Decomposition of Social Networks into Relaxed Cliques

Fabio Furini[1], Timo Gschwind[2], Stefan Irnich[2], Roberto Wolfler Calvo[3]
[1] LAMSADE, Université Paris Dauphine, 75016 Paris, France
fabio.furini@dauphine.fr
[2] Gutenberg School of Management and Economics, Johannes Gutenberg University Mainz, 55128 Mainz, Germany
{gschwind,irnich}@uni-mainz.de
[3] LIPN, Université Paris 13, 93430 Villetaneuse, France
roberto.wolfler@lipn.univ-paris13.fr

Mots-clés : *Graph decomposition, community detection, clique relaxations, social network analysis, branch-and-price.*

1 Introduction

In social network analysis (SNA), relationships between members of a network are encoded in an undirected graph where vertices represent the members of the network and edges indicate the existence of a relationship. One important task in SNA is community detection, that is, clustering the members into communities such that relatively few edges are in the cutsets, but relatively many are internal edges. The clustering is intended to reveal hidden or reproduce known features of the network, while the structure of communities is arbitrary. We propose decomposing a graph into the minimum number of relaxed cliques as a new method for community detection especially conceived for cases in which the internal structure of the community is important. Cliques, that is, subgraphs with pairwise connected vertices, can model perfectly cohesive communities, but often they are overly restrictive because many real communities form dense, but not complete subgraphs. Therefore, different variants of relaxed cliques have been defined in terms of vertex degree and distance, edge density, and connectivity. They allow to impose application-specific constraints a community has to fulfill such as familiarity and reachability among members and robustness of the communities. Standard compact formulations fail in finding optimal solutions even for small instances of such decomposition problems. Hence, we develop exact algorithms based on a Dantzig-Wolfe reformulation and branch-and-price techniques. Extensive computational results demonstrate the effectiveness of all components of the algorithms and the validity of our approach when applied to social network instances from the literature.

2 Contributions and perspectives

In this work, we introduce the problem of decomposing a graph into a minimum number of relaxed cliques as a new method for community detection. While in prior works the resulting clusters generally do not have any structure, the different clique relaxations allow to impose application-specific constraints a cluster has to fulfill. Using the eight types of first-order clique relaxations, we identify new relevant types of decompositions with first-order relaxed cliques. In particular, for non-hereditary relaxed cliques one must distinguish between partitioning and covering the network. Moreover, since a basic requirement for communities is connectivity, we introduce the concept of connected relaxed cliques. As a consequence, decomposing into connected or general relaxed cliques gives rise to different problem variants. Our type of approach is useful in cases where one has a good understanding of what defines a community. For three prominent examples from social network analysis, we demonstrate that decomposition into relaxed cliques reproduces some known features of the network.

From an optimization point of view, decomposing into relaxed cliques is a hard problem. Our exact solution approach is based on branch-and-price, where pricing requires the design of new effective algorithms and branching the development of complete and preferably structure-preserving branching rules. For pricing, we propose a new CB&B for s-club, a modified version of RDS for hereditary relaxed cliques that is able to handle connectivity and negative weights, and new MIP models for s-bundle and k-block. For branching, a comparison of different branching schemes revealed that Ryan-Foster branching is superior although it is not structure-preserving for the pricing problem.

The contributions of this work are the following :

- the formal introduction of various Relaxed Cliques partitioning and covering problems as a new approach for community detection ;
- the presentation of a generic compact formulation of these decomposition problems ;
- the development of branch-and-price algorithms for their exact solution ;
- the introduction of connectivity conditions that each Relaxed Cliques has to respect and the discussion of their implications on pricing algorithms and branching ;
- two new mathematical programming formulations for finding k-blocks and s-bundles, i.e., relaxed cliques defined on the basis of vertex-connectivity ;
- new combinatorial branch-and-bound algorithms for finding maximum-weight s-clubs and connected hereditary relaxed cliques able to cope with general weights ;
- the presentation of branching rules that are structure-preserving in the sense that the pricing problem remains structurally unchanged ;
- the presentation of a comprehensive computational study including the application of the new models and algorithms for detecting communities in some real-world social networks that are intensively studied in the SNA and community-detection literature.

Références

[1] Pattillo, J., N. Youssef, S. Butenko. *On clique relaxation models in network analysis.* European Journal of Operational Research, 226 9-18, 2013.

[2] Trukhanov, S., C. Balasubramaniam, B. Balasundaram, S. Butenko. *Algorithms for detecting optimal hereditary structures in graphs, with application to clique relaxations.* Computational Optimization and Applications, 56, 113-130, 2013.

[3] Evans, T. S. Clique graphs and overlapping communities. Journal of Statistical Mechanics : Theory and Experiment, 2010.

[4] Veremyev, A., V. Boginski. *Identifying large robust network clusters via new compact formulations of maximum k-club problems.* European Journal of Operational Research, 218 316-326. 2012.

[5] Fortunato S. *Community detection in graphs.* Physics Reports, 486, 75–174, 2010.

[6] Aloise, D., S. Cafieri, G. Caporossi, P. Hansen, S. Perron, L. Liberti. *Column generation algorithms for exact modularity maximization in networks*, Physical Review E, 82, 46-112, 2010.