ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 10, October 2013

Multi-layered Graph Clustering in Finding the Community Cores

Zin Mar Yin, Soe Soe Khaing
Faculty of Information, Communication and Technology,
University of Technology, Yatanarpon Cyber City, Myanmar.

Abstract

In the study of complex network, community mining has become popular in recent years. However, almost all existing community detection algorithms are based on a single-layered homogeneous type of graph in which nodes represents individual objects and edges for a single definition of a relationship. In reality, an object has the existence of more than one type of tie. In this paper, we propose a simple idea to find the dense cluster nodes or community cores in multi-layered graph. Our method treats a heterogeneous complex network as a set of homogeneous graphs in ordered to apply the modularity-based single-layered graph clustering method. Then, the community cores are extracted by layered-based intersection approach. With the help of powerful graph exploration and manipulation tool, Gephi, we also presented our experiments in computer-generated networks with graph visualizations.

Keywords: community mining, community cores, graph clustering, heterogeneous network, homogeneous graph.

I. INTRODUCTION

The most effective representation of the interconnections of real world objects is the graph-based network. For example, in the World Wide Web, node represents a single page and edge as a hyperlink. Others are social networks, food webs, metabolic structures, etc. The mathematical formulation and topological investigation of the complex network have a significant benefit to reveal some unexpected relationships of a particular object or a group of them. As crime detection, the chain of offences can be easily detected from a closed cluster of connected persons.

Graph mining methods naturally divides the network of interest into sub-graphs or sub-clusters and the analysis point of view is to investigate those groups for intended purposes. To model this problem, first the structure of a graph is built in accordance with the data acquisition of the problem domain. Then, an appropriate graph partition algorithm is applied to identify the communities. Most of the existing algorithms on network clustering are totally based on a single tie (i.e., one edge definition) to find the subgroup of objects in the homogenous network. However, in real world networks, an object has the existence of more than one type of relationship for its interconnection with other objects. For example, in a student social network, a node becomes an individual student and the relationships between the students become a set of edges, such as same interest of study, interest and hobby, native city or country, religions and culture, etc. Each relation can be configured as a single relation network and this category of social network is defined as multi-relational social network or heterogeneous social network [1]. There is a common sense that a particular tie has a definite interpretation and influence in the network. The importance of each tie is mainly depended on the problem domain of experimenter's interest. Thus, the choice of which relationships are required to investigate should be open to question in this type of data analysis.

More important to consider is how to cluster the subgraph for each types of attributes in the given areas of interest and how to merge the results of them in order to perform analysis effectively. The basic idea of this paper is to find the dense community cores in heterogeneous graph via a simple combination of a series of resulted sub-clusters from each layer of homogenous graph structure.

The paper is organised as follows. In Section 2, a review of the relevant literature in modularity maximization and community studying in homogenous as well as heterogeneous networks are presented. Then, in Section 3, the briefly explanation of graph mining theories and modularity clustering method are also discussed. We introduce the consideration of the multi-layered community mining in Section 4. Our experimental findings with the computer-generated networks are presented in Section 5. Finally, Section 6 includes concluding remarks and future research avenues.

II. RELATED WORK

In the area of heterogeneous graph clustering, the extraction of the most relevant attributes by applying relational mining techniques. And then, a combined relation for investigation the closely connected cluster is presented in [1][2]. Rodriguez and Shinavier studied a path algebra mapping multi-relational networks to single-relational networks and then exposing single-relational network analysis algorithms to find the communities [3].

As a document clustering technique, the concept of community detection in single-edge graphs to multi-edged graphs is generalized by using their variation of information metric in [4]. RMiner algorithm, introduced by Spyropoulou and Bie [5], mines complete connected subgraphs (i.e., clique percolation) in a representation of the database as a K-partite graph and produces the multi-relational patterns. Analyzing the Brazilian scientific community from four different relationship types is constructed by Ströele et al. [6].

In multi-relational network, individual objects are sharing different types of information based on the linkage between them. The importance of an attribute in edge definition is depended on the personal choice. Some consider the attributes x and y is more valuable while some be z. In this paper, we simple investigate each edge type shaping to each layer random graph. Then, separate communities are identified by the Louvain method [7] and extract the dense interconnections from each layer clusters.

III. GRAPH CLUSTERING BY MODULARITY MAXIMIZATION

In the study of complex networks, nodes represent individual objects and edges represent relationships between each pair of objects. A common property of these networks is their community structure which reveals the existence of densely connected groups of vertices, with only sparser connections between groups. This type of graph clustering focuses on the detection and characterization of such network structure. And, this area has received considerable attention over the past few years in sociology and lately data mining as the trend of community detection.

Modularity is a measure of the structure of complex network or graphs. It is done by the division of an intended network into clusters or communities. The value of the modularity is between [-1, 1]. Network with high modularity values means that there exist dense connections between nodes within clusters, but sparse connections between nodes in different modules.

Modularity can also be used as an optimization method for detecting community structure in networks, and it is widely used and well-known as *modularity maximization*. This method discovers the communities by searching over possible divisions of a network for one or more that have particular high modularity. The algorithms are based on approximate optimization methods such as greedy algorithms, simulated annealing, or spectral optimization. These different approaches offer different accuracy and processing speed in their application areas.

For identifying high modularity sub-graphs of large network in short time, Blondel et. al. [7] introduced *the Louvain method* which is a greedy optimization and reveals hierarchies of communities. This method is performed in two steps: looking for small communities by optimizing modularity locally and then, integrating vertices of the same communities and constructing a new network whose nodes are the communities. These steps are repeated iteratively until a maximum of modularity is attained. Although the exact modularity maximization is NP-hard, this method have the time complexity of $O(N \log N)$ where N is the number of nodes in large network.

IV. SYSTEM OVERVIEW

Community mining becomes one of the major trends in graph clustering areas. However, almost all existing community detection algorithms are based on single-layered homogeneous type of graph which means edges represent a definition of only one relationship. In reality, an object may have the existence of more than one meaning of connection and each kind of relationship may play a distinct role in a particular task. Therefore, the proposed system has the notion of heterogeneous complex network as a set of homogeneous type of graph.

Then, each layer network is partitioned by using the Louvain Method and the system produces a series of sub-graphs or sub-clusters.

Finally, for the consideration of heterogeneous network, Cai. et. al, [1] introduced that there are different types of relationships in real-world network and each relation can be treated as a single relational network. Such kind of complex network can be called multi-relational network or heterogeneous network.

In order to find the closely related community cores in heterogeneous type of network, this paper proposed an efficient notion for community core generation. The time complexity of our algorithm has two unfolds: $O(N \log N)$ for the Louvain method, where N is the number of nodes in a network to be clustered by modularization, and $O(n^2)$ for the intersection-based core generation, where n is the number of sub-clusters in each homogeneous graphs as the worse-case. In comparison with other community core extracting algorithms [8], the advantages of our algorithm is that it does not depend on the increasing number of nodes in the intended graph. Moreover, our algorithm can be effectively applied in heterogeneous types of graph with less time-consuming with the help of the Louvain method, which is a homogeneous graph clustering technique, as shown in Fig.1.

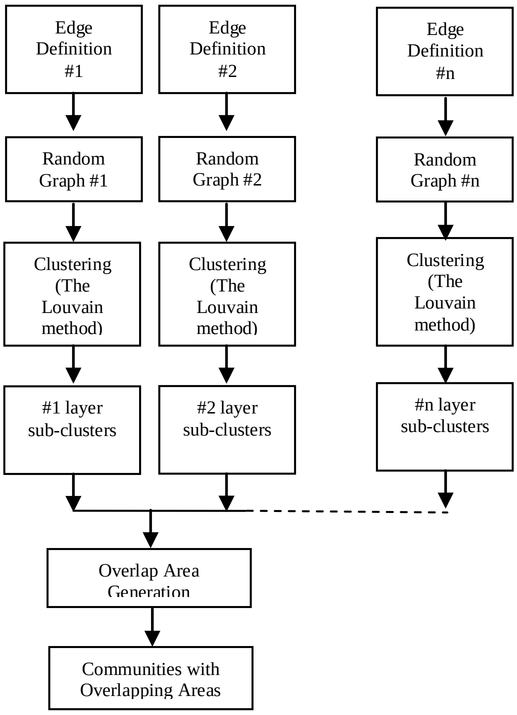

Figure 1: The overview of multi-layered community finding

V. EXPERIMENT WITH SYNTHETIC NETWORKS

To perform the experiment for our proposed system, we generate the three-layered synthetic network with 25 nodes (the label: A to Y). The number edges for each layered is 74, 105 and 121 respectively.

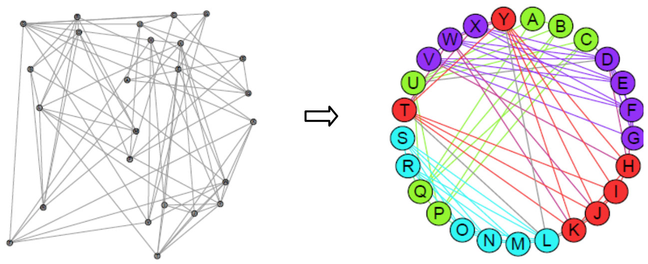

Figure 2: Clustering First-layered Graph

In Fig. 2, the sub-clusters from the first-layered network are mentioned by color separation with modularity values 0.513 at standard resolution limit of the Louvain method 1.0. These sub-clusters are C_{11} = {A, B, C, P, Q, U}, C_{12}= {D, E, F, G, V, W, X}, C_{13} = {H, I, J, K, T, Y} and C_{14}= {L, M, N, O, R, S} respectively.

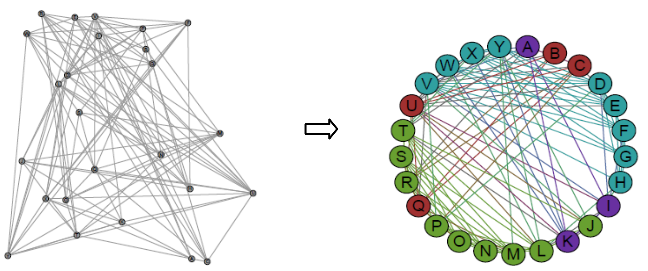

Figure 3: Clustering Second-layered Graph

In Fig. 3, the sub-clusters from the second-layered network are mentioned by color separation with modularity values 0.349 at standard resolution limit of the Louvain method 1.0. These sub-clusters are C_{21} = {A, I, K}, C_{22}= {B, C, Q, U}, C_{23} = {D, E, F, G, H, V, W, X, Y} and C_{24}= {J, L, M, N, O, P, R, S, T} respectively.

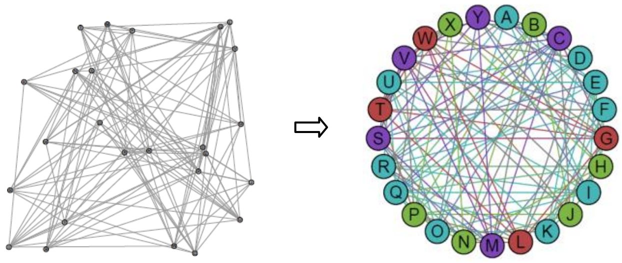

Figure 4: Clustering Third-layered Graph

In Fig.4, the sub-clusters from the third-layered network are mentioned by color separation with modularity values 0.278 at standard resolution limit of the Louvain method 1.0. These sub-clusters are C_{31} = {A, D, E, F, I, K, O, Q, R, U}, C_{32}= {B, H, J, N, P, X}, C_{33} = {C, M, S, V, Y} and C_{34}= {G, L, T, W} respectively.

In the resulted sub-clusters, there are two distinguished form of vertices: (1) some are in same sub-clusters for some type of edge definition, but not all (E.g. A B C) and (2) some are in same sub-clusters for all layers of graph (E.g. D E F, G W), this means that these vertices are *densely connected community cores* in multi-layered consideration.

In order to get these community cores, we apply the following approach which is mainly based on the layered-by-layered intersection of the sub-clusters.

For the case of a sub-cluster, as the first running, this sub-cluster is performed the intersection of all the sub-clusters from second-layered and third-layered. Then, individual nodes and null values are discarded for the absent of edge relationships. This process is repeated iteratively for all the sub-clusters. Finally, we get the vertices which are *densely connected community cores* in all layers of the heterogeneous network. The example of extracting two community cores from all three layers is shown in Fig. 5.

$C_{12} \cap C_{21} = \emptyset$
$C_{12} \cap C_{22} = \emptyset$
$C_{12} \cap C_{23} = D\ E\ F\ G\ V\ W\ X$
$C_{12} \cap C_{24} = \emptyset$
$C_{12} \cap C_{31} = D\ E\ F$
$C_{12} \cap C_{32} = X$
$C_{12} \cap C_{33} = V$
$C_{12} \cap C_{34} = G\ W$
$(C_{12} \cap C_{23}) \cap (C_{12} \cap C_{31}) = \mathbf{D\ E\ F}$
$(C_{12} \cap C_{23}) \cap (C_{12} \cap C_{34}) = \mathbf{G\ W}$

Figure 5: Finding the community cores in all layers from the sub-cluster C_{12}

VI. CONCLUSION AND FUTURE WORKS

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 experimenters' needs or preferences. However, in reality, there exist multiple, heterogeneous networks, each representing a particular kind of relationship, and each kind of relationship may play a distinct role in a particular task. Based on the observation, different relations have different importance with respect to a certain requirement.

In this paper, we propose a simple method for investigating the combination of the relation in each layer which can best meet the experimenters' own perspectives of the important of a particular relationship. With the overlap area generation, better cohesive structures can be achieved for multi-layered community mining.

REFERENCES

[1] Cai, D., Shao, Z., He, X., Yan, X., Han, J.: *Mining hidden community in heterogeneous social networks*. Technical report, Computer Science Department, UIU (UIUCDCS-R-2005-2538, May, 2005)

[2] Cai, D., Shao, Z., He, X., Yan, X., Han, J.: *Community Mining from Multi-relational Networks*. In PKDD, 2005.

[3] Rodrigueza, A. M., Shinavierb, J., *Exposing Multi-Relational Networks to Single-Relational Network Analysis Algorithms.* Journal of Informetrics, vol. 4, no. 1, 29-41, ISSN: 1751-1577, Elsevier, doi:10.1016/j.joi.2009.06.004, LA-UR-08-03931, December 2009.

[4] Rocklin, M., Pinar, A., On clustering on graphs with Multiple edge types

[5] Spyropoulou, E., Bie, D. T., Interesting Multi-Relational Patterns

[6] Ströele, V., Zimbrao, G., Souza, M. J., Modeling, mining and analysis of multi-relational scientific social network. Journal of Universal Computer Science, vol.18, no. 8(2012), 1048-1068

[7] Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast Unfolding of Community Hierarchies in large network, 2008, J. Stat. Mech. P1008.

[8] Gonen, M., Ron, D., Weinsberg, U., Wool, A.,"Finding a Dense-Core in Jellyfish graphs", Computer Networks 52(15):2831–2841, 2008.