Computing
DOI 10.1007/s00607-014-0433-6

A hybrid artificial immune network for detecting communities in complex networks

Amir-Mohsen Karimi-Majd · Mohammad Fathian · Babak Amiri

Received: 14 February 2014 / Accepted: 29 October 2014

Abstract One of the challenging problems when studying complex networks is the detection of sub-structures, called communities. Network communities emerge as dense parts, while they may have a few relationships to each other. Indeed, communities are latent among a mass of nodes and edges in a sparse network. This characteristic makes the community detection process more difficult. Among community detection approaches, modularity maximization has attracted much attention in recent years. In this paper, modularity density (D value) has been employed to discover real community structures. Due to the inadequacy of previous mathematical models in finding the correct number of communities, this paper first formulates a mixed integer non-linear program to detect communities without any need of prior knowledge about their number. Moreover, the mathematical models often suffer from NP-Hardness. In order to overcome this limitation, a new hybrid artificial immune network (HAIN) has been proposed in this paper. HAIN aims to use a network's properties in an efficient way. To do so, this algorithm employs major components of the pure artificial immune network, hybridized with a well-known heuristic, to provide a powerful and parallel search mechanism. The combination of cloning and affinity maturation components, a strong local search routine, and the presence of network suppression and diversity are the main components. The experimental results on artificial and real-world complex networks illustrate that the proposed community detection algorithm provides a useful paradigm for robustly discovering community structures.

A.-M. Karimi-Majd · M. Fathian (✉)
Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran
e-mail: fathian@iust.ac.ir

B. Amiri
The University of Sydney, New South Wales, Australia

Published online: 12 November 2014

Keywords Complex network · Community detection · Mixed integer non-linear programming · Artificial immune network · Modularity-based maximization

Mathematics Subject Classification 91D30 · 90C27 · 68T20

1 Introduction

Recently, complex networks have been studied and developed in many areas of science. Internet, biological networks, power grids, social networks, food webs, and communication networks are well-known examples of complex networks. One of the main issues in modelling such networks is to extract hidden structures, called communities [18]. Detecting communities helps us to make sense of complex networks [9].

Communities consist of objects, called nodes, and their relationships, called edges. They emerge as dense parts in a network while they may have a few relationships to each other. Indeed, communities are latent among a mass of nodes and edges in a sparse network. This characteristic makes the community detection process more difficult. Moreover, we do not usually have any prior knowledge about the number of existent communities and their sizes (i.e., the number of nodes included).

In this paper, we focus on one of the well-known approaches for detecting communities, called modularity-based optimization, initiated from the Newman and Girvan (NG) method [44]. Their method is based on the maximization of a measure, called the modularity function Q, to distinguish the edges of a community from random by means of a heuristic search algorithm. In order to improve the NG method, some researchers proposed to optimize modularity function through a mathematical programming model [1,60,66]. However, solving mathematical models may fail to complete the optimization process for medium and large networks in proper time because of the NP-Hardness of those models, although they find the best solution for small ones. To tackle this issue, researchers proposed to employ different heuristic methods (e.g., [36,57,62]), and many metaheuristic algorithms (e.g., [3,21,23,27,33]).

The Q measure can work well if the respective network is truly modular [51] and communities do not significantly differ in terms of their sizes [31]. Unfortunately, no guarantee exists for network modularity and community uniformity; therefore, this measure may not provide reliable results [15,52]. Consequently, the abovementioned models inherit the stated limitations and drawbacks of the Q function.

Li et al. [34] have proposed another measure called modularity density or D measure in order to cope with this problem. D measure targets the average in-degree (i.e., the average number of edges coming to nodes), denoted by d_{in}, and the average out-degree (i.e., the average number of edges going out of nodes), denoted by d_{out}, of each potential community. This measure aims to find communities so that the maximum of d_{in} and minimum of d_{out} occurs. They also developed a mixed integer non-linear programming (MINLP) model for detecting communities based on the D measure. The number of communities must be known as prior information for their mathematical model. This limitation would be problematic in case of medium and large networks.

In this paper, first, a new mathematical model is proposed to tackle the problems so that automatically discovering the correct number of communities would be possible. Then, as a core contribution of this paper, a new metaheuristic is developed