Discrete Applied Mathematics 160 (2012) 2573–2590

Contents lists available at SciVerse ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

A polyhedral study of the maximum edge subgraph problem

Flavia Bonomo [a,*], Javier Marenco [b], Daniela Saban [c,b], Nicolás E. Stier-Moses [d]

[a] *Depto. de Computación, FCEyN, Universidad de Buenos Aires and IMAS–CONICET, Argentina*
[b] *Instituto de Ciencias, Universidad Nacional de General Sarmiento, Argentina*
[c] *Depto. de Computación, FCEyN, Universidad de Buenos Aires, Argentina*
[d] *Graduate School of Business, Columbia University, USA*

ARTICLE INFO

Article history:
Received 16 April 2010
Received in revised form 25 July 2011
Accepted 8 October 2011
Available online 4 November 2011

Keywords:
Polyhedral combinatorics
Quasi-cliques
Maximum edge subgraph problem

ABSTRACT

The study of cohesive subgroups is an important aspect of social network analysis. Cohesive subgroups are studied using different relaxations of the notion of clique in a graph. For instance, given a graph and an integer k, the *maximum edge subgraph problem* consists of finding a k-vertex subset such that the number of edges within the subset is maximum. This work proposes a polyhedral approach for this NP-hard problem. We study the polytope associated to an integer programming formulation of the problem, present several families of facet-inducing valid inequalities, and discuss the separation problem associated to these families. Finally, we implement a branch and cut algorithm for this problem. This computational study illustrates the effectiveness of the classes of inequalities presented in this work.

1. Introduction

Social network analysis is an important tool to study the relationships and flows between people, organizations, and other entities. Social networks are encoded by graphs with vertices representing the entities, and edges representing interdependencies between them. An important aspect of social network analysis is the detection of *cohesive subgroups*, which are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties [26]. Although one could potentially define such a subgroup with the concept of clique, this does not provide a full picture because in practice missing edges frequently exist but that is not a strong enough reason to claim that two vertices cannot belong to the same group. For example, if two individuals in a social network are not friends because they fought but share most of their friends, they would belong to the same social group. For this reason, it is more satisfactory to define a cohesive subgroup using a relaxation of the definition of clique.

Let $G = (V, E)$ be the graph that represents the social network of interest. One possible approach to study cohesive subgroups is the use of *quasi-cliques*, which are subgraphs with a pre-specified minimum edge density. Here, the density of a subgraph is the quotient between the number of edges with both endpoints in the subgroup and the total number of edges. Formally, given a real number $0 \leq \gamma \leq 1$, a subgraph $G' = (N', E') \subseteq G$ is a γ-quasi-clique if $2|E'|/(|N'|(|N'| - 1)) \geq \gamma$. In the example of the social network of friends referred to earlier, missing relationships would represent that pairs of individuals in the group may not be friends with each other. In earlier work, we strongly relied on quasi-cliques to study the network of bilateral investment treaties signed between countries of the world [24]. Besides having used them to identify cohesive subgroups, they also allowed us to compare different random topologies. Indeed, cliques are inexact because there

* Corresponding author.
E-mail addresses: fbonomo@dc.uba.ar (F. Bonomo), jmarenco@ungs.edu.ar (J. Marenco), dsaban@dc.uba.ar (D. Saban), stier@gsb.columbia.edu (N.E. Stier-Moses).

0166-218X/$ – see front matter
doi:10.1016/j.dam.2011.10.011

is no guarantee that randomly generated edges will form the same cliques as in the original graph. Quasi-cliques, instead, can capture the topological structure regardless of which particular random edges are present in the network.

Since quasi-cliques are characterized by their size and density, they have been studied by considering two formulations that are dual of each other, much like other bi-criteria optimization problems. One can either be given (i) a specified edge density $\gamma \in [0, 1]$, and find the largest γ-quasi-clique, or be given (ii) a size k of a subgraph, and find the densest set of k vertices. The second approach – the one we follow in this paper – is known in the graph and optimization literature as the *maximum edge subgraph problem* (MESP) [4] or *dense/densest/heaviest k-subgraph problem* [12] or *k-cluster problem* [8]. Concretely, given an integer $k < |V|$, the MESP consists of finding a vertex subset $A \subset V$ with $|A| = k$ such that $|E(A)|$ is maximum, where $E(A) = \{ij \in E : \{i, j\} \subset A\}$.

The maximum clique problem reduces to the MESP, hence the latter is NP-hard [3]. Indeed, if the MESP can be solved in polynomial time, then we can polynomially search the greatest k for which the answer to the MESP is a complete subgraph, thus solving the maximum clique problem. This problem remains NP-hard even for some simpler classes of graphs, such as comparability, triangle-free and chordal graphs [8]. The complexity of the problem motivated [1,4,12,15,23] to look for approximation algorithms and heuristics, [8,19,20] to consider restricted classes of graphs, and [6] to introduce several integer programming formulations. In this work, we consider the polytope associated to the formulation *MIP1* of [6]. We introduce several families of facet-inducing valid inequalities and discuss the separation problem associated to these families. Note that [21,22], who independently considered a generalization of our problem, also present families of inequalities that are facet-inducing. Some of those facets are related to ours, as we point out when we present our facets. Although the main contribution of our paper lies with the study of the polytope and its facets, for completeness we also perform a small computational study to test the empirical strength of the families of facets and comment on which are most effective computationally.

The remainder of this paper is organized as follows. Section 2 presents the formulation and studies some of its properties. In Section 3, we introduce some classes of valid inequalities for this formulation, discuss in what situations they induce facets of the associated polytope, and explore the computational complexity of the corresponding separation problems. Section 4 describes the implementation of a branch and cut algorithm based on these classes of valid inequalities and reports on our computational results. Finally, we conclude with some remarks and directions for future research.

2. Integer programming formulation

To represent quasi-cliques, we introduce a binary *vertex variable* x_i for every $i \in V$. We set $x_i = 1$ if and only if the vertex i belongs to the k-subset $A \subset V$, which defines a feasible solution. For every $ij \in E$, we introduce a binary *edge variable* z_{ij} that satisfies that $ij \in E(A)$ when $z_{ij} = 1$. Notice that z_{ij} and z_{ji} denote the same variable because edges are undirected. With these definitions, the maximum edge subgraph problem can be formulated using the following integer program. To be consistent with the literature, we refer to this formulation as *MIP1*.

$$\max \sum_{ij \in E} z_{ij} \tag{1a}$$

$$\text{s.t.} \sum_{i \in V} x_i = k \tag{1b}$$

$$z_{ij} \le x_i \quad \forall ij \in E \tag{1c}$$

$$z_{ij} \le x_j \quad \forall ij \in E \tag{1d}$$

$$x_i \in \{0, 1\} \quad \forall i \in V \tag{1e}$$

$$z_{ij} \in \{0, 1\} \quad \forall ij \in E. \tag{1f}$$

Here, the objective counts the number of edges in the subgraph, the first constraint guarantees that the subgraph has the desired size, and the second and third constraints imply that only edges with both endpoints in the subset can be selected. Note that (1f) can be dropped from the formulation, thus letting the edge variables be continuous and free variables. In an optimal solution, the edge variables will be binary automatically.

We define $P(G, k) \subseteq \mathbb{R}^{|V|+|E|}$ to be the convex hull of the vectors (x, z) satisfying constraints (1b)–(1f). Moreover, we let $P_{LP}(G, k) \subseteq \mathbb{R}^{|V|+|E|}$ be the linear relaxation of $P(G, k)$, as given by constraints (1b)–(1d), $0 \le x_i \le 1$ for $i \in V$, and $0 \le z_{ij} \le 1$ for $ij \in E$. The next result implies that the polytope does not *lose* more dimensions than the single one lost by the equality constraint (1b). This proof uses standard arguments and is therefore omitted.

Theorem 1. *The dimension of the polytope $P(G, k)$ is $|V| + |E| - 1$.*

3. Valid inequalities

Each of the following subsections introduces a class of facet-inducing valid inequalities for $P(G, k)$ and studies the complexity of the associated separation problem. All these classes arise from combinatorial structures in the original graph