Physica A 427 (2015) 289–301

Contents lists available at ScienceDirect

Physica A

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

An ant colony based algorithm for overlapping community detection in complex networks

Xu Zhou [a,b], Yanheng Liu [a,b,c], Jindong Zhang [a,b,*], Tuming Liu [a,b], Di Zhang [a,b]

[a] College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China

[b] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, China

[c] State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130012, China

HIGHLIGHTS

- An ant colony based overlapping community detection algorithm is proposed.
- Redefine heuristic formula to measure similarity between two nodes.
- Label list stored in each node is updated under the guidance of movement of ants.
- Our algorithm performs better than other algorithms in the experiments.

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form 9 December 2014
Available online 11 February 2015

Keywords:
Ant colony algorithm
Overlapping community detection
Heuristic information
Complex networks

ABSTRACT

Community detection is of great importance to understand the structures and functions of networks. Overlap is a significant feature of networks and overlapping community detection has attracted an increasing attention. Many algorithms have been presented to detect overlapping communities. In this paper, we present an ant colony based overlapping community detection algorithm which mainly includes ants' location initialization, ants' movement and post processing phases. An ants' location initialization strategy is designed to identify initial location of ants and initialize label list stored in each node. During the ants' movement phase, the entire ants move according to the transition probability matrix, and a new heuristic information computation approach is redefined to measure similarity between two nodes. Every node keeps a label list through the cooperation made by ants until a termination criterion is reached. A post processing phase is executed on the label list to get final overlapping community structure naturally. We illustrate the capability of our algorithm by making experiments on both synthetic networks and real world networks. The results demonstrate that our algorithm will have better performance in finding overlapping communities and overlapping nodes in synthetic datasets and real world datasets comparing with state-of-the-art algorithms.

1. Introduction

A community is a group of vertices which probably share common properties and play similar roles within the graph representing real systems. Nodes in the same group have a large number of edges between themselves compared with

* Corresponding author.
E-mail address: jdzhang11@163.com (J. Zhang).

http://dx.doi.org/10.1016/j.physa.2015.02.020
0378-4371/

the number of edges with nodes belonging to other groups [1]. Detecting communities in networks provides insight into understanding the functional structures of networks.

Some methods aim at detecting standard partition, that is to say, each vertex in the partition is assigned to a single community [2]. However, vertices are often shared between communities in reality. For example, in social network, a person may have different roles and usually connects to several social groups including family, friends, colleagues or other interest. In a biological network, a node may have a variety of biological functions. Therefore, the study of overlapping community detection has attracted an increasing attention recently, and many algorithms have been presented. These approaches have a widely application in areas such as torrent delay network [3,4], recommendation system [5], and many more.

Although many methods for detecting overlapping community have been presented recently. Among these overlapping community detection algorithms, agent-based algorithms perform quite well due to the way to propagate labels. Community Overlap PRopagation Algorithm (COPRA) is an effective method in the class of agent-based algorithms. However, it updates the label of each node without considering the old information in the last iteration. In order to find a new way to update label of node more accurately, we propose an ant colony based algorithm (AntCBO) for overlapping community detection. Ant colony optimization algorithms (ACO) as a bio-inspired algorithm, has been successfully applied to solve many different problems such as the Job shop scheduling problem [6], feature selection problem [7] and disjoint community detection problem [8–10]. But to the best of our knowledge, it has not been used to solve overlapping community detection problem. Here our main contributions consist of the following three aspects. (1) Extending ACO to solve overlapping community detecting problem. (2) Redefining a heuristic formula to measure similarity between two nodes. (3) Proposing a strategy of updating the label list stored in each node under the guidance of movement of ants.

AntCBO is based on ant foraging model, and ants deposit pheromone on the edge to guide others towards promising areas. We identify ants' initial locations and initialize label list stored in each node based on proposed ants' location initialization strategy. In ants' movement phase, all the ants move simultaneously according to the transition matrix, moreover, ant will take node ID from current node to the next node. The label list stored in every node is the result of cooperation among ants' movement. A post processing executing on the label list is shown to find overlapping community structure naturally. We do both experiments on synthetic datasets and real world datasets. The experiments illustrate that our algorithm has been competitive with state-of-the-art algorithms.

The rest of the paper is organized as follows. Existing techniques for overlapping community detection are discussed in Section 2. A new method called ant colony based overlapping community detection algorithm is presented in Section 3. The experimental results on synthetic datasets and real world datasets with their corresponding analysis respectively are shown in Section 4. Finally, conclusion of the paper and future research directions are given in Section 5.

2. Related work

Uncovering community structure is one of the most important problems in the field of complex networks. Several methods are developed to aim at detecting disjoint partition. However, in real systems vertices often belong to more than one community. Forcing a node into a single community and not allowing for overlap could affect to detect the true underlying community structures [11]. Therefore, the issue of detecting overlapping communities has become quite popular in the last few years, and a number of algorithms have been designed to solve overlapping community detection problem. The main techniques to detect overlapping communities can be divided into 5 categories [12]. It includes clique percolation, local expansion and optimization, link partition, fuzzy detection and agent-based algorithms.

The main representative algorithm of clique percolation is Clique Percolation Method (CPM) proposed by Palla [13]. This algorithm is based on the concept that the internal edges of a community are likely to form cliques due to their high density, and a k-clique community is the largest connected subgraph obtained by the union of a k-clique and of all k-cliques which are connected to it. The limit of CPM is that it can only find big clique and hardly find small cliques. Palla has designed software called CFinder to implement CPM algorithm [14]. Farkas introduced an extension of CPM named CPMw which considers a subgraph intensity threshold for weighted network [15]. CPMw takes a newly k-cliques as a community only if it has larger intensity than a predetermined fixed threshold.

The classical algorithm for local expansion and optimization is Local Fitness Maximization (LFM) [16]. LFM develops a community from a random starting node until the fitness function is not maximized. This method detects both overlapping communities and the hierarchical structure, but it depends on a parameter that controls the size of formed groups. Lee proposed a Greedy Clique Expansion (GCE) algorithm to find the overlapping community structure [17]. It firstly finds some distinct cliques as seeds, and then expands these seeds by greedily optimizing a local fitness function to construct local community. The difference between LFM and GCE algorithm is the selection of initial seeds. Using cliques instead of node will make GCE have better ability in finding highly overlapping structures. Shen proposed a method agglomerativE hierarchiAl clusterinG based on maximaL cliquE (EAGLE) to detect overlapping and hierarchical community structure by using agglomerate framework to produce a dendrogram [18]. All the maximal cliques are taken as initial communities. Then two communities with maximal similarity are merged together. But the drawback of this algorithm is its expensive computation. Baumes proposed the Iterative Scan algorithm (IS) [19]. IS expands seeded small cores by adding or removing nodes until the local density function cannot be improved. The quality of detected communities depends on the quality of seeds. Lancichinetti et al. introduced Order Statistics Local Optimization Method (OSLOM). It is based on the optimization of fitness function [20], but OSLOM usually results in a number of outliers.