Extracting Influential Nodes in Social Networks on Local Weight Aspect

Jun Jun Cheng, China Information Technology Security Evaluation Center, Beijing, China
Yan Chao Zhang, Institute of ICT Security Research, China Academy of Information and Communications Technology, Beijing, China
Xin Zhou, China Information Technology Security Evaluation Center, Beijing, China
Hui Cheng, State Grid Qingdao Power Supply Company, Qingdao, China

ABSTRACT

Studies have shown that influential nodes play an important role in all kinds of dynamic behavior in the complex network. Excavation or recognition of such nodes contributes to the development of application areas such as social network advertising and user interest recommendation. Although some heuristic algorithms such as degree, betweenness, closeness and k-shell (or k-core) can identify influential nodes at the same time, they are disadvantaged in terms of accuracy and time complexity. Based on this, the authors propose a novel local weight index to distinguish the node influence based on the theory of ties strength. This index emphasizes that the node influence is jointly decided by the quantity and quality of the neighbors, and its time complexity is much lower than closeness and betweenness. With the aid of SIR information transmission model, this paper verifies the validity of local weight index.

KEYWORDS

Influential Nodes Extraction, Local Weight Index, Social Networks, Ties Strength

1. INTRODUCTION

In recent years, many scholars are concerned about researches on the social network topology analysis, the relationship between the network topology and dynamic behaviors based on the various types of complex networks (Albert & Barabási, 2002). As is well-known, a small part of the influential nodes play a very important role in the evolution of all kinds of dynamic behaviors in the complex network, such as network cascade (Motter & Lai, 2002), information transmission (Pastor-Satorras & Vespignani, 2002) and nodes synchronization (Zhao et al., 2005), etc. Therefore, excavating key nodes in the network has very important theoretical research value.

In addition, this work also has some important practical applications. For example, some targeted strategies taking into account of some influential nodes can effectively inhibit the spread of rumors and diseases, or create a bigger market for certain services and products in the economic field. In this regard, researchers in related fields from all over the world have done a lot of useful attempts. In the article by D. B. Chen et al. (2012), the authors proposed a new local index combining efficiency and accuracy, and the simulation based on the SIR model in several real networks shows that the index can better identify opinion leaders in the network. In the article by X. Huang et al. (2011), the authors proposed a centricity index called Influence Factor, and then excavated the important executives in the management network of the American corporations. Compared with other centricity indexes,
Influence Factor can recognize influential executives more accurately. In addition, scholars have also done a lot of useful attempts in terms of protecting key areas of the complex network from malicious attacks (Wuellner et al., 2010; Albert et al., 2000) and ranking the reputation of the publications, scientists (Radicchi et al., 2009; Chen et al., 2007) and professional tennis players (Radicchi & Perc, 2011) in the scientific cooperation network.

In the complex network theory, the most direct and convenient index is degree centrality, but this index has some obvious limitations. For example, it may have a larger sphere of influence when information is spread out from an ordinary node who has a friendship with numerous influential nodes though its degree centrality is very small. Moreover, the node power measurement algorithm of some classic global indexes such as betweenness centrality and closeness centrality indexes can’t be applied in the large-scale online social network either. Though they can obtain more excellent evaluation results, their computation complexity is much higher. Recently, in the article by L. Y. Lü et al. (2011), the authors proposed Leader Rank algorithm based on the random walk theory. Compared to the famous PageRank, this algorithm shows a better performance in recognizing the most influential node in the opinion propagation process. Leader Rank and Page Rank are both applicable to the directed network, but they are unsatisfactory in the undirected network in which they will degenerate to degree centrality. Additionally, the article (Kitsak et al., 2010) insist that the most influential node is not hub nodes or larger betweenness and closeness nodes, but the core layer nodes after the k-core decomposition. In summary, how to design an effective and convenient ranking method to identify the influential nodes in the complex network is still a serious problem.

Based on this, this paper proposes a more effective local-weight-style node centrality index (LW) which can approximately measure nodes influence. The core idea of LW index is similar to PageRank in that the quality of the neighbor nodes determines their own level of influence. Our index weakens the interference of the node degree, and at the same time its computational complexity is lower than betweenness and closeness. From the qualitative analysis view, the LW index makes a compromise between performance and consumption. Besides, from the quantitative view, this paper uses SIR model to examine the spread influence of nodes ranked according to different centrality index and verifies the performance of the LW index in two real networks.

This paper can provide reliable algorithm support for the advertising and marketing of achieving higher earnings with a lower cost and provide an effective basis for the spread influence of the quantitative nodes in the concrete network topology. At the same time, this paper contributes to further study the complexity of the network topology features, providing a favorable research tool for complex dynamic network researches.

The rest of this paper is organized as follows: Section 2 introduces the real data sets and the local weight index. Section 3 reviews and analyzes the simulation results. Finally, section 4 summarizes and concludes the study.

2. MATERIALS AND METHODS

2.1. Data Sets

The two data sets used in this paper are both taken from real networks (Zafarani & Liu, 2009). One comes from BlogCataLog, whose homepage will recommend some distinctive published blog and list the most popular and latest blog users. In this data set, each node corresponds to a blog user, and the edge between two nodes represents a friend relationship between two blog users. Another data set comes from Delicious. This website is currently the world’s largest bookmarking website. Basic functions provided through the website: collect, tag, and review, can automatically enable users to easily store, share and discover their favorite website links. The two real networks used in this paper are both undirected graphs. The basic topology characteristics of the two data sets of statistics are shown in Table 1. Among them, |V| represents the nodes total count, |E| is the edges count, d is the
Precise Analysis and Evaluation of OFDM Systems over a Mobile Frequency-Selective Nakagami-m Fading Channels
[www.igi-global.com/article/precise-analysis-evaluation-ofdm-systems/1468?camid=4v1a](www.igi-global.com/article/precise-analysis-evaluation-ofdm-systems/1468?camid=4v1a)

Adjust Fuzzy Model Parameters for Head Election in Wireless Sensor Network Protocols

Handoff Cost Minimization and Planning of Heterogeneous Integrated Overlay Networks: Meta-Heuristics Based Approach
Ayan Paul, Madhubanti Maitra, Swarup Mandal and Samir Kumar Sadhukhan (2012).
*Next Generation Data Communication Technologies: Emerging Trends* (pp. 49-64).
[www.igi-global.com/chapter/handoff-cost-minimization-planning-heterogeneous/61747?camid=4v1a](www.igi-global.com/chapter/handoff-cost-minimization-planning-heterogeneous/61747?camid=4v1a)