Penguins Search Optimization Algorithm for Community Detection in Complex Networks

Mohamed Guendouz, GeCoDe Laboratory, Department of Computer Science, Tahar Moulay University of Saïda, Saïda, Algeria
Abdelmalek Amine, GeCoDe Laboratory, Department of Computer Science, Tahar Moulay University of Saïda, Saïda, Algeria
Reda Mohamed Hamou, GeCoDe Laboratory, Department of Computer Science, Tahar Moulay University of Saïda, Saïda, Algeria

ABSTRACT

In the last decade, the problem of community detection in complex networks has attracted the attention of many researchers in many domains, several methods and algorithms have been proposed to deal with this problem, many of them consider it as an optimization problem and various bio-inspired algorithms have been applied to solve it. In this work, the authors propose a new method for community detection in complex networks using the Penguins Search Optimization Algorithm (PeSOA), the authors use the modularity density evaluation measure as a function to maximize and they propose also to enhance the algorithm by using a new initialization strategy. The proposed algorithm has been tested on four popular real-world networks; experimental results compared with other known algorithms show the effectiveness of using this method for community detection in social networks.

KEYWORDS

Community Detection, Complex Networks, Natural Computing, Penguins Search Optimization Algorithm, Swarm Intelligence

1. INTRODUCTION

In the last decade, complex networks analysis and mining has attracted the attention of many researchers from different domains since a lot of real world problems can be handled as complex networks including problems in: Biology, Social, and Communication. Generally, a complex network is represented by a graph where nodes represent individuals and edges represent the relation between them, for example, individuals could be users of a social networking website, computers or web pages and relations could be respectively friendship, internet connection and URL links. Complex networks are found to group into strongly connected groups in general, these groups have a high density of within-group edges and a lower density of between-group edges – nodes have more connections with in-group nodes than out-group nodes. Detecting these groups also known as clusters, partitions or communities is called community detection which is a hot topic in the field of networks analysis.

Community detection is the process of detecting groups of similar individuals inside a network, which means that individuals who share a number of same characteristics between them are grouped

DOI: 10.4018/IJAMC.2018010101

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
in groups called communities. Community detection algorithms are often provided with a graph $G(V, E)$ where nodes $V$ represent individuals and edges $E$ represent relations between individual. Formally, the task of community detection is to find a set of communities $\{C_i\}_{i=1}^n$ in $G$ such that $\bigcup_{i=1}^n C_i \subseteq V$, community detection is considered as an NP-hard problem due to the fact that neither the size nor the number of communities is known. However, even with these difficulties several algorithms and methods have been proposed to solve this problem. One can classify these methods into different types (Fortunato, 2010); traditional methods can be classified into two groups: Graph Partitioning and Hierarchical Clustering. Graph Partitioning methods cluster a network into a predetermined number of communities, usually with equal size, these methods require the number and the size of communities before partitioning, the Kernighan-Lin algorithm proposed in (Suaris & Kedem, 1988) is one of the earliest methods for graph partitioning. Hierarchical Clustering methods cluster a network into groups of nodes based on their similarity which means that similar nodes are grouped into communities according to a similarity measure. Cosine similarity, Hamming distance and Jaccard (Singhal, 2001) index are frequently used measures. These methods are classified into two categories:

1. **Agglomerative Algorithms:** Initially each node represents a partition of its own, then partitions are successively merged until the desired network partition structure is obtained.

2. **Divisive Algorithms:** All nodes initially belong to one partition, and then the partition is divided into sub-partitions, which are successively divided into their own sub-partitions. This process continues until the desired network partition structure is obtained.

Hierarchical methods have the advantage of not requiring a predefined number and size of partitions. However, they do not provide a way to choose the better partition that represents the community structure of the network from those obtained by the procedure, a detailed classification of community detection algorithms can be found in (Fortunato, 2010). Other algorithms and methods will be discussed later in the next section.

In recent years, natural inspired algorithms have known a great interest by many researchers, several bio-inspired algorithms have been proposed and applied successfully to many real-world optimization problems. Bio-inspired algorithms can be classified into different categories according to their source of inspiration, in general one such algorithm can be inspired from: evolutionary phenomena, collective behavior of creatures (swarm intelligence techniques), physical rules, and human-related concepts.

Nature-inspired algorithms are known for their efficiency in solving optimization problems, they have been applied to many problems like Data Mining (Rana, Jasola & Kumar, 2011; Yeh, 2012), Information Retrieval (Alloui, Boussebough & Chaoui, 2015; Joshi & Srivastava, 2013), Engineering (Fox, Xiang & Lee, 2007). The Penguins Search Optimization Algorithm (PeSOA) is one of the latest proposed nature-inspired metaheuristic optimization algorithms; this algorithm inspired from the hunting phenomenon of penguins was originally proposed by Gheraibia et al. in (Gheraibia, 2013). The Penguins Search Optimization algorithm has been successfully applied to many real-world problems such as Finding Optimal Spaced Seeds (Gheraibia, Moussaoui, Djenouri, Kabir, Yin & Mazouzi, 2015), Vehicle Routing Problem (Ammi & Chikhi, 2016), Travelling Salesman Problem (Mzili, Bouzidi & Riffi, 2015) and recently Association Rules Mining (Gheraibia, Moussaoui, Djenouri, Kabir & Yin, 2016).

The effectiveness of the Penguins Search Optimization algorithm in solving these optimization problems has influenced the authors to propose in this paper a discrete variant of the original algorithm for solving the problem of community detection in complex networks, the essential aim of the authors
A Two-Phase Scheduling Method Combined to the Tabu Search for the DARP
[www.igi-global.com/article/a-two-phase-scheduling-method-combined-to-the-tabu-search-for-the-darp/114203?camid=4v1a](www.igi-global.com/article/a-two-phase-scheduling-method-combined-to-the-tabu-search-for-the-darp/114203?camid=4v1a)

The Analysis of Zero Inventory Drift Variants Based on Simple and General Order-Up-To Policies
[www.igi-global.com/chapter/analysis-zero-inventory-drift-variants/63819?camid=4v1a](www.igi-global.com/chapter/analysis-zero-inventory-drift-variants/63819?camid=4v1a)