Transformation-Based Community Detection from Social Networks

Sungsu Lim° and Jae-Gil Lee
Graduate School of Knowledge Service Engineering, KAIST
{ssungssu, jaegil}@kaist.ac.kr

1. Introduction

In recent years, *social network analysis* has become one of the most attractive issues in machine learning. In particular, *community detection* in networks is a fundamental problem in social network analysis [1]. It is the procedure of finding the community structure, with many edges joining vertices in the same community and relatively few edges joining vertices of different communities. A lot of theories, models, and methods have been actively developed for this purpose. However, due to a wide variety of network structures, there are still challenges of finding community structures from social networks. Among them, we focus on solving two important problems for community detection with various underlying structures: identifying the community structure of a graph when it consists of (i) overlapping community structure and (ii) highly-mixed community structure.

(i) *Overlapping community structure*: For the traditional community detection problem, we assume that an individual participates in at most one community. However, in many social networks, individuals can belong to multiple communities, e.g., friends, family, and colleagues [2]. That is, communities can overlap with each other. Developing an algorithm that offers good performance for overlapping community detection is a challenging problem.

(ii) *Highly-mixed community structure*: The mixing is defined as the fraction of edges that are between different communities. The higher the mixing is, the less detectable the community structure is. The accuracy of the existing algorithms degrades drastically when the mixing exceeds 0.5 [3]. It is important to develop an algorithm for community detection that is robust to high mixing since it occurs frequently as networks become more complicated.

2. Proposed Methods

To tackle the problems, we develop transformation-based community detection algorithms. Our key motivation is that transformation of a given network can give a better structure to identify community structure of an original network, as a kernel trick is used in machine learning. For each problem, our transformation technique converts a graph to a transformed graph that reflects the structure of the original graph and, at the same time, fits better the goal of the problem. We identify the community structure using the transformed graph, and the membership on the transformed graph is translated back to that on the original graph. We illustrate the overall framework in Figure 1.

For the first problem, we present the *link-space transformation*. This transformation converts a given graph to a link-space graph where each vertex of the link-space graph is a link of the original graph and the vertices are connected if the corresponding links in the original graph are connected. To preserve the original graph structure we adopt the link similarity to compute the weights of links of the link-space graph. This transformation provides a weighted graph that is easier to find communities, which

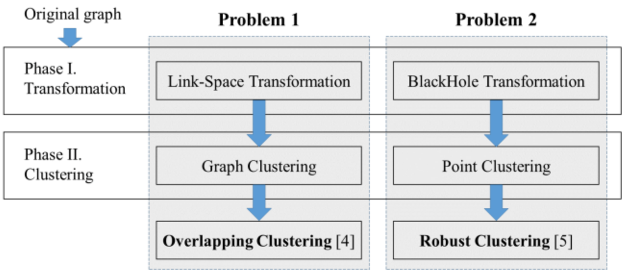

Fig. 1. Framework of the transformation-based community detection

corresponds to the link communities of the original graph and can be converted to the vertex communities whose memberships can overlap. Based on this procedure, we develop two overlapping community detection algorithms: LinkSCAN and LinkSCAN* [4]. The latter is an enhancement of the former for higher efficiency.

For the second problem, we propose a new community detection algorithm for undirected graphs, called BlackHole [5], by importing an embedding technique from graph drawing. The proposed algorithm BlackHole first maps the graph into a new space using the *BlackHole transformation* and then group the vertex positions by a conventional clustering algorithm. This transformation, in fact, is a merger of two widely-accepted building blocks: community detection and graph drawing. Our main contribution is to prove that a common idea in graph drawing, which is characterized by consideration of repulsive forces in addition to attractive forces, improves the clustering tendency of an embedding. As a result, our algorithm has the advantages of being robust especially when the community structure is not easily detectable.

Through extensive experiments, we have shown that LinkSCAN* and BlackHole achieve the accuracy higher than or comparable to the state-of-the-art algorithms for the problems (i) and (ii) respectively. Overall, we believe that our work provided a new framework for transformation-based community detection from social networks.

3. References

[1] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3, pp. 75-174, Feb. 2010.
[2] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, pp. 814–818, Jun. 2005.
[3] A. Lancichinetti and S. Fortunato, “Community detection algorithms: a comparative analysis,” *Physical Review E*, vol. 80, pp. 056117, Nov. 2009.
[4] S. Lim, S. Ryu, S. Kwon, K. Jung, and J.-G. Lee, “LinkSCAN*: Overlapping community detection using the link-space transformation,” in *Proc. IEEE ICDE*, pp. 292-303, 2014.
[5] S. Lim, J. Kim, and J.-G. Lee, “BlackHole: Robust community detection inspired by graph drawing,” in *Proc. IEEE ICDE*, to appear, 2016.