Title of Thesis: Community Structure Analysis in Complex Networks for Ranking Influential Users

Name of the Candidate: Ahmadi Hasan

Name of Guide: Dr. Ahmad Kamal

PhD Program: PhD in Mathematics

Enrollment No.: 10-5278

Notification No.: F.NO. COE/Ph.D./(Notification)/552/2024

Notification Date: 25-01-2024

Name of the Institute: Jamia Millia Islamia

FINDINGS

Network science is a multidisciplinary field dedicated to the analysis of complex networks, their modeling, and the propagation of information between the nodes of a network in order to understand the behavioural aspects of each node in large-scale complex networks. Due to the explosion of large, heterogeneous, and unstructured data from various domains of complex networks, such as social networks, biological networks, or any other real-world networks that comprise real entities, there is a great opportunity to analyze the dynamics of nodes at different levels. This has resulted in the development of computationally intelligent techniques to extract, transform, model, evaluate, and visualize complex network structures.

Chapter 1. Introduction

This chapter is introductory, deals with a general introduction and mathematical foundation of the research topic. The chapter describes briefly the objectives and overview of the thesis.

Chapter 2. State-of-the-Art Techniques in Community Detection and Influential Node Mining

In this chapter, the work is deeply focused on highlighting challenges related to this area of research, including community structure identification and influential node mining. We present a comprehensive survey of various community detection approaches in existing literature, which includes disjoint community detection and overlapping community detection. We also present influential node mining approaches, which mainly aim to address the centrality measures problem in identifying influential nodes. Using communities for influential node mining has been discussed. The final section of this chapter provides a brief overview of a variety of application fields, including fake news classification and propagation path analysis over a news network constructed by measuring cosine similarity in online news articles. The main objective of this chapter is to understand current issues and challenges related to this field of study and subsequently to develop effcient algorithms for mining relevant knowledge from complex networks.

Chapter 3. Community Detection of Weighted Complex Network via Transitive Closure (CoDTC)

This chapter presents a community detection algorithm based on transitive closure that incorporates degree centrality for the selection of initial cluster centers. The novelty of the proposed community detection algorithm is that it does not choose initial cluster centers randomly, which leads to minimizing superfluous runs. Instead, it determines initial cluster centers with the help of the degree of centrality of each node in the underlying network. The other novelty of the proposed community detection algorithm is that it offers the idea of

transitive closure on weighted networks to solve the sparsity issue. This notion is based on imposing a T-norm inequality on the connection weights and providing a method to compute them. The approach demonstrates the effectiveness of the CoDTC method on various real and artificial networks with large and small communities.

Chapter 4. Overlapping Community Detection using Laplacian Eigenmaps and Fuzzy C-Means (LapEFCM)

This chapter proposed an overlapping community detection method in complex networks using a dimension reduction approach. Since communities are not always disjoint and might overlap, the fuzzy clustering algorithm is the most feasible approach to detecting overlapping communities. However, one of the major disadvantages of this approach lies in the prior specification of the number of communities. The LapEFCM method adopts a modularity strategy to obtain a number of overlapping communities, which overcomes the disadvantage of fuzzy clustering.

Chapter 5. Ranking Top-k Influential Nodes in Complex Networks with Random Walk (k-InfNode)

This chapter presents the influential node mining algorithm that captures the dynamics of nodes, which are based on the characteristics of community structure. Inspired by the idea of overlapping nodes that reflect the interaction of nodes and communities over the network, the k-InfNode algorithm adopts a random walk and integrates the local and global properties to determine node significance in a complex network.

Chapter 6. A Network Analysis Approach for Fake News Classification (NetFNC): Case Study of Influential Node Mining

It has been observed that only a small number of influential node mining methods proposed in the literature describe their practical implementations. This chapter demonstrates how social network nodes can be used to detect fake news in the real world. With the advent of online social media networks, news dissemination is readily available, but the content lacks veracity. The interactions generate numerous "fake news" dissemination variations. Fake news on social media has the potential to severely harm society. The availability of heterogeneous data makes it difficult to detect fake news from a technical standpoint. This chapter presents a combination
of network-based and machine learning techniques for classifying fake news articles. In order to identify principal nodes that exert a significant influence on other nodes, extracted news articles are initially modeled as nodes within a news network. Later, the textual content of influential nodes is accessed to identify linguistic features. Using machine learning techniques, network properties and linguistic features are incorporated into a binary classification model to identify fake news articles.

Chapter 7. Conclusion

This chapter concludes our work. We have discussed some unresolved problems and future directions in the area of community structure detection, influential node identification and its application in real-world complex networks.