Accepted Manuscript

Title: A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks

Authors: Mohammad Sattari, Kamran Zamanifar

Please cite this article as: Mohammad Sattari, Kamran Zamanifar, A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks, Journal of Computational Science https://doi.org/10.1016/j.jocs.2018.01.004

A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks

Mohammad Sattari
Department of Computer Engineering, University of Isfahan, Isfahan, Iran
mohammadsattari9220@eng.ui.ac.ir

Kamran Zamanifar[*]
Department of Computer Engineering, University of Isfahan, Isfahan, Iran
zamanifar@eng.ui.ac.ir

Highlights

- We proposed a method for community detection more accurately in dynamic social network.
- We apply the cascade information diffusion model to prevent forming monster communities.
- Our method outperform the dynamic benchmark methods for community detection in dynamic social networks.

Abstract
One of the most important topics in social network analysis is community detection in dynamic social networks. A variety of approaches exists for detecting communities in dynamic social networks, among which the label propagation algorithm (LPA) is the well-known approach. This approach has made remarkable performance, but still has several problems. One of the difficulties of this approach is the new nodes added to the social network graph in the current snapshot has a very slight chance of creating new communities. In fact, these nodes fall under the influence of existing communities. This drawback decreases the accuracy of community detection in dynamic social networks. We propose a new method based on label propagation approach and the cascade information diffusion model in order to solve this difficulty. Here, the newly proposed method, Speaker Listener Propagation Algorithm Dynamic (SLPAD), Dominant Label Propagation Algorithm Evolutionary (DLPAE) and Intrinsic Longitudinal Community Detection (ILCD) on real and synthetic networks are implemented. The findings indicate that the modularity and Normalized Mutual Information (NMI) and also $F1_{AVG}$ of this proposed method is considerably higher than the earlier available methods in most datasets. Therefore, it can be concluded that the proposed method improves the accuracy of community detection in comparison with other available methods.

1. Introduction
In recent years, research on community detection has become a hot topic in social networks [1]. A community is a set of densely connected nodes that have weaker links with the rest of network [2]. Detection communities can be associated with clustering nodes based on the network topology while no information on the size and structure of communities is available [3]. Detecting communities can be both static and dynamic. In static mode, one snapshot is considered. In dynamic mode, more than one snapshot is considered. In fact, the dynamic mode is matched with more realistic results than static mode.

There are many approaches for detecting communities in social networks, among which label propagation approach (LPA) is a simple and time-efficient approach. In LPA, each node is a community with only one node [4]. Several methods based on LPA are proposed in order to detect communities in the static or dynamic networks such as SLPA [5], COPRA [6] and DLPAE [7]. These methods are fast and simple, while over-propagating the labels lead to the formation of monster communities. Formation of this kinds of communities decreases the accuracy of community detection based upon label propagation, especially in dynamic social networks.

For alleviating the limitations of LPA, this paper proposes a method based on the LPA approach and an information cascade model. We used this model based upon the observation that it often has been much more efficient within communities than among communities [8]. We apply the CID (Cascade Information Diffusion) model [9], which is one of the famous information diffusion models. In the CID model, each node has two states. The first state is S_0-state consisting of uninformed nodes, which tend to receive information. The second state is S_1-state consisting of the nodes, which try to cascade information diffusion to S_0-state neighbours. The proposed method has two main stages. The first principal stage consists of applying the CID model, and the second principal

[*]**Corresponding author:**
Kamran Zamanifar, Department of Computer Engineering, University of Isfahan, Isfahan , Iran
Email: zamanifar@eng.ui.ac.ir