Description-Oriented Community Detection using Exhaustive Subgroup Discovery

Martin Atzmuellera, Stephan Doerfela, Folke Mitzlaffa

Summarised: January 9, 2020

1 Justification

This paper focuses on description-oriented community detection using sub-group discovery in order to provide structural and interpretable aspects to communities as sets of nodes. Because the while classic community detection, e. g., [17], for a survey, just identifies sub-groups of nodes with a dense structure, lacking an interpretable description.

The task of identifying communities as sets of nodes together with a description such a community pattern can provide an intuitive description of the community that usually is not achieved by old community mining methods that consider the nodes of a network (e. g., denoting users in a social network) as mere strings or ids.

For tackling this issue, is used a graph structure as well as additional descriptive features of the graph's nodes. Using additional descriptive features of the nodes contained in the network, the task of identifying communities as sets of nodes together with a description, i. e., a logical formula on the values of the nodes' descriptive features, such a community pattern, provides an intuitive description of the community, e. g., by an easily interpretable conjunction of attribute-value pairs.

The descriptive community pattern is built upon these features then describes and identifies a community. The identification of the communities is made according to standard community quality measures, while providing characteristic descriptions of these communities at the same time. Specifically, several optimistic estimates of standard community quality functions are used for efficient pruning of the search space in an exhaustive branch-and-bound algorithm.

1.1 Description-oriented comunity detection

We first introduce description-oriented community detectionandpresentthe COMODO algorithm for obtaining the k-best community patterns using a given community evaluation measure. COMODO is a branch-and-bound algorithm based on an exhaustive subgroup discovery approach.

1.2 Optimistic estimates

For fast description-oriented community detection using COMODO, we propose optimistic estimates [25, 62] which are efficient to compute. We consider a number of standard community quality functions: The segrega- tion index [19], the inverse average ODF (out degree fraction) [38], and the modularity [49]. We discuss the different measures for unweighted and weighted graphs, and extend the optimistic estimates accordingly.

1.3 Evaluation

We evaluate the presented approach using five data sets from three real- world social applications, i. e., from the social bookmarking systems Bib- Sonomy and delicious2, and from the social media platform last.fm3.

We present an algorithm for description-oriented community detection of the top-k communities (described by community patterns) with respect to a number of standard community evaluation functions. The method is based on an adapted subgroup discovery approach [10, 36], and also tackles typical problems that are not addressed by standard approaches for community detection such as patholog- ical cases like small community sizes. We focus on interpretable patterns that can easily be incorporated into a practical application, for example, for recommenda- tions in social bookmarking systems. It is important to note that we focus on static social graphs and do not take the dynamics into account since we aim to charac- terize a given community (allocation) for a given fixed interaction structure. Also, since in practice the entities in a network tend to belong to a number of different communities, the presented method naturally captures overlapping community al- locations. Moreover, in contrast to global approaches, we focus on the discovery of local communities. According to the idea of local pattern mining, e. g., [20], we do not try to find a complete (global) partitioning of the network. Instead, we consider a set of local, potentially overlapping communities. These should be as exceptional as possible with respect to a given community quality measure. We demonstrate our approach on several social media applications such as so- cial networking and social bookmarking systems that provide interaction networks like explicit friendship relations between users. However, the presented approach is not limited to such systems and can be applied to any kind of graph-structured data for which additional descriptive features (node labels) are available, e. g., cer- tain activity in telephone networks or interactions in face-to-face contacts [6] that also utilize tags or topic descriptions for the contained relations.

As an accompanying example, throughout the paper we use the friendship graph of the social bookmarking system BibSonomy1 [15]. In BibSonomy, users can declare their friendship toward other users, thus, creating a directed graph with users as nodes. At the same time, each user collects and tags resources like publications and web pages. Thus, a user's set of tags can be considered as a description of that user's interests. The community mining task here is to find user groups, where users are well connected by their friendship links and share a common interest in one or more features (tags). Overall, the contribution of this paper can be summarized as follows:

2 Main Points

3 Results

-