Community Detection for Social Network Analysis

ABSTRACT

of the **Ph.D.** Thesis

Submitted to

Jamia Millia Islamia

for the award of the Degree of Doctor of Philosophy

Submitted by

Seema Rani

under the supervision of

Dr. Monica Mehrotra

Department of Computer Science

Jamia Millia Islamia

Department of Computer Science

Faculty of Natural Sciences

Jamia Millia Islamia

New Delhi

April, 2019

Community Detection for Social Network Analysis

Abstract

Speedy growth of communication on social media as a result of stress-free and inexpensive means of communication has persuaded to hook up with people across the globe. The communication among people irrespective of place and time leads to the unprecedented growth of social networks. The availability and accessibility of wide-ranging networks have opened up many research avenues for researchers' from multidiscipline. In the present time, the presence of networks in multidiscipline and the demand for their analysis have taken the social network analysis research to astounding scales.

One endeavor of immense significance in social networks is "Community Detection". A community in real-world networks plays a persuasive role in understanding the functionality and organization of complex systems. In real-world networks, members communicate based on likes or dislikes, common or disparate interest etc. Accordingly, they share their membership with one or more groups in the network. Based on the high density of member within a subgroup, the network is partitioned into subgroups. Thus, the community detection problem can be stated as "identifying highly cohesive connected subgroups in the networks".

The highly cohesive connected subgroups in multidiscipline facilitate to solve many problems like identifying crucial members from terrorist networks, viral marketing, recommender system, biological networks, transportation networks, identifying key plugs

in case of natural disaster etc. The researchers are paying continuous attention in the direction of discovering the community structure from complex networks. The current trend directs in the way of finding more efficient and effective methods. This endeavour of social network analysis expedites to simplify the real-life problems.

Community detection problem is witnessed from traditional and optimization problem viewpoint. The communities present in networks often have an obvious interpretation. They accelerate the understanding of the functional organization of the networks. More recently, the focus is on introducing effective and improved community detection algorithms. Primarily, the output of the community detection algorithm for network partitions is evaluated by extensively adopted measure, modularity. In social network analysis, finding network partitions with maximum modularity is an NP-hard problem. Thus, the problem is articulated as an optimization problem. Evolutionary algorithms to deal with optimization problem are one among the prevalent choices. In the last decade, many new nature-inspired algorithms have been introduced and applied for solving different optimization problems.

The rapid growth of networks around the globe and applicability of community detection algorithms by a wide range of disciplines motivated to carry out a study based on enrichments and applications in the last decade. In this thesis, the study falls into two categories with the intent of finding the improved quality of community structure by enhancing the existing community detection method and present variants of a new nature-inspired evolutionary algorithm (BAT).

A centrality named as hybrid influential centrality is introduced for improving the quality of communities. The proposed hybrid influential centrality based Label Propagation Algorithm (LPA) redefines the method of node selection and label selection phenomenon instead of total randomization. The experimental outcomes have been compared with prevailing approaches on the real-world networks under test. The results are improved in terms of increase in modularity value. In the category of evolutionary algorithms, a new variant of discrete bat algorithm (NVDBA) is introduced for finding the communities from networks. The experimental results of NVDBA are obtained for small scale to moderately large size networks. Further, the box-plot analysis and statistical testing are conducted to convey the significant improvement by NVDBA over the well-known algorithms namely LPA and DPSO.

Further, a hybrid bat algorithm is proposed for uncovering the communities from real-world networks. Here, tabu search is embedded for boosting the local search process in the bat algorithm. The proposed hybrid discrete bat algorithm named as DBA-tabu shows better convergence and escalation in modularity value. The experiments have been conducted for Newman benchmark network, LFR benchmark network, and real-world networks. The outcome of DBA-tabu conveys the rationale of hybridization. However, modularity measure is not able to expose the communities of smaller sizes as compare to the size of network i.e. it suffers from the resolution limit problem. So, modular density measure for assessing the goodness of network partitions has been adopted as an optimization function. This measure succeeds in resolving the problem posed by modularity. In continuation, a variant of Bat Algorithm based on modular density

optimization is proposed. The proposed algorithm is tested on real-world networks and compared with existing algorithms in accordance with generally applied evaluation criteria. The final outcome has shown promising and encouraging results. In continuation, box-plot analysis has been conducted to observe the consistency in experimental values during independent runs.