Physica A 563 (2021) 125459

Contents lists available at ScienceDirect

Physica A

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

SAG Cluster: An unsupervised graph clustering based on collaborative similarity for community detection in complex networks

Smita Agrawal [a,*], Atul Patel [b]

[a] Faculty of CSE Department, Institute of Technology, Nirma University, Ahmedabad, India
[b] Faculty of Computer Science, Charotar University of Science and Technology, Changa, India

ARTICLE INFO

Article history:
Received 2 November 2019
Received in revised form 30 April 2020
Available online 17 October 2020

Keywords:
Community detection
Graph clustering
Complex network
K-medoid
Random walk
Collaborative similarity
Distance function

ABSTRACT

Many real-world social networks such as brain graph, protein structure, food web, transportation system, World Wide Web, online social networks exist in the form of a complex network. In such complex networks, pattern identification or community detection requires extra effort in which identifying community is a significant problem in various research areas. Most of the clustering methods on graphs predominantly emphasize on the topological structure without considering connectivity between vertices and not bearing in mind the vertex properties/attributes or similarity-based on indirectly connected vertices. A novel clustering algorithm SAG-Cluster with K-medoids framework presented for detecting communities using a collaborative similarity measure which considers attribute importance in case the pair of disconnected nodes. A novel path strategy using classic Basel problem for the indirectly connected node as well as balanced attribute similarity and distance function is proposed. On two real data sets, experimental results show the effectiveness of SAG-Cluster with the comparison of other relevant methods.

1. Introduction

The amount of data traveling across the internet today is not only massive in terms of size but is complex as well. Nowadays, organizations, institutions, railways, airlines, etc. all of them use the stack of data (i.e., bigdata), which are further used for creating various reports in order to ensure continuity regarding the services that they offer. Many real-world social networks exist in the form of bigdata [1,2]. Bigdata of complex networks include very large scale networks with structured, unstructured, or semi-structured data and a set of the graph [3,4]. Such a complex network is available in the form of a brain graph [5], protein structure [6–8] food web, transportation system [9], world wide web [10,11] etc.

Many scientific and commercial applications have complicated structures than sequential patterns and needs additional effort to analyze it and provides recommendations [12]. Such complicated structure is visualized in the form of trees, graphs, networks, and other complex structures. As a common data structure, the graph has become popular nowadays for bigdata analytics as well. For such graph structure, it is important to model appropriate structures and their interactions, which can be applied to various applications. For these applications, modeling and generating a graph based on their structure, attributes, weight, and direction (whether directed or undirected) becomes an important task.

* Corresponding author.
E-mail addresses: smita.agrawal@nirmauni.ac.in (S. Agrawal), atulpatel.mca@charusat.ac.in (A. Patel).

https://doi.org/10.1016/j.physa.2020.125459
0378-4371/