Physica A 428 (2015) 279–294

Contents lists available at ScienceDirect

Physica A

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

Large-scale community detection based on node membership grade and sub-communities integration

Ronghua Shang [a,*], Shuang Luo [a], Yangyang Li [a], Licheng Jiao [a], Rustam Stolkin [b]

[a] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an, China

[b] School of Mechanical Engineering, University of Birmingham, UK

HIGHLIGHTS

- The proposed algorithm is to detect community structure in large-scale networks.
- It is based on node membership grade and sub-communities integration.
- It introduces two functions based on the local information of each node in networks.
- It employs the algorithm framework resembling label propagation.
- Experimental results indicate the accuracy and efficiency of the proposed algorithm.

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form 28 December 2014
Available online 14 February 2015

Keywords:
Large-scale network
Node membership function
Sub-communities integration
Overlapping community

ABSTRACT

Community detection plays an important role in research on network characteristics and in the mining of network information. A variety of algorithms have previously been proposed, but with the continuous growth of network scale, few of them can detect community structure efficiently. Additionally, most of these algorithms only consider non-overlapping community structures in networks. This paper addresses these problems by proposing a new algorithm, based on node membership grade and sub-communities integration, to detect community structure in large-scale networks. The proposed algorithm firstly introduces two functions based on the local information of each node in networks, namely neighboring inter-nodes membership function f_{MS-NN} and node-to-community membership function f_{MS-NC}. Firstly, local potential's complete sub-graphs are efficiently mined using the function f_{MS-NN}, and then these small graphs are merged into larger ones in light of local modularity. Secondly, incorrectly divided nodes are modified according to function f_{MS-NN}. Additionally, by adjusting the parameters in f_{MS-NC}, we can accurately obtain both non-overlapping communities and overlapping communities. Furthermore, the proposed algorithm employs a framework resembling label propagation, which has low time complexity and is suitable for detecting communities in large-scale networks. Experimental results on both artificial networks and real networks indicate the accuracy and efficiency of the proposed algorithm.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Complex networks are prevalent throughout the natural world, human interactions and computer systems, e.g. the World Wide Web, interpersonal networks, biological networks, and many other examples [1–6]. An important property that exists

* Corresponding author.
E-mail address: rhshang@mail.xidian.edu.cn (R. Shang).

http://dx.doi.org/10.1016/j.physa.2015.02.004
0378-4371/© 2015 Elsevier B.V. All rights reserved.

in many of these networks is community structure [7–9]. A community is a set of nodes that connect more closely with each other than they do with other nodes in different communities [8,9]. Individuals of the same community usually have common characteristics [10,11]. For example, web pages with similar subjects compose a community in the worldwide web network [12]. Additionally, it can be seen that individuals with similar characteristics often share more dense connections with each other than they do with other parts of the same network. Thus, detecting community structures is helpful in understanding the structure and functioning of networks [13] and can also help to detect potentially useful information within a network, through mining relations between individuals.

The problem of community detection is an area of rapidly growing interest within the complex network analysis research community, and a variety of methods have been proposed for its solution. Well-known approaches can be broadly categorized as graph partitioning methods [14–16], hierarchical clustering algorithms [17–19], and evolutionary algorithms [20,21].

Kernighan–Lin algorithm [14] is a well-known graph partitioning method, which works by randomly dividing a network into two communities, and then iteratively exchanging the nodes of the two communities until a modularity measure Q (see [22]) is maximized. Spectral bisection [15] also works by separating the network into two parts, using a Laplace matrix. Both of these algorithms rely on accurate prior knowledge of community size; they can only perform a simple binary division of a network into two communities; also their time complexity is high.

The hierarchical clustering algorithm is based on notions of similarity between the nodes and edge betweenness. This class of algorithms is "hierarchical" in the sense that clusters are recursively merged (agglomerative methods) or split (divisive methods) as one moves up or down the hierarchy respectively. For example, GN [17], proposed in 2002, recursively removes whichever residual edge has the largest edge betweenness, thereby progressively decomposing a network into a number of smaller clusters. However, computing the betweenness of all the edges is time-consuming.

In 2008, Clara Pizzuti [20] first proposed the use of evolutionary algorithms to solve the problem of community detection. The algorithm uses a single objective evolutionary algorithm to optimize community fraction CS as its objective function. Inspired by this method, Gong [23] et al. proposed a memetic algorithm to optimize modularity density D to extract multilevel community structures. In 2013, Shang et al. [24] improved on [23] by incorporating additional kinds of prior knowledge and using simulated annealing as a local search strategy to optimize a modularity measure, Q. In addition to the above methods, other algorithms have recently been proposed for detecting overlapping community structures. Some of these methods firstly extract maximal sub-graphs from the original networks and then merge small sub-graphs according to some index or strategy [25–27]. Other methods detect overlapping nodes in bipartite networks using key bi-communities and free-nodes [28].

However, with the rapid growth in worldwide computer prevalence and connectedness, the corresponding expansion of individual's social circles, and the era of big data, the scale of networks is increasing, engendering a growing need for algorithms which are fast and efficient. In light of this some of the above methods are no longer suitable for community detection in large scale networks, for example the time complexity of KL [14] is $O(n^2)$ and GN [17] is $O(n^3)$. Although evolutionary algorithms have shown potential for finding global optima, and are not constrained by the type of optimization function, they often take a long time to converge. Besides, the detection results still have some room for improvement, such as improving the detection precision and finding more multilevel solutions.

To overcome the limitations of the above algorithms, this paper proposes a large scale community detection algorithm based on node membership grade and sub-communities integration. First, we propose a neighboring inter-nodes membership function $\boldsymbol{f_{MS-NN}}$ to evaluate the closeness of each node with its neighbors. Through merging the couple-node with highest $\boldsymbol{f_{MS-NN}}$ value, this method can quickly find the potential complete sub-graph structures and effectively obtain a preliminary partitioning for the network. Next, those sub-communities achieved from the above steps are integrated by optimizing modularity Q. However, once these sub-communities have been merged together, it is difficult to correct nodes that have been wrongly placed. Therefore we propose another membership function $\boldsymbol{f_{MS-NC}}$ which is used to estimate the intimacy of nodes that connect with different adjacent communities and can modify misclassified nodes, thereby preventing the result from falling into local optima. Finally, through adjusting the parameters of $\boldsymbol{f_{MS-NC}}$, the proposed algorithm can be used to detect overlapping nodes and find overlapping community structures at different levels. Because the proposed algorithm adopts a learning strategy similar to label propagation, which involves only local information in each iteration, our method has low time complexity and is therefore suitably efficient for detecting communities in large and medium scale networks.

The remaining part of this paper is arranged as follows. In Section 2, related algorithms are introduced and the motivations for the proposed algorithm are explained. In Section 3, the details of the proposed algorithm are described. Section 4 presents the results of experiments on both artificial and real networks. Section 5 discusses the results and presents conclusions.

2. Related works and motivation

In this section, we will introduce some related strategies employed for community detection in large scale networks, and discuss the motivations for designing the new algorithm proposed in this paper.

Download English Version:

https://daneshyari.com/en/article/977570

Download Persian Version:

https://daneshyari.com/article/977570

Daneshyari.com