Physica A 453 (2016) 203–227

Contents lists available at ScienceDirect

Physica A

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

A multiobjective evolutionary algorithm to find community structures based on affinity propagation

Ronghua Shang [a,*], Shuang Luo [a], Weitong Zhang [a], Rustam Stolkin [b], Licheng Jiao [a]

[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

- An algorithm called APMOEA is presented for community detection.
- APMOEA takes affinity propagation to preliminarily divide the network.
- APMOEA selects nondominated solutions as its initial population.
- APMOEA finds solutions approximating to the true Pareto optimal front.
- APMOEA uses an elitist strategy to prevent the degeneration.

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form 18 December 2015
Available online 26 February 2016

Keywords:
Complex network
Community detection
Affinity propagation
Multiobjective evolutionary algorithm

ABSTRACT

Community detection plays an important role in reflecting and understanding the topological structure of complex networks, and can be used to help mine the potential information in networks. This paper presents a Multiobjective Evolutionary Algorithm based on Affinity Propagation (APMOEA) which improves the accuracy of community detection. Firstly, APMOEA takes the method of affinity propagation (AP) to initially divide the network. To accelerate its convergence, the multiobjective evolutionary algorithm selects nondominated solutions from the preliminary partitioning results as its initial population. Secondly, the multiobjective evolutionary algorithm finds solutions approximating the true Pareto optimal front through constantly selecting nondominated solutions from the population after crossover and mutation in iterations, which overcomes the tendency of data clustering methods to fall into local optima. Finally, APMOEA uses an elitist strategy, called "external archive", to prevent degeneration during the process of searching using the multiobjective evolutionary algorithm. According to this strategy, the preliminary partitioning results obtained by AP will be archived and participate in the final selection of Pareto-optimal solutions. Experiments on benchmark test data, including both computer-generated networks and eight real-world networks, show that the proposed algorithm achieves more accurate results and has faster convergence speed compared with seven other state-of-art algorithms.

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

http://dx.doi.org/10.1016/j.physa.2016.02.020

1. Introduction

Complex networks are used to represent the relationship between different individuals in a system, such as biological networks, social networks, information transmission networks and so on. In a network, a node represents an individual in a system, and the line connecting two nodes represents the relationship between these two individuals. Complex networks have some basic statistical properties, such as "small world effect [1]" and "scale-free [2]". Another important property is community structure [3–5]. Nodes in the same community are more closely connected with each other than with other nodes in different communities. Community detection is significant in revealing the inherent community structure and analyzing the function of complex networks.

In the past decades, numerous algorithms have been proposed since the significance of mining the community structure of complex networks was realized. Example methods include GN [4], proposed by Girvan and Newman, which is one of the most classical hierarchical clustering algorithms [4–6]. Taking Kernighan–Lin algorithm [7] and spectral bisection algorithm [8] as their representatives, graph partitioning methods try to divide the whole network into a few sub-graphs. Besides, some methods are based on the node similarity [9,10], while others are proposed to find overlapping communities in networks [11–13].

However, these algorithms still have much room for improvement in increasing the accuracy of detection results. Thus, this paper proposes an algorithm which combines a powerful data clustering method, Affinity Propagation (AP) [14], and an evolutionary algorithm to achieve higher detection results within a few iterations. Firstly, this paper employs efficient data clustering methods, such as AP, to preprocess the networks. First a data clustering method is used to obtain initial partition results within a few steps. Next, these results are screened and an evolutionary algorithm, which has the characteristic of global optimization, is used to further improve these results. We demonstrate that this combination of methods produces good results on both real and computer generated network benchmark data.

The remaining parts of this paper are arranged as follows. Section 2 describes related work. Section 3 explains the background and design of the APMOEA algorithm. Section 4 presents experimental results and analysis. Section 5 provides concluding remarks.

2. Related works

In recent years, methods based on the optimization of modularity Q have also been widely used [15–20], such as the Newman greedy algorithm [17], simulated annealing [15] and external optimization [19]. However, it has been proved that maximizing modularity Q is computationally intractable [21]. Because evolutionary algorithms have several good characteristics, there is increasing interest in combining them with other methods in solving this discrete optimization problem. Firstly, objective functions can be optimized by the evolutionary algorithm regardless of whether or not they are continuous. Secondly, the method of multipoint searching in evolutionary algorithm ensures a good global searching ability. Thirdly, evolutionary algorithms have strong combination ability with other algorithms. In 2008, Clara Pizzuti proposed an algorithm named GA-Net [22], which introduced a new objective function termed Community Score (*CS*) and employed an evolutionary algorithm as the optimization method. However, the *CS* criteria only measure the degree of intra-connections in communities without regarding that of inter-connections between communities. Consequently, in 2009 Clara Pizzuti proposed MOGA-Net [23], which is a multiobjective evolutionary algorithm employing the method of NSGA-II [24] and introduces the concept of Community Fitness (*CF*), which is complementary to *CS*, as the second objective function. Through maximizing these two objectives simultaneously, the algorithm obtained a set of solutions which revealed community structure at different hierarchical levels. In 2011, Gong et al. proposed a memetic algorithm (Meme-Net) to detect community structure in networks [25]. The algorithm employed modularity density D [26] as the optimization function and used a hill-climbing strategy as its search strategy. By adjusting the parameter in the objective function D, Meme-Net could find better partitions in networks at different resolutions. On that basis, Shang et al. proposed a method based on Modularity and Improved Genetic Algorithm (MIGA) in 2013 [27]. MIGA adopted modularity Q [3] instead of modularity density D as the objective function to simplify the algorithm and reduced its computational complexity. Meanwhile, using prior information in population initialization makes the algorithm more targeted and accurate in community detection. Moreover, MIGA took simulated annealing as the search algorithm, which has greatly improved the ability of local search. In 2012, Gong et al. put forward an algorithm named MOEA/D-Net [28] based on MOEA/D [29]. It decomposed a two-objective optimization problem into several scalar optimization sub-problems and optimized them simultaneously to obtain a set of solutions approximating the true Pareto-optimal front. Another algorithm, CCDECD [30], which first integrates Cooperative Co-evolution framework into Differential Evolution based Community Detection has been proposed to decompose a network into several smaller parts and optimize them respectively. A bias grouping scheme is employed as the pre-processing method to improve the accuracy of the partition results. Through combining these effective strategies, this method can achieve promising results on larger networks.

However, evolutionary algorithms have some drawbacks, such as low convergence speed, premature convergence and degradation. Aiming at solving some of these problems, this paper presents a Multiobjective Evolutionary Algorithm based on Affinity Propagation (APMOEA). Firstly, the proposed algorithm uses Affinity Propagation (AP) method for the preliminary partition of networks. AP algorithm [14] is a data clustering method proposed by Frey and Dueck in 2007. Through passing messages between data points until a high-quality set of exemplars and corresponding clusters finally emerges, the method