Complex networks such as the Internet, the World Wide Web (WWW), and various social and biological networks, are viewed as large, dynamic, random graphs, with properties significantly different from those of the Erdős-Rényi random graphs. In particular, properties such as degree distribution, network distance, transitivity and clustering coefficient of these networks have been empirically shown to diverge from classical random networks. Existence of communities is one such property inherent to these networks. A community may informally be defined as a locally-dense induced subgraph, of significant size, in a large globally-sparse graph. Recent empirical results reveal communities in networks spanning across different disciplines—physics, statistics, sociology, biology, and linguistics. Various community detection algorithms have been proposed in the literature to detect and extract communities in such networks. A comprehensive taxonomy of the existing community detection algorithms is presented in this work. However, community detection is challenging and even more exigent in case of complex networks.

Given an input graph, algorithms may focus on extracting all the communities. These algorithms are commonly known as community discovery algorithms. Global exploration of these graphs to pull out communities is time and memory consuming. A more confined approach to mine communities in a given network is investigated in this research. Given a graph and a seed vertex, the task is to identify the community that the seed node belongs to. This relatively easier alternative methodology of community detection is referred as community identification. Both community discovery and community identification problems are NP-Complete, but the latter does not require the knowledge of the entire graph. Community identification algorithms exist in the literature, but to a smaller extent. A thorough description and analysis of the existing techniques to identify communities in large networks is presented. Also a novel heuristic for identifying the community to which a given seed node belongs using only its neighborhood information is proposed. We also discuss the various community definitions in the literature and propose an improved definition of a community based on the average degree of the induced subgraph. Next, a faster algorithm to identify community in complex networks based on maximizing the average degree is proposed. The performance of the algorithm on several synthetic and real-world complex networks has been thoroughly investigated.

Major: Computer Science

Educational Career:
Bachelor's of B.E., BS, 2006, Anna University

Committee in Charge:
Dr. Narsingh Deo, Chair, Electrical Engineering and Computer Science
Dr. Charles Hughes, Electrical Engineering and Computer Science
Dr. Ratan Guha, Electrical Engineering and Computer Science
Dr. Mainak Chatterjee, Electrical Engineering and Computer Science
Dr. Yue Zhao, Mathematics

Approved for distribution by Dr. Narsingh Deo, Committee Chair, on March 13, 2012.

The public is welcome to attend.