Fast Algorithm for OVERLAPPING Community Detection

R.Sujatha
Department Of Computer Science
National Engineering College,Kovilpatti
Thootukudi,India

Ms.S.Indumathy M.E.,
Department Of Computer Science
National Engineering College,Kovilpatti
Thootukudi,India

Abstract

Network is a collection of entities that are interconnected with links. The widespread use of social media applications like youtube, flicker, facebook is responsible for evolution of more complex networks. Online social networks like facebook and twitter are very large and dynamic complex networks. Community is a group of nodes that are more densely connected as compared to nodes outside the community. Within the community nodes are more likely to be connected but less likely to be connected with nodes of other communities. Community detection in such networks is one of the most challenging tasks. Community structures provide solutions to many real world problems.

Nowadays, the emergence of online social networks has empowered people to easily share information and media with friends. Interacting users of social networks with similar users and their friends form community structures of networks. Uncovering communities of the online users in social networks plays an important role in network analysis with many applications such as finding a set of expert users, finding a set of users with common activities, finding a set of similar people for marketing goals, to mention a few. Although, several algorithms for disjoint community detection have been presented in the literature, online users simultaneously interact with their friends having different interests. Also users are able to join more than one group at the same time which leads to the formation of overlapping communities.

Thus, finding overlapping communities can realize a realistic analysis of networks. A fast algorithm is proposed for overlapping community detection. In the proposed algorithm, in the first phase, the Louvain method is applied to the given network and in the second phase a belonging matrix is updated where an each element of belonging matrix determines how much a node belongs to a community. Finally, some of the found communities are merged based on the modularity measure. The performance of the proposed algorithm is studied through the simulation on the popular networks which indicates that the proposed algorithm outperforms several well-known overlapping community detection algorithms.

I Introduction

Network is a collection of entities called nodes or vertices which are connected through edges or links. Computers that are connected, web pages that link to each other, group of friends are basic examples of network. Complex network is a group of interacting entities with some non trivial dynamical behavior. There are many types of complex networks for example Social Networks, Technological Networks, Informational Networks, and Biological Networks etc. Social networking is an application that allows users to interact with each other.

The study of complex networks is a youthful and vigorous area of scientific research stimulated largely by the pragmatic study of real-world networks such as computer networks and social networks. Complex network paradigm is one of the modeling tools which have spread through several application fields such as sociology, communication, computer science, biology, and physics and so on during last decades. Community

structure Community is a collection of nodes that are more densely connected as compared to the nodes outside the community. Nodes inside the community have some common properties. There are basically two categories of communities: disjoint community and overlapping community.

The disjoint community refers to crisp arrangement of nodes where node belongs to single community. The overlapping communities are known as fuzzy assignment of nodes where a node may overlap between two or more communities. Community detection is the important aspect of the complex network study. Communities may relate to group of pages of World Wide Web with related topics and to functional modules in metabolic networks.

Community detection: Detection of such community structures in complex network is not an easy task. There are various community detection algorithms available which detects communities in the network. Some of them are Hierarchical clustering algorithm, Divisive algorithm, Kernighan-Lin (KL) Graph Partitioning approach, Multilevel Graph Partitioning. These algorithms detect disjoint communities only. For overlapping community detection other algorithms like Clique Percolation Method (CPM), Eagle algorithm, Cluster Overlap Newman Girvan Algorithm (CONGA), Link Clustering, Label propagation algorithms etc are used.

II Related Works

Community Detection

Detection of such community structures in complex network is not an easy task. There are various community detection algorithms available which detects communities in the network. Some of them are Hierarchical clustering algorithm, Divisive algorithm, Kernighan-Lin (KL) Graph Partitioning approach, Multilevel Graph Partitioning. These algorithms detect disjoint communities only. For overlapping community detection other algorithms like Clique Percolation Method (CPM), Eagle algorithm, Cluster Overlap Newman Girvan Algorithm (CONGA), Link Clustering, Label propagation algorithms etc are used.Social network is a network graph holding nodes with edges connected to it. Community detection aims to identify the community in the graph. Community is a module containing the set of nodes with major activities/interaction/similaritythem.Overlapping communities are possible if any of the nodes has participated in the formation of different modules. Each node's strength or membership value in different modules varies accordingly. Various metrics are utilized to measure the strength of the community. Most popular of it is modularity measure. This modularity is the strength of partition of network as a community.

Existing System

Community detection is important concept in network .To implements various algorithms to identify the nodes and edges. For example Healthcare fraud detection with community detection, Parallel algorithm for community detection with massive network, Redesign coordination algorithm for cyber physical system these three papers use various algorithms that is cluster, parallelism, redesign of coordinate algorithms. These Algorithms are used to identify community in large network but provide result is very slow because calculation of edge between its very difficult.

Disadvantage

In community algorithm,eventhough it yields good results but is very slow because of the computational complexity of edge between calculation.scores has to be re-calculated after every edge removal.

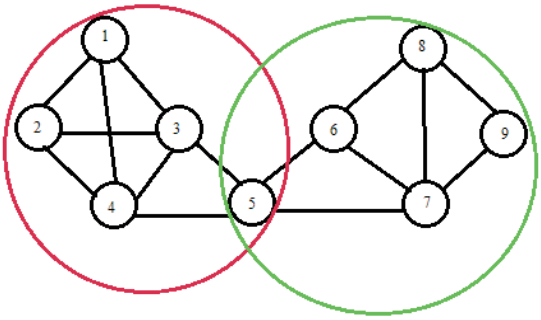
Proposed System

In the first phase, the **Louvain method** is applied to the given network and in the second phase a **belonging matrix is updated** where an each element of belonging matrix determines how much a node belongs to a community. Finally, some of the found communities are merged based on the modularity measure. The performance of the proposed algorithm is studied through the simulation on the popular networks which indicates that the algorithm outperforms several well-known **overlapping community detection algorithms**.

Advantage

Louvain method is applied to detect similer behavior community in large social network.Easy to calculate edge between calculation.

It is today one of the most widely used method for detecting communiies in large network.



Example of Overlapping Community Graph Structure.

Community Detection Algorithm in Social Network

Louvain method

Louvain is greedy optimization that attempts to optimize the modularity of a partition.

The method looks for small community by optimizing modularity locally.

It aggregates nodes belonging to the same community and builds a new networks whose nodes are the communities.

These steps are repeated interactively until a maximum of modularity is attained and a hierarchy of community is produced.

Modularity Optimization

Modularity is one measure of the structure of networks or graphs.It was designed to measure the strength of a network into modules(also called groups,cluster or communities).

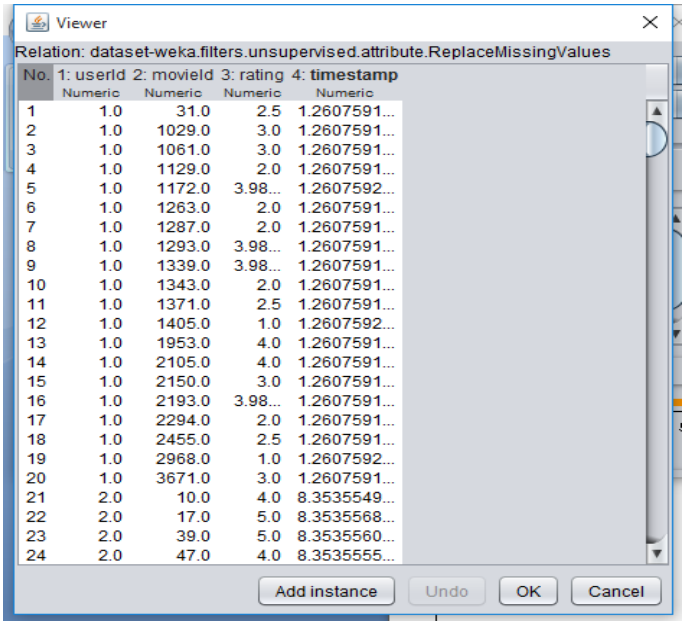
Community detection is the optimization of modularity.Modularity is a scale value between -1 and 1 that measures the density of edges inside communities to edges outside communities.Optimize this value result in the best possible grouping of the nodes of a given network.

Louvain method of community detection first small communities are found by optimizing modularity locally on all nodes.

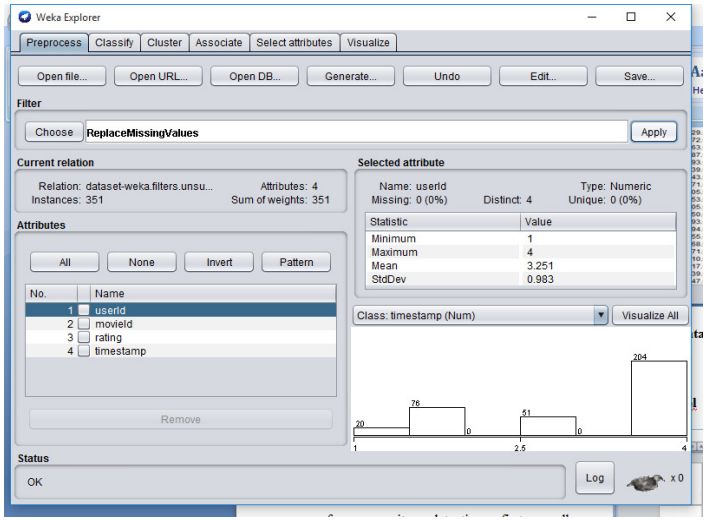
Each small communities grouped in to one node and first step is repeated.

Experimental Result

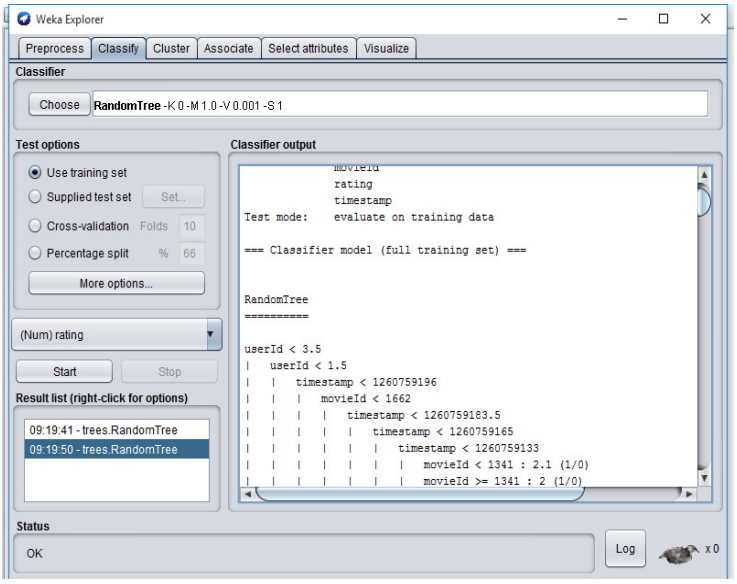
Social media dataset

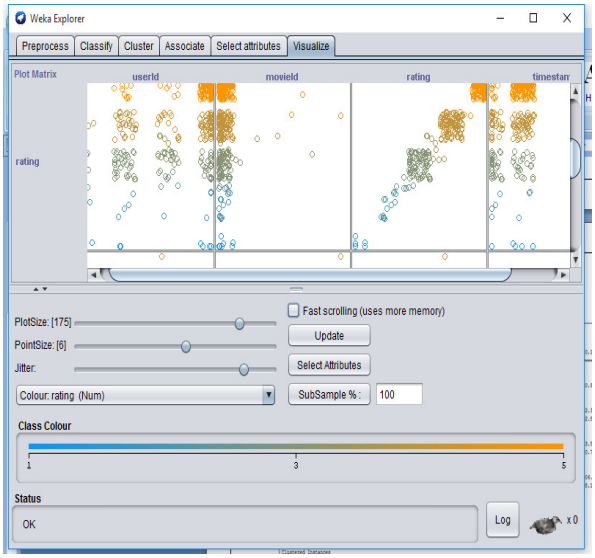


Preprocess dataset

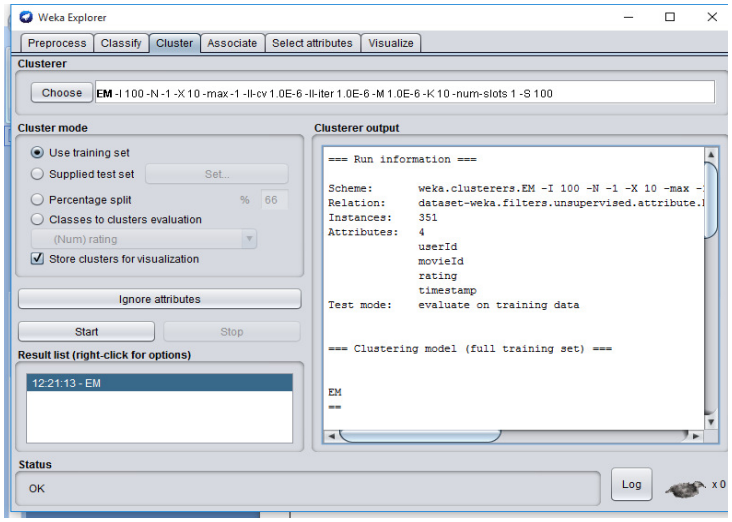


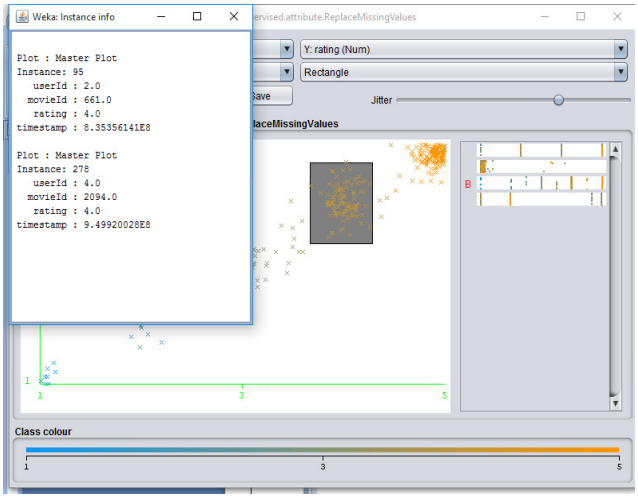
Classification



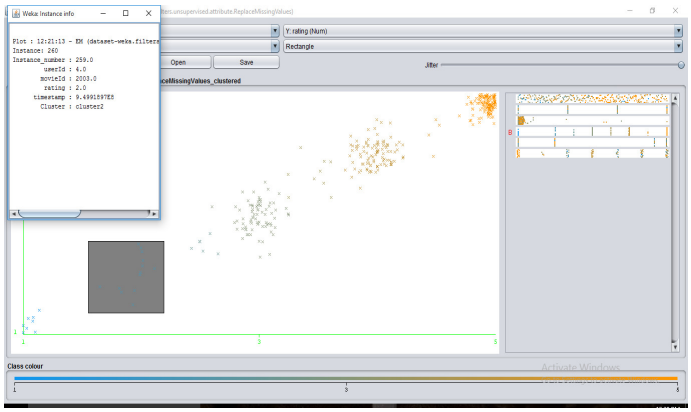


Cluster





Visualize cluster



Conclusion

In this work, several state-of-the-art modularity based community detection algorithms overlapped communities are analyzed.
Quantitative analysis is performed with the modularity score to infer the best available method.

Reference

[1] M E J Newman, "Detecting Community Structure in Networks", *Eur. Phys. J. B 38*, pp 321-330 2004.

[2] M.E.J. Newman and M Girman, "Finding and evaluation community structure in networks", *Physical Review E, 69(2),* 2004.

[3] L. Zhubing, W. Jian, and Li yuzhou, "An Overview on Overlapping Community Detection", *The 7th International Conference on Computer Science & Education (ICCSE 2012)* , Melbourne, Australia , July 14-17, 2012.

[4] M. Girvan and M. Newman, "Community Structure in Social and Biological Networks", *Proceedings of the National Academy of Scinces*, vol. 99, no. 12, pp. 7821– 7826, Jun. 2002.

[5] L.C. Huang, T.J Yen, and S.C.T. Chou, "Community Detection in Dynamic Social Networks".

[6] G. Kangl, M. Diaz, T. Perennou, P. Senac, and L. Xul, "Mobility Model Based on Social Community Detection Scheme", *2011 Cross Strait* Quad-Regional Radio Science and Wireless Technology Conference.

[7]Health Care Fraud Detection with Community Detection Algorithm
Aryya Gangopadhyay,Song Chen
Smart Computing(SMARTCOMP),2016 IEEE.

[8] Engineering Parallel Algorithms for Community Detection in Massive Networks.
Christian L.Staudt,Henning Meyerhenke
IEEE Transactions on Parallel and
distributed Systems(Volume:27,Issue:1,jan. 1 2016).