Aspects of Various Community Detection Algorithms in Social Network Analysis

S

Nicole Belinda Dillen
St. Thomas' College of Engineering and Technology, India

Aruna Chakraborty
St. Thomas' College of Engineering and Technology, India

INTRODUCTION

The 21st century is dominated by social media. Social networking websites like Facebook, Twitter and Myspace connect millions of people from all corners of the world. As a result, the global community is growing at a fast pace with new connections being forged every second. Businesses, government corporations and other organizations are relying increasingly on online networking in order to promote products, conduct surveys and target specific audiences. The immense importance of social media has catalyzed research in social network analysis, a field that had gained prominence in the mid '90s. Social network analysis involves the detection of communities or closely connected groups of individuals, studying the characteristic patterns associated with these communities, identifying important "key" actors in these networks and finally studying the overall behavior associated with people in the network.

BACKGROUND

Community detection in social networks is one of the most important areas of social network analysis. The most widely accepted definition of a community is a closely connected group of individuals with sparse connections to individuals belonging to other communities. With this definition in mind several algorithms have been implemented which serve to identify community structure in social networks.

DOI: 10.4018/978-1-5225-2255-3.ch603

Algorithms such as those proposed by Kernighan (1970) are traditional graph partitioning algorithms which use clustering techniques like *k*-means (MacQueen, 1967) to form graph clusters and can, therefore, be applied on social networks. The problem with these algorithms is that they can only find either a fixed number of communities or community clusters of a certain size, both of which may not really be the best case for real-world scenarios.

A very popular algorithm was developed by Girvan and Newman (2002) and uses a divisive technique in which edges are constantly removed to split a social network into smaller and smaller component networks. This technique gives rise to the natural community structure inherent in the network which can be represented through a community hierarchy. However, certain issues such as time complexity prompted one of the authors to devise an entirely new approach to detect community structure by optimizing a function known as *modularity* (Newman, 2004). This technique paved the way for a number of subsequent algorithms that collectively constitute the class of modularity-optimization algorithms. These include an improvement on Newman's fast algorithm (Clauset, Newman, & Moore, 2004) and an effort to improve community quality by balancing the communities detected by the CNM method (Wakita, & Tsurumi, 2007). The Louvain method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is, perhaps, one of the most popular greedy modularity-optimization algorithms which

is modelled on Newman's original modularity-based algorithm.

Further research developments led to the idea that community structure may not consist of just well-defined "crisp" clusters; rather, most real-world social networks contain communities that naturally overlap to a certain extent. Overlapping community detection algorithms serve to identify this natural community structure and either allocate the overlapping nodes to a single "best-fit" community through modularity optimization (Berti, Sperduti, & Burattin, 2014) or preserve the overlapping structure through the use of fuzzy membership functions (Kundu & Pal, 2015b).

COMMUNITY DETECTION ALGORITHMS: NEW AND OLD

Traditional View of Communities in Social Networks

Most community detection algorithms consider a model in which each individual in a network belongs to a single community. The individuals in this community will have many connections with each other but will have a minimal number of connections to individuals belonging to other communities. In fact, this very phenomenon is exploited by nearly all of the community detection algorithms prevalent in social network analysis.

Overlapping Communities in Social Networks

Contrary to the traditional community model mentioned above, many recent research developments deal with more realistic models in which multiple communities may be assigned to an individual. In real world social networks, it is only natural for an individual to be a part of several communities simultaneously. For example, a student may be part of his/her study group consisting of close friends, and at the same time belong to the school's debate team which is an entirely different community. Many community detection algorithms fail to take this relevant aspect into consideration and are, hence, limited in their approach. The newer models, therefore, describe what are known as "overlapping" communities.

Figure 1. A graph representing the Dolphin Social Network: lines indicate edges and circles indicate vertices