

Description-Oriented Community Detection using Exhaustive Subgroup Discovery

Martin Atzmueller, Stephan Doerfel, Folke Mitzlaff

Summarised by Laura Rodríguez Navas

January 9, 2020

The main contribution of the article

This article focuses on the description-oriented community detection, using exhaustive subgroup discovery to provide a structurally valid and interpretable description for communities (i. e., a set of nodes), with a graph's structures and descriptive features of a graph's nodes. Also, presents the COMODO algorithm that uses an efficient branch and bound method with appropriate pruning techniques to provide the exhaustive subgroup discovery. The article also shows an implementation of COMODO's algorithm that is evaluated by a selectable community quality measure using five real world datasets, and analysed the results.

As a personal opinion, I consider the approach of this article so interesting and introduces an innovation because can solve problems that in contrast to global approaches is usually not achieved by classical community mining methods, that consider the nodes of a network as mere strings or ids. Also the discovered algorithm it is very relevant since can easily be incorporated into practical applications and is not limited to any systems and can be applied to any kind of graph-structured data.

The article structures

The article has the structure that an article should have as a general scheme: the introduction, the methodology, the results of the investigation and the conclusion, as well as a list of references.

First introduce description-oriented community detection and present the COMODO algorithm. Then, summarizes basics of subgroup discovery, and provides general notions of graphs and community quality measures. Specifically for the topic of community quality measures, the article considers three standard community quality measures: The segregation index, the inverse average ODF (out degree fraction) and the local modularity. But is not explained in the introduction section. I think the community quality measures are not well described. It seems to be an optional process, which would be applied after the

COMODO algorithm and yet the experiments are based on these standard community quality measures. In my opinion, it should be described in the related work section. Because the related work section contains the main information of this article in grater detail.

The remainder of the article contains the related work and provides experiments using five data sets and discusses their results in the context of the three real-world applications. Finally, concludes the article with a summary and directions for future research.

Besides from what I have commented previously on the community quality measures, I think it is a well structured article. Although it is a bit long and complicated. Complicated since I have found it is difficult to read, specially the part of the experiments. Because of the number of examples and their simultaneous comparison. I think that with fewer examples it would be easier to understand and clearer.

Table of contents for a presentation

Quantes slides. Quin seria l'índex de continguts. Per a cada slide que hi posaria.