Physica A 442 (2016) 458–466

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Role extraction in complex networks and its application in control of networks

Mingyang Zhou *, Xingsheng He, Zhongqian Fu, Zhao Zhuo

Department of Electronic Science and Technology, University of Science and Technology of China, Heifei 230027, Ahhui, PR China

HIGHLIGHTS

- We utilize nodes' roles to describe differences of nodes in the dynamics.
- According to the behavior differences in dynamics, three roles (Leader, Communicator and Member) are given in advance and they are easier to comprehend and apply into practice.
- Based on community entropy, a fast time saving and unsupervised learning approach is proposed to extract nodes' roles just from network connections.
- Communicator nodes are utilized as pinning nodes and the controllability is enhanced remarkably in networks with strong community structure.

ARTICLE INFO

Article history:
Received 9 January 2015
Received in revised form 10 May 2015
Available online 25 September 2015

Keywords:
Role extraction
Entropy
Pinning control
Complex network

ABSTRACT

Given a large network, dynamics of the network are determined by both nodes' features and network connections. Some features could be extracted from node labels and other kinds of *priori* knowledge. But how to perform the feature classification without *priori* knowledge is a challenge. This paper addresses the key problem: how do we conduct role extraction in networks with only edge connections known? On the basis of behavior differences in dynamics, nodes are classified into three role groups: Leaders(L), Communicators(C) and Members(M). Unlike traditional community detections, we detect overlapping communities by link clustering first and then classify nodes according to the community entropy, which describes the disorder of how many different communities a node connects to. We propose a time saving and unsupervised learning approach for automatically discovering nodes' roles based solely on network topology. The effectiveness of this method is demonstrated on six real-world networks through pinning control. By controlling communicator nodes, the controllability is enhanced and the cost for control is reduced obviously in networks with strong community structure.

1. Introduction

Given a network with only edge connections known, how to capture nodes' dynamical differences (e.g. synchronization, diffusion and information navigation) is an attractive field [1,2]. In the dynamics, a small fraction of nodes dramatically influence the dynamics [3,4]. Thus, finding those important nodes helps understand the mechanism of dynamics. Many real networks follow scale-free (power-law) similar degree distribution [5] and nodes with high degree play an important role in

* Corresponding author.
E-mail addresses: zhoumy2010@gmail.com (M. Zhou), zhzh7532@mail.ustc.edu.cn (Z. Zhuo).

http://dx.doi.org/10.1016/j.physa.2015.08.035
0378-4371/

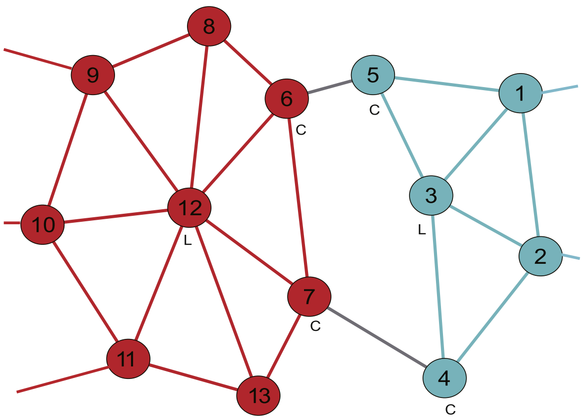

Fig. 1. An artificial network that demonstrates the difference of roles. Leader nodes (L) and Communicator nodes (C) are depicted in the network. Unlabeled nodes are periphery nodes (member).

scale-free model networks [6], where scale-free model networks follow power-law degree distribution, but the connections are completely random. So conventional wisdom suggests that high degree nodes are the key nodes. But recent study finds that some low degree nodes also act as key nodes and many works have concerned on identifying those influential nodes [3,7]. Though practical networks follow scale-free similar degree distribution, they also reveal community structure, assortative mixing pattern, hierarchical organization and so on Refs. [8–14], which influence deeply on the dynamics [3,15–18]. Thus, research about complex networks attracts scientists from multiple disciplines and a large number of methods have been proposed to analyze the relations between the structure and dynamics of networks [8–11]. Among those methods, Page-Rank, random walk and other machine learning methods detect influential nodes well in specific kinds of networks, but not appropriate for all cases [19,20]. Those methods usually optimize an objective function to identify the key nodes [9]. However, it is difficult to determine the optimal objective function and previous functions fail in some networks [9,21].

In practical organizations, some nodes take charge of information exchange in the same community; some are responsible for communication across different communities. Further, inner communication is much stronger than communication across different groups. Based on the idea, Arenas et al. found networks first reach local synchronization (nodes in the same communities reach the same state first), and then collective synchronization (all nodes reach the same state) [15]. Jesús et al. studied synchronization paths and found synchronization paths varied in networks with different community structure [22]. Wang et al. investigated pinning control in scale-free model networks and found that large-degree selection performs better than random selection [4]. But in real-world networks, multiple large degree nodes performs badly in the spreading process due to overlapping influence of the nodes [3]. Thus, we need to explore the dynamical differences for each nodes. Since the dynamical difference could be distinguished by nodes' roles, we could investigate the dynamical differences from the perspective of nodes' roles. But nodes' roles are inaccessible without *priori* knowledge. So determining the underlying roles provide an easy access to comprehend the dynamics.

In our framework, inspired by practical social organization [1], we classify nodes into three kinds of roles: Leader (L), Communicator (C), Member (M, peripheral nodes in communities). Practical organizations are usually made up of groups (communities) and each group has several principals (*i.e.* Leader). Principals mainly take charge of the inner communication in the same group and some special people are in charge of communication across different groups. For example, in the synchronization paths, principals take the role of Leader and nodes who are in charge of communication across different groups take the role of Communicator. The other nodes are Members (peripheral nodes). Leader nodes promote information exchange extensively in the same community, which results in local synchronization. While communicator nodes benefit information spreading across the whole networks that leads to collective synchronization. As Fig. 1 shows, nodes 3 and 12 act as leaders in communities and mainly take charge of communication in the communities. Nodes 4, 5, 6 and 7 are mainly responsible for information transfer among different communities and become communicators. The remaining nodes are peripheral nodes. Fig. 1 shows that leader nodes mainly links to nodes within the same community, while communicator nodes mainly connects to different communities. Inspired by the phenomenon, we first detect overlapping communities, and then distinguish nodes' roles according to their connections to different communities. To verify the efficiency of role extraction, the communicator nodes, which bridge multiple communities and exert great influence on information exchange across different communities (influential bridge nodes), are chosen as driver nodes in pinning control. Those driver nodes achieve better controllability than that of conventional large-degree selection in networks with strong community structure, but much lower cost, which demonstrates the effectiveness of our proposed method.

The rest of the paper is organized as follows: In Section 2, role extraction method is introduced first. Then, in Section 3, we apply the proposed method into six real-world networks and extract roles for nodes; later, the communicators are selected as driver nodes in pinning control which performs better than that of large-degree selection. The conclusion is given at last.