Physica A 388 (2009) 1706–1712

Contents lists available at ScienceDirect

Physica A

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

Detect overlapping and hierarchical community structure in networks

Huawei Shen [a,b], Xueqi Cheng [a,*], Kai Cai [a], Mao-Bin Hu [c]

[a] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, PR China
[b] Graduate University of Chinese Academy of Sciences, Beijing 100049, PR China
[c] School of Engineering Science, University of Science and Technology of China, Hefei 230026, PR China

ARTICLE INFO

Article history:
Received 3 September 2008
Received in revised form 8 December 2008
Available online 16 December 2008

PACS:
89.75.Hc
05.10.-a
87.23.Ge
89.20.Hh

Keywords:
Community detection
EAGLE
Modularity
Overlapping & hierarchical community
Maximal clique

ABSTRACT

Clustering and community structure is crucial for many network systems and the related dynamic processes. It has been shown that communities are usually overlapping and hierarchical. However, previous methods investigate these two properties of community structure separately. This paper proposes an algorithm (EAGLE) to detect both the overlapping and hierarchical properties of complex community structure together. This algorithm deals with the set of maximal cliques and adopts an agglomerative framework. The quality function of modularity is extended to evaluate the goodness of a cover. The examples of application to real world networks give excellent results.

1. Introduction

Many complex systems in nature and society can be described in terms of networks or graphs. Examples include the Internet, the world-wide-web, social and biological systems of various kinds, and many others [1–3]. In the past decade, the theory of complex networks has attracted much attention. Complex networks are usually characterized by several distinctive properties: power law degree distribution, short path length, clustering and community structure. The problem becomes important because a complex system's dynamics is actually determined by the interaction of many components and the topological properties of the network will affect the dynamics in a very fundamental way. Therefore, an efficient and sound approach that can capture the topological properties of network is needed.

Identifying the community structure is crucial to understand the structural and functional properties of the networks [4–6]. Many methods have been proposed to identify the community structure of complex networks [7–12]. One can refer to Ref. [13] for reviews. These methods can be roughly classified into two categories in terms of their results, i.e., to form a partition or a cover of the network.

The first kind of method produces a partition, i.e., each vertex belongs to one and only one community and is regarded as equally important. Different from the classical graph-partition problem, the number of communities and the size of each community are prior unknown. Newman et al. propose a quality function Q, namely *modularity*, to evaluate the goodness of a partition [9]. A high value of Q indicates a significant community structure. Several community detection methods have been proposed by optimizing modularity [11,14,15]. Generally, this kind of method is suitable for understanding the entire

* Corresponding author.
E-mail addresses: shenhuawei@software.ict.ac.cn (H. Shen), cxq@ict.ac.cn (X. Cheng).

0378-4371/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2008.12.021

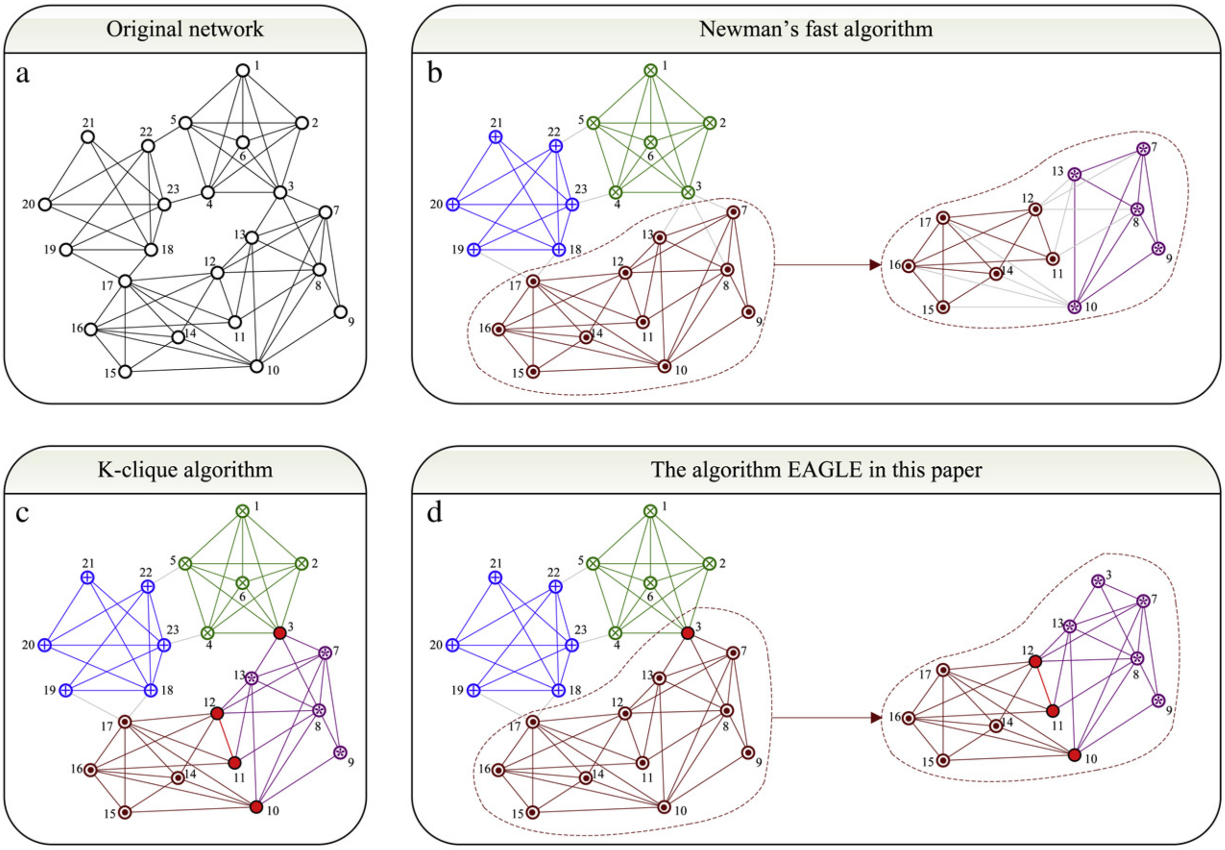

Fig. 1. Comparison of community structure found by different algorithms. Different communities are rendered in different colors. Edges between communities are colored in light gray. Overlapping regions between communities are emphasized in red. (a) The schematic network. (b) The hierarchical community structure found by Newman's fast algorithm. This algorithm is chosen as a representative of the first kind of algorithm. (c) The overlapping community structure found by the k-clique algorithm as a representative of the second kind of algorithm. (d) The overlapping and hierarchical community structure found by the algorithm EAGLE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structure of networks, especially for the networks with a small size. Recently, some authors [16,17] have pointed out that the optimization of modularity has a fundamental drawback, i.e., the existence of a resolution limit.

The second kind of method aims to discover the vertex sets (i.e., communities) with a high density of edges. In this case, overlapping is allowed, that is, some vertices may belong to more than one community. Meanwhile, some vertices may be neglected as subordinate vertices. Therefore, these methods result in an incomplete cover of the network. Numerous methods have been proposed, based on *k-clique* [8], *k-dense* [25] or other patterns. Unfortunately, there is no commonly accepted standard to evaluate the goodness of a cover up to now. Compared with the partition methods, this kind of method is appropriate to find the cohesive regions in large scale networks.

In real networks, communities are usually overlapping and hierarchical [8,18–20]. Overlapping means that some vertices may belong to more than one community. Hierarchical means that communities may be further divided into sub-communities. The two kinds of existing methods, as mentioned above, investigate these two phenomena separately. The first kind of method can be used to explore the hierarchical community structure, however, it is unable to deal with overlaps between communities. The second kind of method can uncover overlapping community structure of networks, but it is incapable of finding the hierarchy of communities. Recently, Lancichinetti et al. made a pioneering attempt on the detection of both overlapping and hierarchical community structures in complex networks [21]. They tried to detect the overlapped communities in the network based on the local optimization of a fitness function. Their method can uncover the hierarchical relation between these overlapped communities around a particular node. The remained problem lies in that the detection of the hierarchy of all overlapped communities in the network is not guaranteed due to the random choice of seed nodes.

In this paper, a new algorithm EAGLE (agglomerativE hierarchicAl clusterinG based on maximaL cliquE) is presented to uncover both the overlapping and hierarchical community structures of networks. This algorithm deals with the set of maximal cliques and adopts an agglomerative framework. The effectiveness is demonstrated by applications to two real world networks, namely the word association network and the scientific collaboration network.

In Fig. 1, we use a schematic network to illustrate what EAGLE can do and compare it with the two kinds of existing method. Fig. 1(a) depicts the schematic network. We construct this network according to the schematic network in Ref. [8],