2-Edge Connected Balanced Subgraphs for Correlation Clustering Problem

Nejat Arinik, Rosa Figueiredo, Vincent Labatut

2-Edge Connected Balanced Subgraphs for Correlation Clustering Problem

Nejat Arınık[1], Rosa Figueiredo[1], Vincent Labatut[1]

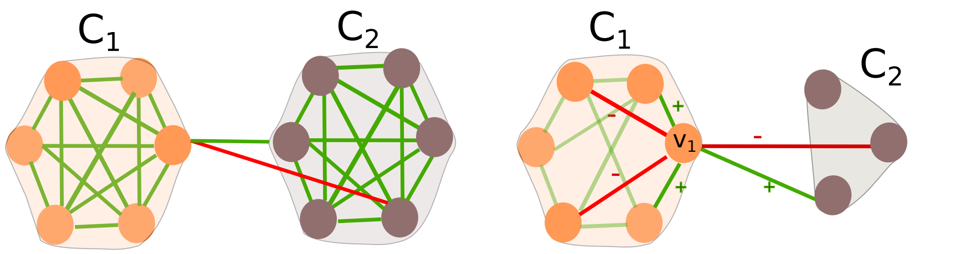
Laboratoire Informatique d'Avignon, France
{firstname.lastname}@univ-avignon.fr

Mots-clés : *Correlation Clustering, Structural Balance, Subgraph Connectivity.*

1 Introduction

A signed graph, whose links are labeled as positive (+) and negative (−), is considered *structurally balanced* [3] if it can be partitioned into a number of clusters, such that *positive* (*negative*) links are located *inside* (*in-between*) the clusters. Due to the imbalanced nature of real-world networks, various measures have been defined to quantify the amount of imbalance. Such measures are expressed relatively to a graph partition, so processing the graph balance amounts to identifying the partition corresponding to the lowest imbalance measure. A well-known measure among them corresponds to the definition of the *Correlation Clustering (CC) problem*, and it consists in counting the numbers of misplaced (w.r.t. structural balance definition) links [3].

One issue of the CC problem is that it is solely based on misplaced links, and it may not always reflect the real situations. One extreme but plausible case is the network instance of two positive groups where they are connected by the same number of positive and negative links, and this produces two possible optimal solutions. We see such an example in Figure 1a. Two positively well connected groups are tied through one positive and one negative links, resulting in two optimal partitions : 1) 2 clusters (C_1 and C_2), 2) a single cluster. Although the first partition makes more sense, the CC problem does not show any preference towards this partition. A similar situation can also be imagined at the node level, as illustrated in Figure 1b. Node v_1 is connected to C_1 and C_2 by the same proportion of positive and negative links, nonetheless, it has more positive links towards C_1. Likewise, this produces two optimal partitions.



(a) Two optimal solutions : 1) a single cluster, 2) clusters C_1 and C_2

(b) Two optimal solutions : 1) $v_1 \in C_1$, 2) $v_1 \in C_2$

FIG. 1 – Illustrative examples. Positive (resp. negative) links are represented by green (resp. red) color. In Figure 1b, positive links located inside C_1 are shown transparently to highlight the situation of node v_1.

To overcome these situations and those being in similar fashion one needs to extend the CC problem by adding additional topological constraints. The topology of the network is an important aspect in graph theory, and it is nearly always present in practical situations (e.g. in social dynamics). Indeed, topology constitutes the base of the problems in social networks (e.g. community detection), since they rely solely on this type of criteria. In this work, we are inspired by the studies ensuring network robustness (e.g. single failure in telecommunication, robustness to the damage by mutation or viral infection in biological networks), and we propose to extend the CC problem with 2 positive edge connectivity requirement at cluster level. Hence, the network topology provides for at least two diverse positive paths between each pair of nodes in the same cluster. In literature, although 2-edge connectivity constraint is well studied (for non-signed networks), to the best of our knowledge, the only work considering 2-edge connectivity constraint for the CC problem is on planar graphs (i.e. graph that can be drawn on the plane without intersection of its edges) [4], but not for the general case.

2 Method

We propose three different ILP formulations to solve the problem exactly. These are two flow-based and one cut-based formulations. Compared to the other formulations, the cut-based one has an exponential number of inequalities to guarantee the bi-connectivity of the clusters. Therefore, we need to develop a procedure that allows to separate unfeasible solutions in an efficient way. Moreover, we develop a branch-and-cut algorithm based on the appropriate state-of-the-art valid inequalities, specifically those for signed networks. Indeed, the fact that the graph is signed encodes an additional information of negative links, and more efficient cutting strategies can be designed. For instance, Miyauchi and Sukegawa [5] show that a portion of the transitivity constraints are redundant. Finally, we discuss the computational results.

3 Application

We apply our method to a subset of the 7$^{\text{th}}$ term European Parliament dataset presented in [1, 2]. Our goal with bi-connectivity constraints at cluster level is to detect groups of deputies with stronger cohesion thanks to our proposed network topology. We also compare our results with those in [1, 2] and interpret them.

Références

[1] N. Arinik, R. Figueiredo, and V. Labatut. Signed graph analysis for the interpretation of voting behavior. In *International Conference on Knowledge Technologies and Data-driven Business - International Workshop on Social Network Analysis and Digital Humanities*, Graz, AT, 2017.

[2] N. Arinik, R. Figueiredo, and V. Labatut. Multiple partit. of multiplex signed networks. *Soc Netw*, 2019.

[3] J. A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20(2) :181–187, 1967.

[4] Philip N. Klein, Claire Mathieu, and Hang Zhou. Correlation Clustering and Two-edge-connected Augmentation for Planar Graphs. In *32nd International Symposium on Theoretical Aspects of Computer Science (STACS 2015)*, volume 30 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 554–567, 2015.

[5] Atsushi Miyauchi and Noriyoshi Sukegawa. Redundant constraints in the standard formulation for the clique partitioning problem. *Optimization Letters*, 9(1) :199–207, 2015.