International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Detecting Overlapping Nodes in MLM Chain Network

Sukhada Vader[1], Mugdha Kirkire[2], Rajvardhan Babar[3]

[1]Department of Computer Engineering, GHRCEM, Wagholi, Pune, India

[2]Professor, Department of Computer Engineering, GHRCEM, Wagholi, Pune, India

[3]Department of Computer Engineering, Shivaji University, Kolhapur

Abstract: ***When a particular node of a network concurrently belongs to more than a few communities, overlapping communities is form which is an open challenge related to community detection. This article evaluates the state-of-the-art in overlapping community detection algorithm in MLM Chain networks. Community discovery in MLM Chain networks is an interesting problem or task in complex networks, with the number of applications, particularly in the social and information networks knowledge extraction tasks. A community which is also referred or considered as a cluster or module is usually consisting of a group of nodes with more connections between its members than between its members and the remainder of the network. Clustering of social networks is an important task for analysis. A variety of clustering algorithms have recently been given to handle data that is not linearly separable. Customary graph partitioning algorithms are be unsuccessful to get the concealed knowledge present in modular structure appear, because they impose a top-down global view of a network. Due to the difficulties in detected communities and the limitations of scalable algorithms, the problem of overlapping community detection in large networks is remains an open problem. To avoid this problem or to get rid it; we proposed a method called Detecting Overlapping Node Structure a connection based algorithm proposed for discovering high quality overlapping structures in MLM Chain networks. The main idea of this paper is to find the seed by applying algorithm and then expanding these seeds. In our method, communities are allowed to overlap because communities are formed by adding peripheral nodes to cores. Here we focus on commonly used method commonly known as the tree edit distance. a polynomial-time algorithm is used for ordered trees compute it. Proposed method is fast, very limited parameter dependent and only requires local knowledge about the network chain. Furthermore, the community structures discovered is deterministic.***

Keywords: Overlapping, MLM Networking, Community, Network Chain

1. Introduction

Community or modular structure is considered as the important characteristics of real-world social networks and in MLM Chain networks as it frequently accounts for the functionality of the system. In many social and information networks, these communities get overlap naturally [2]. Clustering of social networks is an important task for analysis. A variety of clustering algorithms have recently been given to handle data for small scale data set. But as the dataset increase it becomes very difficult to determine communities due to the limitations of scalable algorithms for large data set.

Numerous methods have been proposed for both efficient and effective community detection. The previous technique like spectral clustering, Random walks, differential equations, modularity maximization and statistical mechanics have all been used to determine community. This type of detection methods assumes that the network is partitioned into dense regions in which nodes have more connections to each other than to the rest of the network. Therefore, existing clustering or community algorithms are not a good option for clustering large-scale social networks.

In this novel method we propose a technique which first searches a node in a given dataset and then arranges their neighbors according to them by a simple search algorithm. Then a random local search algorithm takes two nodes which are selected randomly which belong to different communities and their community ID is exchanged. The process is repeated till a maximum modularity is discover and then all links between both communities are removed. The proposed method is fast, very limited parameter dependent and only requires local knowledge about the network chain. Furthermore, the community structures discovered is deterministic.

2. Background

When a particular node of a network concurrently belongs to more than a few communities overlapping communities is form which is an open challenge related to community detection. Number of methods is used for identifying overlapping communities like CPM for identifying overlapping communities. Later on, methods like MOSES and SHRINK were Introduce to identify overlapping communities in social network which are based on label propagation. None of the mentioned methods consider the hierarchical structure of communities. However, to provide proper information about the modular structure, it is desirable to detect overlapping communities along with their hierarchical organization [12].

[13] Present a method which allowing overlapping community Detection based on the principle of edge between's introduced by Newman and Girvan. This is more suitable for social media, called Overlapping Community Detection Algorithm (OCDA). OCDA allows nodes to belong to more than one community, which is accordingly more appropriate and valid for the study of social networks. This method first identifying all possible central nodes, then

trying to expand communities and optimize this partition in the last phase. A simplify and expand method on relations between vector based and graph-based clustering to offer a unifying mathematical connection between kernel k-means and graph clustering objectives. [3] In exacting, a weighted form of the kernel k-means purpose is equal to the mathematically equivalent to a universal weighted graph collect objective. A kernel k-means algorithm is used to calculate the clustering in nodes and it can also be used where eigenvector computation is not possible. Kernel k-means is more desirable than spectral methods

Earlier to avoid the overlapping community detection, [14] proposed an algorithm using a seed expansion technique. The key idea this algorithm is that to find good seeds, and then avariciously develop these seeds based on a community metric. Within this seed expansion technique, they also investigate to overcome from the problem of how to determine good seed nodes in a graph. In short, they expand new seeding strategy for a modified PageRank clustering system that optimizes the conductance community achieve.

A technique based on the approach of matrices, moments, and quadrature developed in the numerical linear algebra community for approximating the commute time and Katz score between a pair of nodes. This approach is mainly based on the Lanczos process provide with upper and lower bounds on an estimate of the pair wise scores. In this process to approximate the commute time a conjugate gradient algorithm is used or for Katz scores a technique based on exploiting an empirical localization property of the Katz matrix is used. This method also used algorithms for personalized PageRank to compute Katz scores [6].

A method called Overlapping Community Detection by Collective Friendship Group Inference is proposed in [9], which is based on collective viewpoint of individuals for group detection that finds communities. The main idea of this method is that by the way of its ego net, each node in the network knows who is in its friendship groups. Therefore, by collectively each individual's views of friendship groups, communities can be exposed. In this method they used the term "friendship group." to represent the small clusters, extracted from ego nets, which containing the central node and its connected neighbors.

3. Proposed Methodology

To community structure improve the quality and optimizes modularity factor a random local search agent is used. We introduce our overlapping community detection algorithm, NISE which consists of four phases: filtering, seeding, seed expansion, and propagation. In the filtering phase, we remove regions of the graph that are trivially separable from the rest of the graph. In the seeding phase, we find good seeds in the filtered graph, and in the seed expansion phase, we expand the seeds using a personalized PageRank clustering scheme. Finally, in the propagation phase, we further expand the communities to the regions that were removed in the filtering phase.

1) Filtering phase:
Is this process used to identify regions of graph to overlapping algorithmic solutions, for that graph partitioning method can be begin for connected component for separate partitioning. Combine two disconnected components into a single partition in order to satisfy a balance constraint on the partitioning. To separate partitioning has convert into another portioning which remove filter path or node.

1) Seeding phase
We are find communities which have different types of graph identification. In seed phase we get biconnected and seed in filtered graph. cited core graph, we find seeds in this filtered graph. The goal of an effective seeding strategy is to identify a diversity of vertices, each of which lies within a cluster of good conductance. This identification should not be too computationally expensive.

a) Graclus Centers:
It has apply fast partitioning and high quality graph. In graph set has no of good, bad seeds are available. This graph has find out different types of clustering communities. To find out each and every vertex and communities distance using adjacent graph. Finally we compute minimum distance set which has stored good seeds.

b) **Spread Hubs**:
Which has independent vertex for decreasing node degree. In hubs find minimal distance node and cluster which inverse propositional to node degree. In good cluster find out high vertex small distance to other vertex.

2) Seed Expansion Phase
In this phase we calculate Pagerank, accuracy for expand cluster around those seeds. To achieve used to Personalized PageRank(PPR) called as random walk. a particular algorithm to compute a personalized PageRank vector, followed by a sweep over all cuts induced by the vector, will recognize a set of good conductance inside the graph. We can use multiple node in graph and neighborhood entire seed node as restart node.

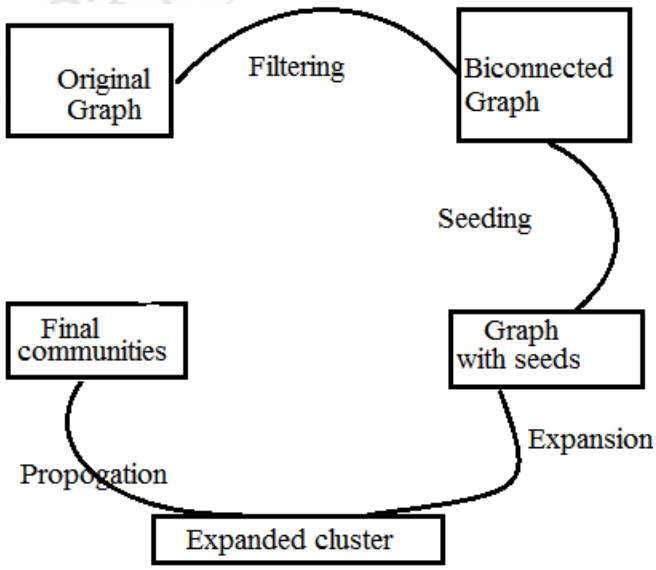

Figure 1: System flow

The above fig shows the system flows. Which have using to filtering, seeding, Expansion and propagation module find out final communities?

4. Algorithm and Mathematical Evolution

1. Let Vertex and edges are set of graphs G
$$G = (V, E)$$
2. Apply filtering for calculate biconnected Graph BG.
$$BG(V, E) = \int_{i=1}^{node} Filtering(G(V_i, E_i))$$
3. Combine two disconnected component in to single partitioning in filtering.
4. Find every vertex and communities distance using adjacent graph
$$Graclus = \min_{i=0} Dist(V, C_i)$$
Where V=Vertex,
C_i=Communities.
5. Find minimal distance node and cluster which inverse propositional to node degree in Spread Hubs.
$$spread = \min_{i=0} \left(\frac{1}{Dist(G)}\right)$$
6. Find graph identification as seeding in step 4 and 5
$$Seeding = Graclus + spread$$
7. Find final cummunities
$$communities = PPR(seeding)$$
Where, PPR= personalized PageRank vector.

5. Dataset Evolution

In this system we have used to different social networking, product networking datasets. Which has include Flicker, Live Journal , MySpace. This datasets has no of vertex and edges are available.

1) Flicker Datasets: A network data set crawled from Flicker. Both the contact network and selected group association information are integrated. It has included 80513 Nodes and 5899882 edges are connected to each other. It has include 4 files node.csv, edges.csv, group.csv, group-edges.csv.
2) Orkut: Orkut is a free on-line social network where customer form friendship every one. Orkut also allows consumer form a group which other associate can then join. We consider such user-defined collection as ground-truth society. We provide the Orkut companionship social network and ground-truth society. It has 3072441 Node and 117185083 Edges are available.

6. Software Requirement Specification

The software requirement section includes all the information to run the project Overlapping Social Network Detection for MLM Chain Network. The document describes the issues related to the system and what actions are to be performed by the development team in order to come up with a better solution. A minimum Dual Core of 2.2 GHZ processor with 2 GB ram and 100GB hard disk is require to run the project comfortably. For this project we have used platform C# with technology ASP using C#, for that we have used Visual Studio 2010 and database SQL Server

7. Result Analysis

Graph		Oslom	demon	bigclam	nise
Flickr	coverage (%)	N/A	N/A	50.11	91.70
	no. of clusters	N/A	N/A	14,600	16,047
Amazon	coverage (%)	98.03	77.18	98.3	98.3
	no. of clusters	16,882	105,000	24,890	27,582
Orkut	coverage (%)	N/A	N/A	80.43	100
	no. of clusters	N/A	N/A	24,795	25,228

We contrast our NISE method with the other society discovery algorithm such as demon, bigclam and oslom by using real time dataset like Orkut , Amazon and Flicker. First we account the figure of group and the graph contact of each method as shown in table 2. The graph exposure shows the number of vertices allocate to clusters. As we can manage the number of seeds in NISE and number of *k* i.e. number of cluster in bigclam we have set 14,600, 24,890 for Amazon and for orkut it is 24,795. As we remove all the clustering using NISE algorithm it shows that the return numbers are slightly smaller than the number of *k*. here we use top down hierarchy so the number of cluster before filtering is greater than or equal to $2^{[\log k]}$. The technique oslom and demon evaluate the number of cluster based on the dataset themselves. Although these methods are fail to evaluate for Flicker and Orkut.

8. Conclusion

The advantage of our technique is that it is quicker than other technique which is best in MLM Chain networks. Maybe shockingly, the major dissimilarity in cost between utilizing "graclus focuses" for the seeds and the other seed decisions does not come about because of the cost of running Graclus. Or maybe, it combines on the grounds that the customized PageRank addition system takes more time for the seeds picked by Graclus. At the point when the PageRank extension method has a bigger info set, it be inclined to take longer, and the "graclus focuses" seeding methodology produce larger input sets because of the neighborhood inflation and because the central vertices of clusters are likely to be high degree vertices. To address the relationship between our results and some prior observations on overlapping communities.

References

[1] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: the state of the art and comparative study," ACM Computing Surveys, vol. 45, no. 4, 2013.
[2] J. J. Whang, X. Sui, and I. S. Dhillon, "Scalable and memoryefficient clustering of large-scale social networks," in Proceedings of the 12th IEEE International Conference on Data Mining, 2012, pp. 705–714.
[3] S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors: A multilevel approach," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 11, pp. 1944–1957, 2007.

[4] H. H. Song, B. Savas, T. W. Cho, V. Dave, Z. Lu, I. S. Dhillon, Y. Zhang, and L. Qiu, "Clustered embedding of massive social networks," ACM SIGMETRICS Performance Evaluation Review, vol. 40, no. 1, pp. 331–342, 2012.
[5] B. Abrahao, S. Soundarajan, J. Hopcroft, and R. Kleinberg, "On the separability of structural classes of communities," in Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data mining, 2012, pp. 624–632
[6] F. Bonchi, P. Esfandiar, D. F. Gleich, C. Greif, and L. V. Lakshmanan, "Fast matrix computations for pair wise and column wise commute times and Katz scores," Internet Mathematics, vol. 8, no. 1-2, pp. 73–112, 2012.
I. M. Kloumann and J. M. Kleinberg, "Community membership identification from small seed sets," in Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data mining, 2014, pp. 1366–1375.
[7] J. Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in Proceedings of the 6th ACM International Conference on Web Search and Data Mining, 2013, pp. 587–596.
[8] B. S. Rees and K. B. Gallagher, "Overlapping community detection by collective friendship group inference," in International Conference on Advances in Social Networks Analysis and Mining, 2010, pp. 375–379.
[9] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, "Demon: A local-first discovery method for overlapping communities," in Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data mining, 2012, pp. 615–623.
[10] Sanghamitra Bandyopadhyay, Garisha Chowdhary, and Debarka Sengupta "FOCS: Fast Overlapped Community Search" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 11, NOVEMBER 2015.
[11] Sajid Yousuf Bhat and Muhammad Abulaish "HOCTracker: Tracking the Evolution of Hierarchical and Overlapping Communities in Dynamic Social Networks",IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 4, APRIL 2015.
[12] Feyza Altunbey, Bilal Alatas "Overlapping Community Detection in Social Networks Using Parliamentary Optimization Algorithm" International Journal of Computer Networks and Applications Volume 2, Issue 1, January - February (2015).
[13] Joyce Jiyoung Whang David F. Gleich, and Inderjit S. Dhillon "Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING
[14] Qinna Wang and Eric Fleury "Overlapping Community Structure and Modular Overlaps in Complex Networks" © Springer Science+Business Media Dordrecht 2013.