



A polyhedral study of the maximum edge subgraph problem

Flavia Bonomo¹

*CONICET and Computer Science Dept., University of Buenos Aires
Buenos Aires, Argentina*

Javier Marengo²

Daniela Sabán²

*Sciences Institute, National University of General Sarmiento
Malvinas Argentinas, Argentina*

Nicolás Stier-Moses³

*Graduate School of Business, Columbia University
New York, NY, USA*

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 in 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.

Keywords: Polyhedral combinatorics, Maximum edge subgraph problem

1 Introduction

Social network analysis (SNA) 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 in SNA is the detection and analysis of *cohesive subgroups*, which are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties [8]. Cohesive subgroups are studied using different relaxations of the definition of clique. *Quasi-cliques*—subgraphs with a pre-specified minimum edge density—are one of the most popular relaxations. The detection of quasi-cliques is crucial in [7] for studying the network of bilateral investment treaties. In this case, quasi-cliques are used both in the analysis of cohesive subgroups and as an instrument to evaluate differences in the topology of random graphs.

There are two main approaches to study quasi-cliques: (a) given a specified edge density $\gamma \in [0, 1]$, find the largest vertex set which is γ -dense and, (b) given a size k , find the densest set of k vertices. The second approach is known in the graph and optimization literature as the *maximum edge subgraph problem* (MESP) or *dense/densest/heaviest k -subgraph problem*. Formally, given a graph $G = (V, E)$ and an integer k , the MESP consists in finding a vertex subset $A \subseteq V$ with $|A| = k$ and such that $|E(A)|$ is maximum, where $E(A) = \{ij \in E : i \in A \text{ and } j \in A\}$. The maximum clique problem clearly reduces to the MESP, hence the latter is NP-hard [2]. Approximation algorithms for the MESP have been presented in [1,4,5,6] while [3] introduces several integer programming formulations for this problem. In this work, we study the polytope associated to the formulation *MIP1* of MESP introduced by [3], introduce several families of facet-inducing valid inequalities, and discuss the separation problem associated to these families.

2 Integer programming formulation

An instance of MESP is given by a graph $G = (V, E)$, the cardinality k of the subset $X \subseteq V$ that we want to find, and the minimum density $\gamma \in [0, 1]$. We introduce a binary variable x_i that represents whether $i \in X$ for every $i \in V$, and a binary variable z_{ij} that represents whether $ij \in E(X)$ for every

¹ Email: fbonomo@dc.uba.ar.

² Email: {jmarenco,dsaban}@ungs.edu.ar.

³ Email: stier@gsb.columbia.edu

$ij \in E$. Note that since edges are undirected, z_{ij} and z_{ji} denote the same variable. Using these variables, the maximum edge subgraph problem can be formulated as follows:

$$\begin{aligned} \max \quad & \sum_{ij \in E} z_{ij} \\ \text{s.t.} \quad & \sum_{i \in V} x_i = k \end{aligned} \tag{1}$$

$$z_{ij} \leq x_i \quad \forall ij \in E \tag{2}$$

$$z_{ij} \leq x_j \quad \forall ij \in E \tag{3}$$

$$x_i \in \{0, 1\} \quad \forall i \in V \tag{4}$$

$$z_{ij} \in \{0, 1\} \quad \forall ij \in E \tag{5}$$

Note that binary constraints (5) can be removed since variables z_{ij} must take integer values in any optimal solution. We define $P(G, k) \subseteq \mathbf{R}^{|V|+|E|}$ to be the convex hull of vectors (x, z) satisfying constraints (1)-(5). (Proofs are deferred to the full version of the paper because of lack of space.)

Theorem 2.1 $\dim(P(G, k)) = |V| + |E| - 1$.

3 Valid inequalities

We let $N(i) = \{j \in V : ij \in E\}$ be the *neighborhood* of a vertex i , for $i \in V$, and $\delta(i) = \{ij : ij \in E\}$ be the edges incident to i . For $i \in V$, we define

$$\sum_{j \in N(i)} z_{ij} \leq (k - 1)x_i \tag{6}$$

to be the *neighborhood inequality* associated with the vertex i .

Theorem 3.1 *The neighborhood inequality (6) is valid for $P(G, k)$. If $|N(i)| \geq k$ and $|V| \geq k + 2$, then (6) defines a facet of this polytope.*

Note that the family of neighborhood inequalities is composed of just $|V|$ inequalities. Hence, these inequalities can be added to constraints (1)-(5) to strengthen the formulation.

For $i \in V$ and $A \subseteq V \setminus \{i\}$ with $|A| = k - 2$, we define

$$\sum_{j \in A} x_j + \sum_{j \in N(i) \setminus A} z_{ij} \leq (k - 2) + x_i \tag{7}$$

to be the *extended neighborhood inequality* associated with the vertex i and the set A .

Theorem 3.2 *The extended neighborhood inequality (7) is valid for $P(G, k)$. Moreover, if $|N(i) \setminus A| \geq 2$ and $|V| \geq k + 2$, then (7) is facet-inducing for $P(G, k)$.*

It is interesting to observe that inequalities (6) and (7) can be generalized into a single family. To this end, for $i \in V$ and $A \subseteq V \setminus \{i\}$ with $|A| \leq k - 2$, we define

$$\sum_{j \in A} x_j + \sum_{j \in N(i) \setminus A} z_{ij} \leq |A| + (k - |A| - 1)x_i \tag{8}$$

to be the *generalized neighborhood inequality* associated with the vertex i and the set A .

Theorem 3.3 *The generalized neighborhood inequality (8) is valid for $P(G, k)$. Moreover, if $|N(i) \setminus A| \geq k - |A|$ and $|V| \geq k + 2$, then (8) is facet-inducing for $P(G, k)$.*

Theorem 3.4 *The generalized neighborhood inequality can be separated in $O(n^2 \cdot \log(n))$ time*

Let $A \subseteq V$ be a vertex subset and let $B \subseteq E(V \setminus A)$ be a nonempty maximal matching of $E(V \setminus A)$. In this setting, we define

$$\sum_{i \in A} x_i + \sum_{ij \in B} z_{ij} \leq \frac{|A| + k - 1}{2} \tag{9}$$

to be the *matching inequality* associated with the set A and the matching B . Let $V(B) \subseteq V$ be the set of endpoints from the edges of B .

Theorem 3.5 *The matching inequality (9) is valid for $P(G, k)$ if $|A| < k$ and $|A| + k$ is odd. In addition, if $|B| > (k - |A| - 1)/2$, then (9) is facet-defining as long as $|A| \leq k + 3$ or $|V| \geq k + 3$.*

Theorem 3.6 *The matching inequalities can be separated in $O(|V|^3)$ time if k is odd, and can be separated in $O(|V|^4)$ time if k is even.*

Let $A \subseteq V$ be a vertex subset such that $|V \setminus A| = k - 1$ and let $T \subseteq E(A)$ be an acyclic edge subset. We define

$$1 + \sum_{ij \in T} z_{ij} \leq \sum_{i \in A} x_i \tag{10}$$

to be the *forest inequality* associated with the vertex set A and the edge set T .

Theorem 3.7 *The forest inequality (10) is valid for $P(G, k)$. In addition, if for every $ij \in E(A) \setminus T$ there exists a path in T from i to j of length at most $k - 1$, then (10) is facet-defining.*

Theorem 3.8 *The separation of forest inequalities is NP-hard.*

For a subgraph $H \subseteq G$, denote by $d_H(i)$ the degree of the vertex i in the subgraph H , that is, $d_H(i) = |\delta_H(i)|$. Let $A \subset V$ be a vertex subset with $|A| < k$, and let $T = (V_T, E_T) \subseteq G \setminus A$ be a spanning tree on $k - |A| + 1$ vertices. We define

$$\sum_{i \in A} x_i + \sum_{ij \in E_T} z_{ij} \leq |A| + \sum_{i \in V_T} (d_T(i) - 1)x_i \tag{11}$$

to be the *tree inequality* associated with the set A and the tree T . Notice that if $A = \emptyset$, the inequality still holds.

Theorem 3.9 *The tree inequality (11) is valid for $P(G, k)$. Furthermore, if $|V| > 2k - 2 - \ell$, where ℓ is the number of leaves of T , then (11) induces a facet of this polytope if and only if T is not a star.*

Notice that if T is a star, then the tree inequality is dominated by (8). A special case of (11) arises when T is a path and $A = \emptyset$. In this case, the inequality (11) holds for every path of length at least k and is facet-defining for all paths of length ℓ such that $k \leq \ell \leq 2k - 2$.

Theorem 3.10 *The separation of tree inequalities is NP-hard.*

Let $M \subseteq E$ be a matching such that $0 < |M| < (k - 1)/2$ and let $T = (V_T, E_T) \subseteq G$ be a tree on $k - 2|M| + 1$ vertices such that $V_T \subseteq V \setminus V(M)$. We define

$$\sum_{pr \in M} z_{pr} + \sum_{ij \in E_T} z_{ij} \leq |M| + \sum_{i \in V_T} (d_T(i) - 1)x_i \tag{12}$$

to be the *disjoint matching inequality* associated with the tree T , and the edge matching M .

Theorem 3.11 *The disjoint matching inequality (12) is valid for $P(G, k)$. Furthermore, if $|V| \geq k + 2$ and $|V| > 2k - 2|M| - \ell$ where ℓ is the number of leaves of T , then (12) induces a facet of $P(G, k)$.*

Theorem 3.12 *The separation of disjoint matching inequalities is NP-hard.*

4 Concluding remarks

In this work we have presented a polyhedral study of the maximum edge subgraph problem, by introducing seven families of valid inequalities. These results show that the associated polytope admits interesting facets arising from simple combinatorial structures, and we conjecture that many of the families introduced in this work can be generalized further. We have also analyzed the complexity of the associated separation problems for these families and identified some cases that are NP-hard. We plan to extend the polyhedral study by testing the strength of the presented families of valid inequalities, and by assessing their performance in a branch and cut environment.

Acknowledgements. Partially supported by ANPCyT PICT-2007-00518 and 00533, and UBACyT Grants X069 and X606 (Arg.) and the Columbia University Center for International Business Education and Research (USA).

References

- [1] Y. Asahiro, K. Iwama, H. Tamaki, and T. Tokuyama, *Greedily Finding Dense Subgraphs*, Journal of Algorithms, 34 (2000), 203-221.
- [2] Y. Asahiro, R. Hassin, K. Iwama, *Complexity of finding dense subgraphs*, Discrete Applied Mathematics, 121 (2002), 15-26.
- [3] A. Billionnet, *Different formulations for solving the heaviest k -subgraph problem*, Information Systems and Operational Research, 43-3 (2005), 171-186.
- [4] U. Feige, G. Kortsarz, D. Peleg, *The Dense k -Subgraph Problem*, Algorithmica 29 (2001) 410-421.
- [5] Han, Q. Ye, Y. and Zhang, J. 2000, *Approximation of Dense- k -Subgraph*, Working Paper, Department of Management Sciences, Henry B. Tippie College of Business, The University of Iowa.
- [6] F. Roupin and A. Billionnet, *A deterministic approximation algorithm for the Densest k -Subgraph Problem*, International Journal of Operational Research, 3-3 (2008), 301-314.
- [7] D. Saban, *Analysis and Models of Bilateral Investment Treaties using a Social Networks Approach*, MSc. Thesis (2009), Computer Science Dept., University of Buenos Aires, Buenos Aires, Argentina.
- [8] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*, Cambridge University Press, (1994).