

Chapter 8

A Survey on Overlapping Communities in Large-Scale Social Networks

S Rao Chintalapudi
JNTUK, India

H. M. Krishna Prasad M
JNTUK, India

ABSTRACT

Social network analysis is one of the emerging research areas in the modern world. Social networks can be adapted to all the sectors by using graph theory concepts such as transportation networks, collaboration networks, and biological networks and so on. The most important property of social networks is community, collection of nodes with dense connections inside and sparse connections at outside. Community detection is similar to clustering analysis and has many applications in the real-time world such as recommendation systems, target marketing and so on. Community detection algorithms are broadly classified into two categories. One is disjoint community detection algorithms and the other is overlapping community detection algorithms. This chapter reviews overlapping community detection algorithms with their strengths and limitations. To evaluate these algorithms, a popular synthetic network generator, i.e., LFR benchmark generator and the new extended quality measures are discussed in detail.

DOI: 10.4018/978-1-5225-2805-0.ch008

1. INTRODUCTION

With the advent of internet, websites and social networks generate large amount of data. Social Network Analysis (Aggarwal 2014) (Reza et al., 2014) is a traditional research area that faces the challenge of big data in the recent years. As social networks are getting more popular, analyzing such networks has become one of the most important issues in various areas. In the era of big data, the amount of available data is growing unprecedentedly. Thus, data analysis techniques need very scalable approaches that can cope with huge network datasets. The significance of big data has attracted the concern of governments, companies and scientific institutions. The voluminous data available in social network sites and web like Facebook, Twitter, Instagram, Linkedin, Weibo, World Wide Web and Wikipedia can be treated as bi data. This data can be represented as a graph/ Network, where nodes denote persons or pages, while edges represent the relationship between persons or pages. This relationship represents following in Twitter, friendship in Facebook, professional connections in LinkedIn, hyperlinks in WWW. This data may constitute several communities based interactions between people or entities. The members of community have some common interests such as movies, travel, photography, music, novels ... etc and hence, they tend to interact more frequently within the community than the outside. Finding Communities in social networks is similar to Graph clustering problem, hence, it can also be called as Big graph clustering with respect to larger networks.

The basic objective of the chapter is to answer the following three questions in detail.

1. What is community detection?
2. How can we detect overlapping communities?
3. How can we evaluate detected overlapping communities?

This chapter is organized as follows. Section 2 discusses about the preliminaries of community detection. The several popular overlapping community detection algorithms with their strengths, limitations are discussed in section 3. In section 4, the synthetic networks and real world networks with their quality measures such as Overlapping Normalized Mutual Information (ONMI), Omega index, F-Score, Overlap Modularity that can be used for evaluating overlapping community detection algorithms were discussed. Finally, conclusions and the future research directions were discussed in section 5.

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/a-survey-on-overlapping-communities-in-large-scale-social-networks/185984

Related Content

From Personal to Mobile Healthcare: Challenges and Opportunities

Elena Villalba-Mora, Ignacio Peinado and Leocadio Rodríguez-Mañas (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 2415-2428).

www.irma-international.org/chapter/from-personal-to-mobile-healthcare/150272

Gridifying Neuroscientific Pipelines: A SOA Recipe and Experience from the neuGRID Project

David Manset (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 1893-1915).

www.irma-international.org/chapter/gridifying-neuroscientific-pipelines/73547

OCL Constraints Checking on NoSQL Systems Through an MDA-Based Approach

Fatma Abdelhedi, Amal Ait Brahim and Gilles Zurfluh (2021). *International Journal of Data Warehousing and Mining* (pp. 1-14).

www.irma-international.org/article/ocl-constraints-checking-on-nosql-systems-through-an-mda-based-approach/272015

Sentiment, Stance, and Intent Detection in Turkish Tweets

Dilek Küçük (2021). *New Opportunities for Sentiment Analysis and Information Processing* (pp. 206-217).

www.irma-international.org/chapter/sentiment-stance-and-intent-detection-in-turkish-tweets/286912

Understanding the SNN Input Parameters and How They Affect the Clustering Results

Guilherme Moreira, Maribel Yasmina Santos, João Moura Pires and João Galvão (2015). *International Journal of Data Warehousing and Mining* (pp. 26-48).

www.irma-international.org/article/understanding-the-snn-input-parameters-and-how-they-affect-the-clustering-results/129523