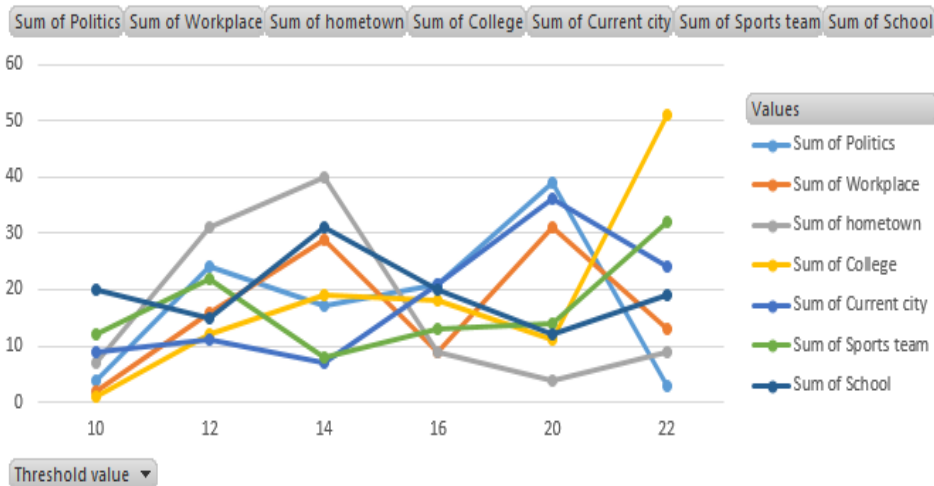


Fig. 4 number of suggestions for each community varies for different threshold value in a social network for fixed number of nodes.

These are the two results which are produced in our research and the output produced by our approach shows that overlapping communities are detected successfully.

VI. Conclusion and Future Scope

Our research is based on the detection of overlapping communities with the help of label sharing approach. We described earlier that in this approach we made a separate table which maintains the record of all the communities present in the network and also the number of nodes belonged to those communities. A number of members are suggested from each community on the behalf of the members' interest. The result produced by this approach shows that overlapping communities are successfully detected. In future, we will try to improve the time for producing the output.

References

- [1] Dr. Kulvinder Singh, Dr. Sanjeev Dhawan, Vinay, "Study of Community Detection in Social Networks using Label Sharing Approach", International Journal of Research in Information Technology, Vol. 2, Issue 4, pp. 65- 70, 2014
- [2] Ding Xiao, Nan Du, Bin Wu and Bai Wang "Community Ranking in Social Network", Second IEEE International Multi-symposium on Computer and Computational Sciences, pp. 322-329, 2007

- [3] Guilan Hu, Xiaochun Cai “*Research on Feature Optimization Method in Personal Information Mining*”, proceedings of WASE International Conference on Information Engineering, Vol. 1, pp. 656-659, 2009
- [4] Sorensen, L. “*User managed trust in Social Networking– comparing Facebook, MySpace and LinkedIn*”, proceedings of Wireless VITAE 1st International Conference, pp. 427-431, May 2009
- [5] Steve Gregory “*Finding overlapping communities in networks by label propagation*”, New Journal of Physics, Vol. 12, pp.1-26, 2010
- [6] Vinay, Sumit, Jai Parkash, “*A Review Paper for Detection of of Overlapping Communities in Complex Networks*”, International Journal of Computer Science and Information Technologies, Vol. 5 (3), pp. 3339-3341, 2014
- [7] U Kang and Christos Faloutsos “*Big Graph Mining: Algorithms and Discoveries*”, ACM SIGKDD Explorations Newsletter , Vol. 14, no. 2, pp.29-36, 2012
- [8] Lovro Subelj and Marko Bajec “*Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction*” Phys. Rev. E 83(3), 036103, pp:- 1- 13 March, 2011
- [9] S. Simranjit, M. Nikunj and M. NishantSardar Patel Institute of Technology, Mumbai, India “*A Multifaceted approach towards Friend Recommendation in Social Network*”, International Journal of Computer Science and Telecommunications Volume 3, Issue 11, November 2012
- [10] Brian Dickinson, Benjamin Valyou “*A Genetic Algorithm for Identifying Overlapping Communities in Social Networks Using an Optimized Search Space*”, Social Networking, Vol. 2, pp. 193-201, 2013
- [11] S. Thayananthporkalanchiam, Mrs. G. Umarani Srikanth, S. Rajakuncharam, Veeraraghavapuram “*MOGA-Net: A MultiObjective Genetic Algorithm-Net to Find Communities in Complex Networks*”, International Journal of Computer Science and Management Research, Vol. 2 Issue 2, pp. 1601-1606, 2013