

Studies on Efficient Clustering Algorithms for Signed Graphs

孫, 舒蓉

<https://hdl.handle.net/2324/7182290>

出版情報 : Kyushu University, 2023, 博士（経済学）, 課程博士
バージョン :
権利関係 :



氏 名 : Shurong Sun

Name

論 文 名 : Studies on Efficient Clustering Algorithms for Signed Graphs

Title. (符号付きグラフに対する効率的なクラスタリングアルゴリズムに関する研究)

区 分 : 甲

Category

論 文 内 容 の 要 旨

Thesis Summary

The rise of digital networks has sparked a boom in the study and analysis of complex systems. Social network analysis is a building block of the field and has become an indispensable tool for understanding complex patterns of relationships and interactions within social structures. Community structure is an important feature of complex networks, indicating that nodes are gathered into several groups. Community detection, also called graph partition, is a worldwide task in network analysis, where the goal is to find groups of nodes that are, in some sense, more similar to each other than the other nodes. By analyzing the results, it is possible to reveal hidden relations among the nodes in the network. Clustering in graphs has become an essential technique for studying complex network systems and has garnered widespread attention and applications in various real-world scenarios. With the increasing richness of social media interactions and the additional information about the types of interactions present in today's social networks, researchers have turned to richer interpretations of edges in graph networks and have recently turned their attention to using signed graphs. A signed graph is a graph in which each edge has a positive or negative sign. Many real-world relationships can be represented by signed graphs with positive and negative links such as like/dislike, trust/distrust, or friends/enemies. Unlike traditional graphs, where edges represent only connections, signed graphs with positive or negative weighted edges usually contain more information and status. The presence of positive and negative links brings unique challenges and implications to signed graph clustering. Various clustering algorithms for signed graphs have been proposed in recent years. Inspired by these state-of-the-art methods, studies on efficient clustering algorithms for signed graphs are conducted in this thesis. This thesis comprises six chapters.

Chapter 1 begins by providing definitions of several types of graphs along with related mathematical notations. This is followed by an introduction to the background of graph clustering, highlighting how graph clustering has garnered significant attention in various fields. The chapter then introduces the background of signed graph clustering. With the rise of digital networks, researchers have increasingly focused on using signed graph networks to represent complex interactions on social media platforms.

Chapter 2 begins with an overview of essential preliminaries, foundational concepts, and relevant prior works. The chapter begins with two fundamental theories important to signed graphs: balance theory and status theory. The conventional spectral clustering algorithm is introduced with examples. Subsequently, spectral clustering algorithms for signed graphs are discussed. In the presentation of related work, the section 2.3.1 begins with an introduction to unsigned graph clustering. This is followed by an exploration of

signed graph analysis and a review of further research on signed graph spectral clustering. The basic concept and theory of k -core decomposition are then clarified, highlighting its related works. Finally, the section 2.3.5 discusses the basic concept and definition of network motif, introducing motif-based graph clustering.

Chapter 3 analyzes the existing clustering algorithms on synthetic graphs generated by the Signed Stochastic Block Model (SSBM). SSBM, a generative model for random graphs with community structures, is often utilized as a benchmark in network analysis and community detection algorithms. The chapter begins with an introduction to the Signed Stochastic Block Model, followed by a description of the accuracy evaluation index used in section 3.2, the Adjusted Rand Index (ARI). Subsequently, existing algorithms for numerical experiments are introduced. As the adjacency matrix, the symmetric signed normalized Laplacian matrix, and the signed normalized cut were already introduced in Chapter 2 when introducing the traditional spectral clustering algorithm, and this chapter provides a detailed introduction to the balanced normalized cut and the signed positive over negative generalized eigenproblem. The ARI values of the synthetic networks for each algorithm under SSBM with different parameters are presented, along with the corresponding computation times. The chapter concludes with an analysis of the results of numerical experiments under various conditions.

In Chapter 4, an efficient clustering algorithm for signed graphs is proposed. Observations from the previous chapter indicate that existing clustering algorithms face challenges when applied to complex real-world networks, which has been a significant motivation for this research. Recent studies on the higher-order organization of complex networks, particularly at the level of small network subgraphs, have provided substantial inspiration. This method is extended to signed graph clustering. The chapter then introduces the definition and related mathematical notations of signed motifs. Following this, the details of the proposed algorithm are presented. The process begins with identifying signed motifs in the signed graph, followed by conducting the k -core decomposition algorithm on the graph. Next, the signed motif-based clustering algorithm is implemented on the top k -core subgraph. Subsequently, the remaining nodes in the rest of the $k-1$ -core subgraphs are grouped into the clusters identified in the previous step. The chapter concludes with the presentation of a pseudocode for the proposed algorithm.

Chapter 5 implements numerical experiments on real-world datasets and evaluates the performance of the algorithm proposed in Chapter 4. The chapter starts with a description of the datasets used and includes example plots for several small-scale networks. In the absence of ground truth, the modularity value is employed as a metric to assess the accuracy of algorithms. The chapter then compares the proposed algorithm with the previously introduced algorithms, focusing on accuracy (modularity) and execution time. Based on these results, a detailed analysis is presented. The numerical experiments indicate that while the performance on relatively small-scale real-world networks is comparable to other algorithms, the proposed algorithm performs well in some larger-scale real-world networks. It outperforms others in efficacy and exhibits lower computation time in most scenarios.

Chapter 6 provides a comprehensive summary of this thesis's main findings and conclusions.