Physica A 443 (2016) 70–85

Contents lists available at ScienceDirect

Physica A

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

Targeted revision: A learning-based approach for incremental community detection in dynamic networks

Jiaxing Shang [a,*], Lianchen Liu [a], Xin Li [b], Feng Xie [a], Cheng Wu [a]

[a] Department of Automation, Tsinghua University, Beijing 100084, PR China

[b] Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong Special Administrative Region

HIGHLIGHTS

- We propose a learning-based targeted revision (LBTR) approach for efficient incremental community detection.
- We provide mathematical analysis on how the vertex classifier can affect the community detection time complexity.
- Experiment results show that our approach can significantly reduce the running time while maintaining high community detection quality.
- To make our approach effective, one should increase the precision of the vertex classifier while keeping recall at a reasonable level.

ARTICLE INFO

Article history:
Received 1 April 2015
Received in revised form 12 May 2015
Available online 30 September 2015

Keywords:
Incremental community detection
Dynamic networks
Targeted revision
Computational complexity

ABSTRACT

Community detection is a fundamental task in network analysis. Applications on massive dynamic networks require more efficient solutions and lead to incremental community detection, which revises the community assignments of new or changed vertices during network updates. In this paper, we propose to use machine learning classifiers to predict the vertices that need to be inspected for community assignment revision. This learning-based targeted revision (LBTR) approach aims to improve community detection efficiency by filtering out the unchanged vertices from unnecessary processing. In this paper, we design features that can be used for efficient target classification and analyze the time complexity of our framework. We conduct experiments on two real-world datasets, which show our LBTR approach significantly reduces the computational time while keeping a high community detection quality. Furthermore, as compared with the benchmarks, we find our approach's performance is stable on both growing networks and networks with vertex/edge removals. Experiments suggest that one should increase the target classification precision while keeping recall at a reasonable level when implementing our proposed approach. The study provides a unique perspective in incremental community detection.

1. Introduction

Many real-world systems can be represented as networks, such as social networks [1], biological networks [2], citation networks [3], etc. Complex networks often exhibit a sense of community structure, where vertices form groups and have much denser connection within groups than between groups [4]. Community structure is a basic structural property of networks and can be used in various applications. For example, communities in protein interaction networks can be used to

* Corresponding author. Tel.: +86 15110099654.
E-mail address: shangjiaxing@gmail.com (J. Shang).

http://dx.doi.org/10.1016/j.physa.2015.09.072

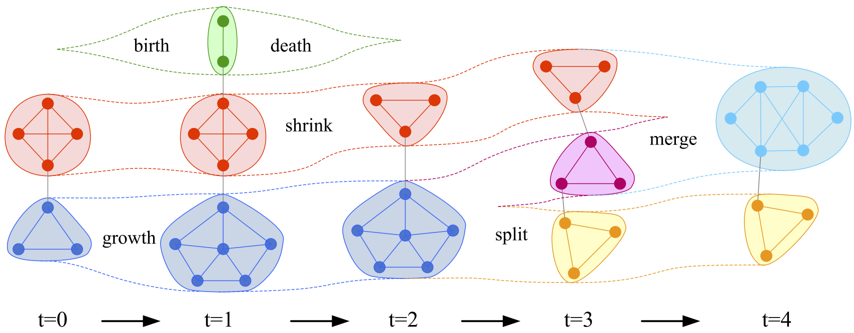

Fig. 1. Community evolution in a dynamic network.

predict the protein functions [5]. Communities in Webpage networks can be used for topic identification [6]. Previous works also studied the impact of community structure on epidemic spreading [7–9]. Community detection is a fundamental task in complex network analysis nowadays [4].

There are many community detection algorithms. A comprehensive review can be found in Ref. [10]. However, these algorithms generally consider networks to be static, while most real-world networks change over time. Recently, community detection in dynamic networks has attracted attention [11], which can help us understand how communities evolve [12–15], e.g., birth, death, growth, contracting (shrink), merging, splitting, etc., as shown in Fig. 1.

Community detection in dynamic networks can be addressed by applying static community detection algorithms multiple times on snapshots of the networks. However, it is more cost effective to incrementally revise the community structure of the old network when networks are updated [16–18], i.e., incremental community detection. Incremental community detection is more valuable when the networks are in mega-scale or change frequently, which is often the case in real networks.

Note that the essence of incremental community detection is the revision of community structure, if we can identify the high-risk vertices that need to be inspected, we argue it is possible to improve the efficiency of incremental algorithms. Holding this conjecture, in this paper, we propose a learning-based targeted revision (LBTR) approach. In this approach, we first classify the vertices that need to be revised. Then, we revise the community assignments of such vertices according to local modularity maximization [19]. In this paper, we provide mathematical analysis on how the classifier can affect the community detection time complexity. We conduct experiment on two real-world datasets to evaluate our approach, which shows that our proposed approach can significantly reduce the running time while maintaining community detection quality. Furthermore, as compared with the benchmarks, we find our approach's performance is stable on both growing networks and networks with vertex/edge removals. Experiments also suggest that, to make our approach effective, one should increase the precision of the vertex classifier while keeping recall at a reasonable level.

The rest of this paper is organized as follows. Section 2 reviews the literature on community detection. Section 3 elaborates the preliminaries of the problem to be addressed. Section 4 introduces our LBTR approach. Section 5 gives the evaluation framework, including the datasets, evaluation metrics, baseline methods, and experimental procedure. Experiment results are presented in Section 6. Section 7 concludes this paper.

2. Literature review

2.1. Community detection studies on static networks

There are many algorithms on community detection in static networks. Reader may refer to Ref. [10] for a comprehensive review. Here we only review a small number of papers that are more relevant to our paper.

Modularity-based algorithms are one major type of community detection algorithms, although they may have a resolution limit problem [20]. Modularity, first proposed by Newman et al. [21], is the fraction of connections within communities subtracts the expected fraction of connections within communities when vertices are randomly connected. Generally, a higher modularity value corresponds to a better community structure. Thus, the problem of community detection can be transformed to optimizing modularity. In Ref. [22] Clauset et al. proposed a greedy search algorithm (the CNM algorithm) to address this problem. It gradually merges communities that lead to the largest gain in modularity until convergence. In Ref. [23] Wakita et al. improved the time efficiency of the CNM algorithm by merging the communities in a more balanced way. To the best of our knowledge, the fastest modularity-based static algorithm is the Louvain algorithm [19]. The algorithm uses greedy strategies in a local manner. Initially, each vertex is put in a singleton community. Then the algorithm iteratively moves each vertex to its neighbor communities to maximize the gain in modularity. The procedure is then repeated at community level. Since moving a vertex to its neighbor community can be computed in $O(1)$ time, the algorithm is very efficient. The quality of the detected community structure also outperforms the CNM algorithm, as evaluated by the modularity measure.