Community Detection in Social Networks using Improvised Evolutionary Algorithms

Abstract of PhD Thesis

As a Part of Degree of Doctor of Philosophy

In

(Computer Sciences & Engineering)

By

Harish Kumar Shakya

Under the Supervision of

Dr. Bhaskar Biswas, (Associate Professor)

Department of Computer Science & Engineering,
Indian Institute of Technology,
(Banaras Hindu University)
Varanasi- 221005

Enrolment: 341685 **January 2018**
Roll Number: 11600EN003

ABSTRACT

The concept of social network has become popular after websites like Facebook and Google+ emerged and become a part of our everyday life. The two main properties of the social networks are entities and the relationships within these entities participating in the network. Entities might be "people" and relationships might be the "friendship" of these people like on Facebook and like most of the other social websites but they are not limited to "people" and "friendship". Entities might be entirely different e.g., organizations, websites and relationships might be something else e.g., business, trade, collaboration. Relationships can be all-or-nothing as in Facebook that you are friend with someone or not, or can be a degree as in Google+. Although social networks and their analysis has been a very popular research area in sociology [60][68] for decades, recent revolution on the internet and computer applications have made huge amount of real world data available to analyze and process for researchers. Real world networks can be very large in size, even reaching billion of vertices so there is a need for changing how to handle analyzing and processing networks and a large number new methods have been produced.

In this work, an improvised algorithm, Modified Crossover Opposition Based Genetic Algorithm (MCOBGA), for community detection has been proposed. As name suggests, Genetic Algorithm has been used to discover community structure in social networks. Modified crossover and opposition based initialization have been deployed along with GA to improve the quality of community structures. Initialization of the population through Opposition based learning ensures improved selection of initial population, whereas modified crossover transmits information for improved community structures. The evaluations of proposed algorithm have been done on real-world networks and experimental results show that MCOBGA has very competitive performance compared with GA with vertex similarity applied to community detection, which have been most similar approach to the proposed algorithm. Experimental results not only demonstrates improvement on convergence rate of the algorithm, but also communities discovered by proposed algorithm (MCOBGA) is highly inclined towards quality, compared to its counterpart.

Community detection has become a significant research track for data mining in social networks. Evolutionary algorithm-based approaches, along with many other existing community detection techniques are extensively utilized. However EA-based methodologies are inclined to populace corruption and local convergence. Developing more proficient evolutionary algorithms thus becomes essential. This proposes a community detection algorithm based on the genetic algorithm using the regenerative technique. In this algorithm, the Community detection process is done by regeneration of the population and selection. The similarity of vertexes is defined as local fitness function; the community quality increment is used as a screening criterion for evolutionary operators. Populations are stored according to their involvement and diversity, making evolution still more different. In the entire process, a random regeneration strategy is employed to maintain population diversity. This approach not only makes sure the explore space of operators but also reduces the population degradation. The experimental results demonstrate that the Regenerative Genetic algorithm could recognize the group structure with higher cluster accuracy and lower computational cost when compared with existing algorithms.

A new fuzzy genetic algorithm proposed for community identification in social networks. In this proposal, we have used matrix encoding that enables traditional crossover between individuals and mutation takes place in some of the individuals. Matrix encoding determines which node belongs to which community. Using these concepts enhance the overall performance of any evolutionary algorithms. In this experiment, we used the genetic algorithm with the fuzzy concept and compared to other existing methods like as crisp genetic algorithm and vertex similarity based genetic algorithm. We employed the three real world dataset strike, Karate Club, Dolphin in this work. The usefulness and efficiency of proposed algorithm are verified through the accuracy and quality metrics and provide a rank of proposed algorithm using multiple criteria decision-making method.

Community structure identification is an important task in social network analysis. Social communications exist with some social situation and communities are a fundamental form of social contexts. Social network is application of web mining and web mining is also an application of data mining. Social network is a type of structure made up of a set of social actors like as persons or organizations, sets of pair ties, and other interaction socially between actors. In recent scenario community detection in social networks is a very hot and dynamic area of

research. In this work, we have used improvised genetic algorithm for community detection in social networks, we used the combination of roulette wheel selection and square quadratic knapsack problem. We have executed the experiment on different datasets i.e. Zachary's karate club [31], American college football [39], Dolphin social network [32] and many more. All are verified and well known datasets in the research world of social network analysis. An experimental result shows the improvement on convergence rate of proposed algorithm and discovered communities are highly inclined towards quality.

Community detection is the fundamental issue in the social network analysis. In present scenario, complex network analysis is sizzling research area. We proposed the new idea of finding the fuzzy community detection in social network with the help of permanence concept with node similarity based genetic algorithm. In this proposed work we found the both disjoint community and the overlapping community detection then we compares it as quality wise and accuracy wise with the help of some functions. We utilized the disjoint community structure as an input for our base algorithm. We employed the artificial datasets and the real world datasets for our experiment. In this proposal we compare the disjoint community and the fuzzy community for our proposed algorithm with the help of some quality and accuracy based parameters. The proposed method eliminates the need of a separate algorithm for fuzzy community structures. It fulfills the role of both disjoint community detection and fuzzy community detection without adding any extra step of genetic algorithm.

Community detection algorithms in social networks using Differential evolution have been studied. For this, experiments are performed using differential evolution algorithm with multiple objective functions and the results are analyzed for the number of datasets. Differential evolution algorithm uses modularity as a fitness function. In this effort, the different objective functions are used such as conductance, normalized cut, internal density, average degree, expansion, Cut ratio. In this aspect, the approach used here is to find the best competitive variation of differential evolution algorithm for the several conditions and on the different type of datasets. The best one is selected for specific conditions. So a survey of community detection in social networks and evolutionary algorithms is presented along with the experimental results which are based on multiple datasets.

Most of the real world networks we encounter today are complex networks and one of the important characteristics of these networks is the community structure. Identifying communities in a complex network is classified as computably hard and thus many meta-heuristic approaches have been proposed in the past. We propose an improved differential evolution based algorithm which exploits the structural similarity of the network to generate a better initial population leading to a more accurate identification of communities. We have tested our algorithm on various well-known real world and artificial networks.

The concept of the opposition has an old history in sciences and other fields. This is the first time to contribute in social networks for community detection to enhance an optimizer. In this proposal, present a novel scheme to make the differential evolution algorithm faster. The proposed opposition based DE employs opposition based optimization for population initialization. Opposite numbers have been utilized to improve the convergence rate of traditional DE. The proposed algorithm also employs the tournament selection method for mutation and accelerates the differential evolution. We have to employed the combination of opposition based learning and Tournament selection method with DE for community detection in social networks. Combination of technique are optimized the results.

In this proposal, we used the differential evolution algorithm for community detection in complex network. DE is a evolutionary technique i.e. swarm techniques. We improved the classical version of Differential evolution for optimization. In this work, we modified the DE with the help of some other concepts i.e. initialization process modified by opposition based concept and selection process done by the tournament method. A new DE algorithm employed the real world and artificial datasets for community detection in social networks.

In recent scenario, DE is a very well known algorithm for optimization after my experiment gain the new height in social network analysis and community detection. In this experiment, DE is good for accuracy-wise and quality-wise for community identification. According to results new DE is better performance compare to other evolutionary algorithm.