Physica A 448 (2016) 68–80

Contents lists available at ScienceDirect

Physica A

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

Integrated structure investigation in complex networks by label propagation

Tao Wu [a,b,d,*], Yuxiao Guo [a], Leiting Chen [a,b,d], Yanbing Liu [c]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[b] Institute of Electronic and Information Engineering in Dongguan, University of Electronic Science and Technology of China, Chengdu 611731, China
[c] Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[d] Digital Media Technology Key Laboratory of Sichuan Province, Chengdu 611731, China

HIGHLIGHTS

- Define several novel and effective influence measures.
- Explore influence distribution to propose a parameter-free propagation strategy.
- An integrated network structure investigation algorithm is proposed.
- Provide a soft-partitioning solution for overlapping nodes.

ARTICLE INFO

Article history:
Received 27 September 2014
Received in revised form 22 October 2015
Available online 30 December 2015

Keywords:
Hierarchical and overlapping community
Hubs and outliers
Label propagation

ABSTRACT

The investigation of network structure has important significance to understand the functions of various complex networks. The communities with hierarchical and overlapping structures and the special nodes like hubs and outliers are all common structure features to the networks. Network structure investigation has attracted considerable research effort recently. However, existing studies have only partially explored the structure features. In this paper, a label propagation based integrated network structure investigation algorithm (LINSIA) is proposed. The main novelty here is that LINSIA can uncover hierarchical and overlapping communities, as well as hubs and outliers. Moreover, LINSIA can provide insight into the label propagation mechanism and propose a parameter-free solution that requires no prior knowledge. In addition, LINSIA can give out a soft-partitioning result and depict the degree of overlapping nodes belonging to each relevant community. The proposed algorithm is validated on various synthetic and real-world networks. Experimental results demonstrate that the algorithm outperforms several state-of-the-art methods.

1. Introduction

Complex networks provide a way to represent complex systems of interacting objects, where nodes denote the objects and edges describe the interactions between them. For example, in ecological systems, nodes represent lives and edges represent dependencies, and in protein association networks, nodes represent proteins and edges represent physical

* Corresponding author at: School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China.
E-mail address: wutaoadeny@gmail.com (T. Wu).

http://dx.doi.org/10.1016/j.physa.2015.12.073
0378-4371/

interactions. Common sense dictates that the network structure is characterized by at least three common features: communities, hubs and outliers. Network communities often correspond to groups of nodes that share a common property, role or function. The nodes within one community are more likely to interact to each other than to the rest of the network. Community structure allows for understanding the network functions that cannot be studied when considering only individual object or the entire system. Moreover, hubs are the nodes that bridge multiple communities, and hubs identification is crucial to network exploration. For example, hubs in epidemiology network could be the nodes that spread epidemics across groups, and immunizing the hubs could help to prevent the spread of epidemics. In addition, the identification of outliers that are marginally connected with the community members is also crucial to networks, as they can be used for abnormal nodes detection in complex networks. The network structure investigation is an important aspect of complex networks research.

In the issue of network structure investigation, early work mainly focused on the community detection in naive networks (non-hierarchical and non-overlapping structures). Many classical methods have been proposed to detect the community structure of complex networks, including division methods [1,2], agglomerative methods [3,4], matrix related methods (non-negative matrix factorization [5], spectral method [6]), model-based methods (label propagation methods [7,8], mixture models [9], stochastic block models [10,11]), etc. In a word, the traditional detection problem about the community structure in naive networks is likely to be properly solved in the last decade.

However, the problem of community detection becomes much harder for the networks with complex communities where hierarchical and overlapping structures emerge at the same time and tangle with each other. In real-world networks, communities are always nested, and the networks exhibit hierarchical community structure, such as the organization of a large company. Moreover, to use a social metaphor, common sense goes that people can belong to different social communities, depending on their friends, professions, hobbies, etc. In network terms, each node can be shared between communities, forming overlapping communities. Thus, the networks demand methods that are able to detect complex community with hierarchical and overlapping structures. So later work paid close attention to the community detection in hierarchical and overlapping cases, which generally includes two main genres: hierarchical community detection [12–14] and overlapping community detection [15–22], although several researchers made a pioneering attempt on hierarchical and overlapping community detection [23–25].

Based on the above discussion, most of the existing methods detect community by considering the hierarchical and overlapping structures separately, and there are few integrated algorithms for hierarchical and overlapping community detection. Meanwhile, in the issue of network structure investigation, there are few methods focusing on hubs and outliers identification, although hubs and outliers are important, common features of complex networks. Thus, network structure investigation should also take hubs and outliers identification under consideration. Therefore, how to discover the hierarchical and overlapping communities, hubs and outliers comprehensively and efficiently remains an important task to date.

To better capture the network structures, an intuitive idea is that one should analyze networks in a local view. Thus the computation of the community memberships and the roles of each node can be based only on local information, which can avoid the complex organizational structures with a variety of properties. In this paper, we develop a label propagation based integrated network structure investigation algorithm (LINSIA). The basic idea behind LINSIA is that there has a network topology related equilibrium state between the nodes' community label choices, and the reasonable structure division can be derived based on the label equilibrium state. LINSIA controls the label propagation process to discover multi-scale communities at different aggregation levels, and a community on a large scale corresponds to multiple underlying communities. The main contributions include: (1) LINSIA can reveal hierarchical and overlapping communities, as well as hubs and outliers; (2) LINSIA can provide insight into the label propagation mechanism and propose a parameter-free solution that requires no prior knowledge; (3) LINSIA can give out a soft-partitioning result and depict the degree of overlapping nodes belonging to each relevant community. Experimental results demonstrate that the algorithm outperforms several state-of-the-art methods.

The rest of the paper is organized as follows. Section 2 introduces some influence measures. Section 3 proposes a new label propagation strategy. Section 4 presents the LINSIA algorithm. Our algorithm is tested on a diverse set of networks in Section 5. Discussion and Conclusion are given in Section 6.

2. ENCoreness based influence measures

Before we proceed, it is worthwhile to introduce some necessary definitions for LINSIA. The node influence represents node importance in the full network, and the label influence measures the popularity of each community label in label voting process. The node influence and the label influence are all regulated by an adaptive variable α. In order to compute node importance accurately, node influence should consider node's global role and its local topology information comprehensively. Thus, it exploits extended neighborhood coreness (ENCoreness) centrality [26] to measure node's global importance and includes node degree to capture node's local topology.

Definition 2.1 (*Node Influence*)**.** Let $NI(i)$ indicate the influence of node i, and it is defined as

$$NI(i) = ENCoreness(i) + \sum_{j \in N(i)} \frac{ENCoreness(j)}{degree(j)^{\alpha}} * w_{i,j} \tag{1}$$