PREFACE

Complex networks such as social networks exhibit disproportionate connections among different nodes, resulting in densely interconnected groups of nodes. These highly connected groups of nodes within the network are referred as communities, which have significant role in understanding and uncovering various functional properties of the system. Identification of communities has grown as one of the major research topics in social network analysis. In this thesis, broadly four different aspects related to communities are studied: 1) assurance of accuracy and role of nodes, 2) network diversity, 3) application of communities, and 4) evaluation of communities.

Communities are investigated from the perspective of ego network to explore the role of nodes in community formation. The notion of mutual interest in the relationship is introduced by extending the personalized view of ego network, and defined two properties: Reachability and Isolability. Exploring these two properties, an Ego Network Based Community detection algorithm (ENBC) is proposed for identifying communities. Incorporation of Reachability confines the decision on whether a node would belong to a community to the node itself, which is decided only by the members of the community in most of the existing algorithms. Enhancing the role of individual nodes in community detection process shows highly accurate communities.

Further exploring the role of nodes, a fuzzy agglomerative community detection algorithm (FuzAg) is proposed to identify both disjoint and overlapping communities. The algorithm utilizes the Reachability property to compute membership degree of nodes to different communities. The notion of self-membership is introduced in addition to the membership of different communities. The essence of self-membership is to ascertain opportunity to all the nodes in growing their own community. A node having sufficient self-membership degree is referred as an anchor; subsequently the node gets a chance

to grow new community around it within the network. Incorporation of both Reachability property and the notion of self-membership deepen the role of nodes in community detection process, resulting in highly accurate communities.

Nowadays, multiple connections between two nodes exist in some networks, which are referred as multiple featured networks. An improved Particle Swarm Optimization (PSO) is proposed to identify communities in such networks. A novel cognitive avoidance mechanism is introduced in the standard PSO to improve its performance. In this setting, each particle maintains its worst value along with the personal best and global best. With this known worst value, the particles try to avoid further movement towards it, having the sense that solutions near the worst one may not be suitable.

An application of community detection problem is also studied. A Community-based Link Prediction (CLP) algorithm is proposed to predict missing links involving community information. The locality of the nodes associated with a connection is considered to assign weights to the connection, while edge centrality measures are considered to define importance of neighbors of a node. Incorporation of community information in link prediction ensures positive influence on the scores if the associated nodes belong to same community, otherwise influences negatively.

Finally, the problem of community evaluation is addressed. Evaluation of communities has two aspects: 1) validation metrics and 2) evaluation methodology. A set of three quality metrics is proposed. These metrics have the ability to assure accuracy better than the existing metrics. On the ground of evaluation methodology, a framework is proposed to analyze Relative Inclination Towards Accuracy (RITA) of a set of community detection algorithms. RITA analysis gives an intuition about how likely an algorithm would produce accurate communities. Another methodology is proposed to analyze evolutionary optimization algorithms and community detection algorithms. A shifting mechanism of regression line is developed by harnessing the concepts of quantile-quantile plot and linear regression analysis. Proposed methodology is designed based on the newly developed shifting mechanism to compare performance of different algorithms visually.