Physica A 392 (2013) 1215–1231

Contents lists available at SciVerse ScienceDirect

Physica A

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

Community detection based on modularity and an improved genetic algorithm

Ronghua Shang [*], Jing Bai, Licheng Jiao, Chao Jin

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, PR China

ARTICLE INFO

Article history:
Received 18 April 2012
Received in revised form 8 July 2012
Available online 16 November 2012

Keywords:
Complex network
Community detection
Prior information
Genetic algorithm

ABSTRACT

Complex networks are widely applied in every aspect of human society, and community detection is a research hotspot in complex networks. Many algorithms use modularity as the objective function, which can simplify the algorithm. In this paper, a community detection method based on modularity and an improved genetic algorithm (MIGA) is put forward. MIGA takes the modularity Q as the objective function, which can simplify the algorithm, and uses prior information (the number of community structures), which makes the algorithm more targeted and improves the stability and accuracy of community detection. Meanwhile, MIGA takes the simulated annealing method as the local search method, which can improve the ability of local search by adjusting the parameters. Compared with the state-of-art algorithms, simulation results on computer-generated and four real-world networks reflect the effectiveness of MIGA.

1. Introduction

With the development of computer technology, complex networks are widely applied in all walks of life. Community detection is one of the hot topics in the study of complex networks, and the notes of different properties and different types constitute the structure called community in the network. Communities are groups of nodes having dense intra-connections and sparse inter-connections, which is one of the characteristics of complex network [1]. Community detection is important for understanding network structures and analyzing network characteristics. Community structure analysis is widely used in biology, physics, computer graphics and sociology.

In order to accurately analyze the community structure in networks, many excellent community detection methods have been put forward. Pothen, Simon and Liou proposed the spectral bisection method, which is a method based on hierarchical clustering to detect community structure in networks [2]. Newman and Girvan proposed the GN algorithm, which is a kind of split method [3]. Newman put forward the modularity Q [4] and proposed a fast algorithm based on the GN algorithm, which is a kind of condensing algorithm. Newman also proposed a spectrum algorithm based on the concept of the modularity matrix [5]. Leicht and Newman extended the spectrum algorithm based on the modularity matrix to the directed network [6]. Rosvall and Bergstrom proposed a community detection method based on information theory. It is a method which takes the modularity of the network as a lossy compression of the network structure, and then the problem of community detection is converted into a foundation problem in information theory [7]. Raghavan, Albert, and Kumara put forward the LPA algorithm [8]. Palla, Dereny and Farkas proposed a group penetration algorithm, which is the first algorithm that can detect overlapping communities [9]. On that basis, Kumpula, Kivea and Kaski proposed the SCP algorithm [10] and Duch and Arenas put forward the EO algorithm [11]. Via an improved spectral method, Xie et al. proposed the detection of community structure in a network, which proved to be successful in clustering nodes [12]. For

---

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

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2012.11.003

huge real-world networks, Wu et al. proposed an efficient overlapping community detection method, which is better than other algorithms both in the quality of results and in computational performance [13]. Zhang et al. proposed a fuzzy analysis of community detection in complex networks, which can quickly produce the desired results [14]. Faqeeh et al. proposed a community detection method based on the ''clumpiness'' matrix, which gives accurate results for many computer-generated and real-world networks [15]. Pan et al. proposed a detecting community structure via node similarity, which is rather efficient in discovering the community structure in complex networks [16].

In recent years, researchers have gradually tended to use artificial intelligence technology to improve the accuracy of community detection. One of the fastest development methods in recent years is to optimize the modularity to find an ideal community structure [17,18]. One of the most widely used methods is the genetic algorithm (GA).

The GA is a famous adaptive heuristic search algorithm inspired by the evolutionary ideas of natural selection and genetics and pioneered by Holland at the University of Michigan [19]. The GA is designed to simulate processes in natural system necessary for evolution and to follow the principles first laid down by Charles Darwin of survival of the fittest. When a GA is used to solve a problem, first, a population of chromosomes or individuals should be maintained where each chromosome represents a potential solution to the problem. Second, each chromosome is evaluated to give some measure of its fitness. And then some chromosomes undergo stochastic transformations by means of genetic operators (crossover and mutation) to form new chromosomes called offspring, where crossover creates new chromosomes by combining parts from two chromosomes and mutation creates new chromosomes by making changes in a single chromosome. Then, a new population is formed by selecting the more fit chromosomes from the parent population and the offspring population. Further iterations are carried out until the stopping criterion is satisfied [20].

The GA has been widely studied and applied in many fields in the engineering world. Goldberg studied GAs in search, optimization and machine learning [20]. Koza proposed Genetic Programming, and has published two books [21]. McKinney proposed groundwater management models and obtained the solutions by using a GA [22]. Schwefel proposed Evolution Strategies for Evolution and Optimum Seeking [23]. Reeves used a GA to solve flowshop sequencing [24]. Yang used a GA to select such subsets and to achieve multicriteria optimization with the features [25]. Jones et al. used a genetic algorithm for flexible docking [26]. Deb et al. successfully applied a GA to multiobjective optimization problems [27]. Stoico et al. designed a genetic algorithm for the 1D electron gas [28]. Wu et al. modeled interaction networks in genetic algorithms and analyzed the scale-free properties of information flux networks [29]. In recent years, many researchers have proposed algorithms [30,31] based on the GA for community detection. The existing algorithms based on the GA for community detection have some advantages such as parallel search and some drawbacks such as slow convergence and low accuracy. Theory and experiments have shown that a GA can find the local optimal solution and can hardly find the global optimal solution.

Gong et al. have proposed a memetic algorithm (MA) for community detection in networks based on the GA [32]. In MAs, a meme is defined as the learning process capable of performing local refinements for an individual. MAs are named differently such as genetic local searchers and Lamarckian genetic algorithms. An MA for community detection [32] has been proposed to optimize the modularity density. The modularity density includes a tunable parameter. The algorithm uses a hill-climbing strategy for the local search procedure. From the optimization point of view, the MA has greatly improved the detection effect of the GA and has improved the accuracy rate of community detection. However, the modularity density D brings in a parameter λ [33]. Only by adjusting λ can one find relatively ideal detection results [32]. Meanwhile, the hill-climbing strategy is a simple greed research algorithm, which selects the optimal solution as the current solution from the adjoining solution space of the current solution each time. The algorithm will end when a local optimal solution is found. The hill-climbing strategy can be easily realized. Its main disadvantage is low efficiency, local optimum and not always searching the global optimum. The simulated annealing method can move around with a probability, namely the method will accept a worse solution than the current one with a probability. The algorithm will perhaps jump away from the local optimum and get the global optimum after moving several times. So this paper takes the simulated annealing method as the local search method.

In order to improve the precision ratio for community detection and the ability of handling community detection problems, this paper proposes a community detection method based on the modularity and an improved genetic algorithm (MIGA). First, the calculation of the modularity Q is fast and simple, which can well guide the performance of the results for community detection. So MIGA takes the modularity Q as the objective function which can simplify the algorithm. Second, MIGA uses prior information (the number of community structures) in the initialization, which makes the algorithm more targeted and improves the stability and accuracy of community detection. On the one hand, the number of community structures is known for most real-world networks. The algorithm can immediately use prior information. On the other hand, for some networks whose classes are unknown, we can get the general classes of the network by using the traditional GA or the other state-of-art algorithms. Then, we select the numbers around the number of obtained general classes. Last, because the simulated annealing method has the advantages of good local searching ability of the network and lower computational complexity, MIGA takes the simulated annealing method as the local search method which can improve the ability of local search by adjusting the parameters. MIGA detects some classic networks, such as computer-generated networks, Zachary's karate club [34], the dolphin social network [35], American college football [3] and books about US politics [5,36]. Simulation results show that MIGA can greatly improve the stability of the detection results and the accuracy of the final results.

The structure of this paper is as follows. Section 2 expounds the content of community detection based on the modularity and an improved genetic algorithm. Section 3 presents the simulation experiments and results analysis. Some classic networks are detected by MIGA, the MA, and the GA, respectively, and experimental results are analyzed. Section 4 presents the summary.