Chapter 11
Evolutionary Computation Techniques for Community Detection in Social Network Analysis

Abhishek Garg  
Indian Institute of Technology (BHU) Varanasi, India

Anupam Biswas  
Indian Institute of Technology (BHU) Varanasi, India

Bhaskar Biswas  
Indian Institute of Technology (BHU) Varanasi, India

ABSTRACT

Community detection is a topic of great interest in complex network analysis. The basic problem is to identify closely connected groups of nodes (i.e. the communities) from the networks of various objects represented in the form of a graph. Often, the problem is expressed as an optimization problem, where popular optimization techniques such as evolutionary computation techniques are utilized. The importance of these approaches is increasing for efficient community detection with the rapidly growing networks. The primary focus of this chapter is to study the applicability of such techniques for community detection. Our study includes the utilization of Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) with their numerous variants developed specifically for community detection. We have discussed several issues related to community detection, GA, PSO and the major hurdles faced during the implication of evolutionary approaches. In addition, the chapter also includes a detailed study of how these issues are being tackled with the various developments happening in the domain.

1. INTRODUCTION

Nowadays, representation of links among various objects present within the data in the form of network is very common in many domains. Such network representation in modern-day data which include social network (Scott, 2012), ecological network (Newman, 2012), biological network (Sah et al., 2014),
Evolutionary Computation Techniques for Community Detection in Social Network Analysis

biochemical network (Bennett et. al., 2014), citation network (Chen and Redner, 2010), web network (Xiaodong et. al., 2008) and ego network (Biswas and Biswas, 2015) etc are very complex. Analysis of such data with network representation eases the understanding of complex relationships among the objects of the data. The core of such analysis is to understand the relationships through different properties of networks such as network transitivity (Han et. al., 2015), power-law degree distribution (Dorogovtsev and Mendes, 2013), and community structure (Fortunato and Castellano, 2012). In this chapter, we limit our discussion only to community structure property. Incorporation of community structure property in the analysis of such complex networks requires detection of communities. The community detection problem is closely related to the clustering problem that is basically an unsupervised learning of groups based on some similarity measures (Schaeffer, S. E., 2007). Therefore, often community detection problem is also referred as a graph clustering.

Numerous approaches have been developed for detecting communities efficiently. These approaches can be categorized broadly into four categories: Node centric, Group centric, Network centric and Hierarchy centric community detection. In node centric community detection, each node or object of a community has to follow certain properties. Clique percolation method is a node centric approach that is based on complete mutuality of nodes. In group centric community detection, group as a whole has to follow certain properties without looking into details of nodes. One such property is group-density (Pattillo et. al., 2013) that must be greater than certain threshold. In network centric community detection, whole network is partitioned into various groups by considering connections of whole network. A function is defined covering entire network and optimized that function. Popular methods such as modularity maximization approaches (Newman, 2004; Clauset et. al., 2004; Blondel et. al., 2008; Medus et. al., 2005; Duch and Arenas, 2005) and spectral clustering (Chauhan et. al., 2009; Arenas et. al., 2006; Shen and Cheng, 2010; Richardson et. al., 2009) are fall under this category. In hierarchy centric community detection, hierarchical structure of communities is formed by analyzing network at different resolution, where hierarchy is represented with a dendrogram. Agglomerative clustering approaches and divisive clustering approaches (Schaeffer, S. E., 2007) are the examples of hierarchy centric community detection approaches. Among these approaches, the network centric approaches are the essence of this chapter where requires optimization of a function bound to the network.

The function that is defined in network centric approaches has to be optimized efficiently to get effective communities. Here, two major questions needed to be resolved. First question is about how to design a suitable function for obtaining better communities. Follow-up second question is about how to optimize the defined function. One approach for resolving the first question is through rigorous study of network and required communities. Another alternative, easy, and widely utilized approach is simply use the quality metrics. Numerous quality metrics such as modularity (Newman and Girvan, 2004), conductance (Kannan et. al., 2004), Silhouette Index (Tan et. al., 2006), cut ratio (Fortunato, 2009), community score (Pizzuti, 2008) are already proposed. These quality metrics are actually defined for measuring the quality of communities detected with any algorithm. Newman is the first to utilize these metrics directly for optimization. He has proposed a greedy approach that optimizes the quality metric modularity (Newman, 2004). Such optimization has been proved as NP-hard problem (Brandes et. al., 2006). Thus for resolving second question several approximation methods (Newman, 2004; Blondel et. al., 2008) has been proposed to detect communities efficiently. Evolutionary computation techniques are in the forefront to solve NP-hard problem. These approaches are competent to detect comparatively better quality communities then other approaches in less time (Tasgin et. al., 2007; Xiaodong et. al., 2008). We consider two of such techniques namely Genetic Algorithm (GA) and Particle Swarm Opti-