A polyhedral study of the maximum edge subgraph problem

Flavia Bonomo [1]

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

Javier Marenco [2] Daniela Sabán [3]

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

Nicolás Stier [4]

Columbia University, USA

Abstract

The study of cohesive subgroups is a relevant aspect of social network analysis. Cohesive subgroups are studied using different relaxations of the definition of clique in a graph, one of them generating the *maximum edge subgraph problem.* Given a graph and an integer k, this problem consists in finding a k-vertex subset such that the number of edges within the subset is maximum. This problem is NP-hard, and in this work we start an integer programming approach by studying the polytope associated to a straigthforward integer programming formulation. We present several families of facet-inducing valid inequalities for this polytope, and we discuss the separation problem associated to restrictions of some of 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 represented using graphs, where vertices represent the entities and edges represent one or more specific types of interdependency 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 are one of the most popular relaxations, a *quasi-clique* being a subgraph with a pre-specified edge density. 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 analisys 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], and [3] introduces several integer programming formulations for this problem. In this work we are interested in the straigthforward formulation referred as MIP1 in [3], and we present an initial study of the associated polytope.

2 Integer programming formulation

Let $G = (V, E)$ be a graph. For every $i \in V$, we introduce the binary *vertex variable* x_i such that $x_i = 1$ if and only if the vertex i belongs to the k-subset $A \subseteq V$ defining the feasible solution. For every $ij \in E$, we introduce the

[1] Email: fbonomo@dc.uba.ar - Partially supported by ANPCyT PICT-2007-00518 and 00533, and UBACyT Grants X069 and X606 (Arg.).
[2] Email: jmarenco@ungs.edu.ar
[3] Email: dsaban@ungs.edu.ar
[4] Email: stier@gsb.columbia.edu

binary *edge variable* z_{ij} such that $z_{ij} = 0$ if $ij \notin E(A)$. Since the edges are undirected, we consider z_{ij} and z_{ji} as the same variable. Under these definitions, the maximum edge subgraph problem can be formulated as follows:

$$\max \quad \sum_{ij \in E} z_{ij}$$

$$\sum_{i \in V} x_i = k \tag{1}$$

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

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

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

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

Note that the constraints (5) can be relaxed to being free variables for $ij \in E$, as these variables take integer values in any optimal solution. We define $P(G,k) \subseteq \mathbf{R}^{|V|+|E|}$ to be the convex hull of the vectors $(x,z) \in \mathbf{R}^{|V|+|E|}$ satisfying constraints (1)-(5).

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

3 Valid inequalities

For $i \in V$, we define $N(i) = \{j \in V : ij \in E\}$ to be the *neighborhood* of the vertex i, and we call $\delta(i) = \{ij : ij \in E\}$ to the set of 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 by $|V|$ inequalities only, hence these inequalities can be added to the initial integer programming formulation in a practical environment.

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

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

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 $|N(i)\backslash A| \geq 2$ 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 $A \subseteq V\backslash\{i\}$ with $|A| \leq k-2$, we define

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

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 $|N(i)\backslash A| \geq k - |A|$ and $|V| \geq k+2$, then (8) is facet-inducing for $P(G,k)$.*

Note that the generalized neighborhood inequalities restricted to $A \subseteq N(i)$ can be separated in polynomial time: fix a vertex $i \in V$ and, for every $j \in N(i)$, add j to A if and only if $x_j > z_{ij}$. If the resulting inequality is not violated, then no generalized neighborhood inequality associated with the vertex i and having $A \subseteq N(i)$ is violated.

For $B \subseteq E$ and $i \in V$, we define $\delta_B(i) = \delta(i) \cap B$ to be the set of edges incident to i in the edge set B. Let $T = (V_T, E_T) \subseteq G$ be a spanning tree on $k-1$ vertices. Let $pr \in E$ such that $p, r \in V\backslash V_T$. We define

$$z_{pr} + \sum_{ij \in E_T} z_{ij} \;\leq\; 1 + \sum_{i \in V_T} (|\delta_T(i)| - 1)x_i \tag{9}$$

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

Theorem 3.4 *The disjoint edge inequality (9) is valid for $P(G,k)$. Furthermore, if $|V| > 2k-2-l$, where l is the number of leaves of T, then (9) induces a facet of $P(G,k)$.*

Note that the disjoint edge inequalities restricted to the case where T is a star can be separated in polynomial time: for every vertex $i \in V$ and every edge $pr \in E$ with $pr \notin \delta(i)$, consider the vertices in $N(i)\backslash\{p,r\}$ in non-increasing order of $x_j - z_{ij}$, insert into A the first $k-2$ vertices in such ordering, and determine whether the disjoint edge inequality associated with the vertex i, the set A, and the edge pr is violated or not.

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

$$\sum_{i \in A} x_i + \sum_{ij \in B} z_{ij} \;\leq\; \frac{|A| + k - 1}{2} \tag{10}$$

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 (10) is valid for $P(G, k)$. In addition, if $|A| \leq k$ and $|A| + k$ is odd, then (10) is facet-defining as long as $A \cup V(B)$ is strictly contained in V or $|A| + 2|B| \geq k + 3$.*

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 $A \subset V$ be a vertex subset with $|A| < k$, and let $T = (V_T, E_T) \subseteq G \backslash A$ be a spanning tree on $k - |A| + 1$ vertices. We define

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

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

Theorem 3.7 *The tree inequality (11) is valid for $P(G, k)$. Furthermore, if $|V| > 2k - 2 - l$, 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 $k \leq \ell \leq 2k - 2$.

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

$$1 + \sum_{ij \in T} z_{ij} \;\leq\; \sum_{i \in A} x_i \tag{12}$$

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

Theorem 3.8 *The forest inequality (12) is valid for $P(G, k)$. In addition, if for every $ij \in E(A) \backslash T$ there exists a path in T from i to j of length at most $k - 1$, then (12) is facet-defining.*

4 Concluding remarks

In this work we have presented an initial 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 further generalized. We plan to extend these polyhedral results and to further analyze the complexity of the associated separation problems, in order to identify the NP-complete cases and develop and test suitable separation heuristics.

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, v43 i3 (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, De- partment 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).