Towards a Generalization of the Motzkin-Straus Theorem to Edge-weighted Graphs

Massimiliano Pavan, Marcello Pelillo, and Enrico Jabara

Dipartimento di Informatica
Università Ca' Foscari di Venezia
Via Torino 155, 30172 Venezia Mestre, Italy
{mapavan,pelillo,jabara}@dsi.unive.it

Abstract. Given an unweighted undirected graph G with adjacency matrix A, the Motzkin-Straus theorem establishes a remarkable connection between the local/global solutions of the following quadratic program:

$$\begin{array}{ll} \text{maximize} & \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}/2 \\ \text{subject to} & \boldsymbol{e}^{\mathrm{T}} \boldsymbol{x} = 1,\ \boldsymbol{x} \in \mathbb{R}^n_+ \end{array}$$

where $\boldsymbol{e} = (1, ..., 1)^{\mathrm{T}}$, and the maximal/maximum cliques of G. Given an edge-weighted undirected graph G and the corresponding weight matrix A, in this paper we address the following question: What kind of (combinatorial) structures of G are associated to the (continuous) local solutions of our quadratic program? We show that these structures correspond to a "weighted" generalization of maximal cliques, thereby providing a first step towards an edge-weighted generalization of the Motzkin-Straus theorem. Potential applications of our result are in the areas of computer vision and pattern recognition.

1 Introduction

Let $G = (V, E)$ be a simple undirected (and unweighted) graph, where $V = \{1, \dots, n\}$ is the vertex set and $E \subseteq \binom{V}{2}$ is the edge set, $\binom{V}{2}$ denoting the set of all two-element subsets of V. A *clique* of G is a subset of V in which all vertices are pairwise adjacent. A clique C is called *maximal* if no strict superset of C is a clique, i.e., no vertex external to C is adjacent to more than $|C| - 1$ vertices of C (here, and in the sequel, $|C|$ denotes the cardinality of a set C). If, in addition, no external vertex of C is adjacent to exactly $|C| - 1$ vertices of C, then a maximal clique C will be called *strictly maximal*. Finally, a maximum cardinality clique (or, simply, a *maximum clique*) is a clique whose cardinality is the largest possible. Clearly, a maximum clique is also maximal but the converse need not be true. The maximum size of a clique in G is called the *clique number* (of G) and is denoted by $\omega(G)$. The problem of determining a maximum clique of a graph (or finding its clique number) is a classical combinatorial optimization problem with a variety of practical applications. It is known to be NP-hard for arbitrary graphs and, according to recent theoretical results, so is the problem of approximating it within a constant factor. We refer to [4] for a recent review concerning algorithms, applications and complexity issues of this important problem.

Let

$$\Delta = \{\boldsymbol{x} \in \mathrm{I\!R}^n : \boldsymbol{x} \geq \boldsymbol{0} \text{ and } \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = 1\}$$

be the standard simplex of $\mathrm{I\!R}^n$, where $\boldsymbol{e}$ is a vector of appropriate length consisting of unit entries (hence $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = \sum_i x_i$), and consider the following quadratic function which is sometimes called the *Lagrangian* of graph G:

$$f_G(\boldsymbol{x}) = \sum_{\{i,j\}\in E} x_i x_j \ . \tag{1}$$

In 1965, Motzkin and Straus [10] established a remarkable connection between the maxima of the Lagrangian of a graph over Δ and its clique number. Specifically, if $\boldsymbol{x}^*$ is a maximizer of f_G in Δ, they showed that the clique number $\omega(G)$ of G is related to $f_G(\boldsymbol{x}^*)$ according to the following formula:

$$\omega(G) = \frac{1}{1 - 2f_G(\boldsymbol{x}^*)} \ . \tag{2}$$

Additionally, they proved that a subset of vertices S is a maximum clique of G if and only if its (unweighted) *characteristic vector* $\boldsymbol{x}^{S,\mathrm{u}}$, which is the vector in Δ defined as:

$$x_i^{S,\mathrm{u}} = \begin{cases} 1/|S|, & \text{if } i \in S \\ 0, & \text{otherwise} \end{cases}$$

is a global maximizer of f in Δ.[1] Recently, a similar connection has been established between *maximal* cliques of G and *local* maximizers of f_G [14, 7].

The connection established by the Motzkin-Straus theorem between a purely combinatorial problem and a continuous one, although probably motivated by genuine mathematical curiosity, is indeed a fruitful one both from a theoretical and a practical standpoint. For example, in their original paper Motzkin and Straus used their result to provide an alternative proof of Turán's classical theorem, and Wilf [16] derived from it new bounds on the clique size. Computationally, the result also has proven to be extremely useful since it motivated several clique-finding heuristics (see, e.g., [5, 6, 11, 13, 9])

Recently, several generalizations of the Motzkin-Straus theorem have been proposed including, for example, Bomze's spurious-free regularized version [2], Gibbons et al.'s extension to vertex-weighted graphs [7], based on a suggestion by Lovász, and Sós and Straus' generalization to hypergraphs [15].

In this paper, we provide a first step towards a further generalization of the Motzkin-Straus theorem. Specifically, given an edge-weighted simple undirected graph $G = (V, E)$ with (positive) weight function $w : E \to \mathrm{I\!R}_+^*$, we consider the *weighted Lagrangian* of G defined as:[2]

$$f(\boldsymbol{x}) \equiv f_{G,w}(\boldsymbol{x}) = \sum_{\{i,j\}\in E} w(i,j) x_i x_j \tag{3}$$

[1] Actually, in their original paper, Motzkin and Straus proved just the "only-if" part of this theorem. The converse direction is however a straightforward consequence of their result [14].

[2] By convention, we write $w(i, j)$ instead of $w(\{i, j\})$.

and we address the following question: What kind of combinatorial structures of G, if any, are associated to the maximizers of its weighted Lagrangian $f_{G,w}$ on the simplex Δ? We show that such structures do exist and correspond to a "weighted" generalization of maximal (and maximum) cliques, thereby providing a first step towards an edge-weighted generalization of the Motzkin-Straus theorem.

Potential applications of our result are in the areas of computer vision and pattern recognition, where graphs are employed as a convenient means of representing high-level pictorial information [1]. Here, the vertices of these graphs typically represent features such as lines, corners, regions, etc., and the edges represent possible relations among them. A fundamental problem in these areas is to match a new image, or image representation, against a model, for recognition purposes. An elegant approach to solve this matching problem is to construct an auxiliary graph structure, called the *association graph*, from the two graphs being matched and then to look for maximum cliques in it [1]. In the association graph, the vertices correspond to putative associations between vertices of the graphs being matched, and the edges represent compatibility relations between associations. Graphs arising in these applications, however, often contains numerical attributes on their edges, and this leads to an edge-weighted association graph where the weights on the edges correspond to a measure of the "similarity" between attribute vectors.

2 Preliminaries

Let $G = (V, E)$ be an edge-weighted simple undirected graph of order n and let $A = (a_{ij})$ be the corresponding *weighted adjacency matrix*, which is the $n \times n$ symmetric matrix defined as:

$$a_{ij} = \begin{cases} w(i,j), & \text{if } \{i,j\} \in E \\ 0, & \text{otherwise.} \end{cases}$$

We shall consider the following standard quadratic program:[3]

$$\begin{array}{ll} \text{maximize} & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \in \Delta \end{array} \tag{4}$$

whose objective function is nothing but the weighted Lagrangian of G, as defined in (3).

A point $\boldsymbol{x}^* \in \Delta$ is said to be a *global* solution of program (4) if $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$, for all $\boldsymbol{x} \in \Delta$. It is said to be a *local* solution if there exists an $\varepsilon > 0$ such that $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Delta$ whose distance from $\boldsymbol{x}^*$ is less than ε, and if $f(\boldsymbol{x}^*) = f(\boldsymbol{x})$ implies $\boldsymbol{x}^* = \boldsymbol{x}$, then $\boldsymbol{x}^*$ is said to be a *strict* local solution.

Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, the *support* of $\boldsymbol{x}$ is defined as the set of indices corresponding to its non-zero components, that is:

$$\sigma(\boldsymbol{x}) = \{i \in V \ : \ x_i \neq 0\} \tag{5}$$

[3] This terminology is due to Bomze [3].

A point $\boldsymbol{x} \in \Delta$ satisfies the Karush-Kuhn-Tucker (KKT) conditions for problem (4), i.e. the first-order necessary conditions for local optimality [8], if there exist $n+1$ real constants (Lagrange multipliers) $\mu_1, \ldots, \mu_n$ and λ, with $\mu_i \geq 0$ for all $i = 1 \ldots n$, such that:

$$(A\boldsymbol{x})_i - \lambda + \mu_i = 0 \tag{6}$$

for all $i = 1 \ldots n$, and

$$\sum_{i=1}^{n} x_i \mu_i = 0 \ . \tag{7}$$

Note that, since both x_i and μ_i are nonnegative for all $i = 1 \ldots n$, the latter condition is equivalent to saying that $i \in \sigma(\boldsymbol{x})$ implies $\mu_i = 0$. Hence, the KKT conditions can be rewritten as:

$$(A\boldsymbol{x})_i \begin{cases} = \lambda, & \text{if } i \in \sigma(\boldsymbol{x}) \\ \leq \lambda, & \text{otherwise} \end{cases} \tag{8}$$

for some real constant λ (indeed, it is immediate to see that $\lambda = \boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}$). A point $\boldsymbol{x} \in \Delta$ satisfying (8) will be called a *KKT point* throughout.

We shall find it useful to introduce some new notations. If $S \subseteq V$, we denote by A_S the submatrix of A formed by the rows and the columns indexed by the elements of S. Additionally, we define the matrix B_S as:

$$B_S = \begin{pmatrix} 0 & \boldsymbol{e}^{\mathrm{T}} \\ \boldsymbol{e} & A_S \end{pmatrix}$$

and, assuming $S = \{i_1, \ldots, i_m\}$ with $i_1 < \cdots < i_m$, the matrix ${}^{j}B_S$ is defined to be:

$${}^{j}B_S = \begin{pmatrix} 0 & & & & \boldsymbol{e}^{\mathrm{T}} & & & & \\ \boldsymbol{e} & A_S^1 & A_S^2 & \cdots & A_S^{j-1} & \boldsymbol{0} & A_S^{j+1} & A_S^{j+2} & \cdots & A_S^n \end{pmatrix}$$

where A_S^i denotes the i-th column of A_S. With these notations, note that the KKT equality conditions in (8) amount to saying that there exists a real number λ such that:

$$B_\sigma \, (\lambda, x_{i_1}, \ldots, x_{i_m})^{\mathrm{T}} = (1, 0, \ldots, 0)^{\mathrm{T}} \tag{9}$$

where $\sigma = \sigma(\boldsymbol{x}) = \{i_1, \ldots, i_m\}$ with $i_1 < \cdots < i_m$.

3 The Combinatorics of Standard Quadratic Programs

Our goal in this paper is to provide a combinatorial interpretation of the solutions of program (4). Basically, our main result is centered around the intuitive idea that the assignment of edge-weights induces, in some way to be described, an assignment of weights on the vertices.

To grasp the intuition behind this idea, consider the graph in Figure 1 and the subgraph induced by the set $S = \{1, 2, 3\}$. Observe that the edges incident to vertex 1 are the lightest ones (within the subgraph), the heaviest ones are incident to vertex 3 and those incident to 2 are the lightest as well as the heaviest ones. This induces a sort of natural ranking among the vertices of S, which is captured by the notions introduced below.

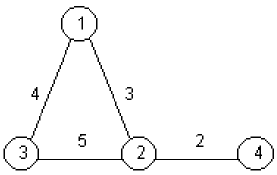

Fig. 1. An example edge-weighted graph.

In order to formalize how such induction of node-weights takes place, we need some preliminary definitions. Let $S \subseteq V$ be a non-empty subset of vertices and $i \in V$. The *(average) weighted degree* of i w.r.t. S is defined as:[4]

$$\mathrm{awdeg}_S\,(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij} \ . \tag{10}$$

Observe that $\mathrm{awdeg}_{\{i\}}\,(i) = 0$ for any $i \in V$. Moreover, if $j \notin S$ we define:

$$\phi_S\,(i,j) = a_{ij} - \mathrm{awdeg}_S\,(i) \ . \tag{11}$$

Note that $\phi_{\{i\}}(i,j) = a_{ij}$, for all $i, j \in V$ with $i \neq j$. Referring to the graph in Figure 1 we have:

$$\phi_{\{1,2\}}\,(1,3) = a_{13} - \mathrm{awdeg}_{\{1,2\}}\,(1) = 4 - \frac{0+3}{2} = \frac{5}{2} \ .$$

We are now in a position to formalize the notion of induction of node-weights.

Definition 1. *Let $S \subseteq V$ be a non-empty subset of vertices and $i \in S$. The* weight *of i w.r.t. S is*

$$\mathrm{w}_S\,(i) = \begin{cases} 1, & \textit{if } |S| = 1 \\ \displaystyle\sum_{j \in S \setminus \{i\}} \phi_{S \setminus \{i\}}\,(j,i)\, \mathrm{w}_{S \setminus \{i\}}\,(j)\,, & \textit{otherwise.} \end{cases} \tag{12}$$

Moreover, the total weight *of S is defined to be:*

$$\mathrm{W}(S) = \sum_{i \in S} \mathrm{w}_S(i) \ . \tag{13}$$

[4] Note that, although we consider graphs with no self-loops, in computing the average weighted degree of $i \in S$ we divide by $|S|$ instead of $|S| - 1$, and this can lead to somehow counterintuitive results. For example, referring to the graph in Figure 1, $\mathrm{awdeg}_{\{1,2\}}(1) = \frac{0+3}{2} < 3$.

Note that $\mathrm{w}_{\{i,j\}}(i) = \mathrm{w}_{\{i,j\}}(j) = a_{ij}$, for all $i, j \in V$ $(i \neq j)$. For example, in Figure 1 we have:

$$\begin{aligned}\mathrm{w}_{\{1,2,3\}}(3) &= \phi_{\{1,2\}}(1,3)\,\mathrm{w}_{\{1,2\}}(1) + \phi_{\{1,2\}}(2,3)\,\mathrm{w}_{\{1,2\}}(2) \\ &= [4-(0+3)/2]3 + [5-(0+3)/2]3 = 18 \ .\end{aligned}$$

The following useful lemma characterizes the notions introduced above in terms of determinants. The proof proceeds by induction and exploits elementary properties of the determinant (see [12] for details).

Lemma 1. *Let $S = \{i_1, \ldots, i_m\} \subseteq V$ be a non-empty subset of vertices and, w.l.o.g., assume $i_1 < \cdots < i_m$. Then, we have:*

$$\mathrm{w}_S(i_h) = (-1)^m \det\left({}^h B_S\right), \tag{14}$$

for any $i_h \in S$. Moreover:

$$\mathrm{W}(S) = (-1)^m \det(B_S) \ . \tag{15}$$

The next lemma, which follows easily from the previous one, provides an alternative way of computing the $\mathrm{w}_S(i)$'s.

Lemma 2. *For any non-empty subset of vertices $S \subseteq V$ and $i \in S$ we have:*

$$\mathrm{w}_S(i) = \sum_{j \in S \setminus \{i\}} (a_{ij} - a_{hj})\,\mathrm{w}_{S \setminus \{i\}}(j) \tag{16}$$

for all $h \in S \setminus \{i\}$.

The next definition is instrumental to characterize the solutions of our standard quadratic program (4) in terms of combinatorial structures of G.

Definition 2. *We say that a non-empty subset of vertices S admits a* weighted characteristic vector $\boldsymbol{x}^S$ *if it has non-null total weight* $\mathrm{W}(S)$*, in which case we set:*

$$x_i^S = \begin{cases} \frac{\mathrm{w}_S(i)}{\mathrm{W}(S)}, & \text{if } i \in S \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

For example, consider the graph in Figure 1 and the set $S = \{1, 2, 3\}$. Since $\mathrm{w}_S(1) = 10$ and $\mathrm{w}_S(2) = 16$, we have:

$$\boldsymbol{x}^S = (10/44, 16/44, 18/44, 0)^{\mathrm{T}} \approx (0.23, 0.36, 0.41, 0)^{\mathrm{T}}$$

and this confirms our intuitive ranking of the vertices in S based on the weights of the incident edges. The total weight of a subset of vertices $\mathrm{W}(S)$ depends clearly on the topology of the graph and on the weights assigned to the edges, and can well be zero. Hence, a subset of vertices need not always admit a characteristic vector. For example, it is simple to see that any subset of at least two mutually non-adjacent vertices does not admit a characteristic vector.

The next two results establish useful connections between KKT points of program (4) and weighted characteristic vectors.

Lemma 3. *Let $\sigma = \sigma(\boldsymbol{x})$ be the support of a vector $\boldsymbol{x} \in \Delta$ which admits weighted characteristic vector $\boldsymbol{x}^\sigma$. Then $\boldsymbol{x}$ satisfies the KKT equality conditions in (8) if and only if $\boldsymbol{x} = \boldsymbol{x}^\sigma$. Moreover, in this case, we have:*

$$\frac{\mathrm{w}_{\sigma \cup \{j\}}(j)}{\mathrm{W}(\sigma)} = (A\boldsymbol{x})_j - (A\boldsymbol{x})_i = -\mu_j \tag{18}$$

for all $i \in \sigma$ and $j \notin \sigma$, where the μ_j's are the (nonnegative) Lagrange multipliers of program (4) as defined in Section 2.

Proof. Note that conditions (9), which are equivalent to the KKT equality conditions in (8), can be regarded as a system of linear equations in the unknowns λ and x_i's ($i \in \sigma$). For Lemma 1, the system has a unique solution since $\det(B_\sigma) \neq 0$. Hence, supposing $\sigma = \{i_1, \ldots, i_m\}$ and, w.l.o.g., $i_1 < \ldots < i_m$, from Cramer's rule and Lemma 1 we have:

$$x_{i_h} = \frac{\det\left({}^h B_\sigma\right)}{\det(B_\sigma)} = \frac{(-1)^m \mathrm{w}_\sigma(i_h)}{(-1)^m \mathrm{W}(\sigma)} = \frac{\mathrm{w}_\sigma(i_h)}{\mathrm{W}(\sigma)}$$

for any $1 \leq h \leq m$. Therefore $\boldsymbol{x} = \boldsymbol{x}^\sigma$.

The fact that $(A\boldsymbol{x})_j - (A\boldsymbol{x})_i = -\mu_j$, for $i \in \sigma$ and $j \notin \sigma$, follows immediately from equation (6). Finally, applying Lemma 2 we obtain:

$$\begin{aligned}
\frac{\mathrm{w}_{\sigma \cup \{j\}}(j)}{\mathrm{W}(\sigma)} &= \frac{\sum_{h \in \sigma}(a_{jh} - a_{ih})\mathrm{w}_\sigma(h)}{\mathrm{W}(\sigma)} \\
&= \sum_{h \in \sigma} a_{jh} x_h^\sigma - \sum_{h \in \sigma} a_{ih} x_h^\sigma \\
&= (A\boldsymbol{x}^\sigma)_j - (A\boldsymbol{x}^\sigma)_i
\end{aligned}$$

which concludes the proof, since $\boldsymbol{x} = \boldsymbol{x}^\sigma$. □

Proposition 1. *Let $\boldsymbol{x} \in \Delta$ be a vector whose support has positive total weight $\mathrm{W}(\sigma)$, and hence admitting weighted characteristic vector $\boldsymbol{x}^\sigma$. Then, $\boldsymbol{x}$ is a KKT point for (4) if and only if the following conditions hold:*

1. $\boldsymbol{x} = \boldsymbol{x}^\sigma$
2. $\mathrm{w}_{\sigma \cup \{j\}}(j) \leq 0$, *for all* $j \notin \sigma$.

Proof. Vector $\boldsymbol{x}$ satisfies the KKT conditions (8) if and only if $\boldsymbol{x} = \boldsymbol{x}^\sigma$ (cfr. Lemma 3) and $(A\boldsymbol{x})_j \leq (A\boldsymbol{x})_i$ for any $j \notin \sigma$ and $i \in \sigma$, but from (18) the latter condition amounts to saying that $\mathrm{w}_{\sigma \cup \{j\}}(j) \leq 0$, since $\mathrm{W}(\sigma) > 0$. □

The notion of a weighted characteristic vector is too weak to characterize (strict) local solutions of program (4). A strengthened, but still insufficient, notion is given by the following:

Definition 3. *A non-empty subset of vertices S is said to be* good *if:*

1. $\mathrm{w}_S(i) > 0$, *for all* $i \in S$

2. $\mathrm{W}(T) > 0$*, for any non-empty* $T \subseteq S$

A good subset of vertices S *is said to be* maximally good*, if no strict superset of* S *is good.*

It is evident that any good subset of vertices admits weighted characteristic vector, and it always belongs to the simplex Δ. Note that the subset of vertices $S = \{1, 2\}$ in Figure 1 is good for $\mathrm{w}_S(1) = \mathrm{w}_S(2) = 3 > 0$.

By using the bordered Hessian test [8] from nonlinear programming, it is possible to prove the following instrumental result (cfr. [12]).

Lemma 4. *Let* $S \subseteq V$ *be a subset of* m *vertices. Then,* $\mathrm{W}(T) > 0$ *for any non-empty subset* $T \subseteq S$ *if and only if* A_S *is negative definite in the subspace* $\{\boldsymbol{y} \in \mathbb{R}^m : \sum_{i=1}^m y_i = 0\}$*.*

The following is our first connection between solutions of the continuous problem (4) and combinatorial structures of G.

Lemma 5. *If a vector* $\boldsymbol{x}^* \in \Delta$ *is a strict local solution for program (4), then its support* $\sigma = \sigma(\boldsymbol{x}^*)$ *is good.*

Proof. From the strict local maximality of $\boldsymbol{x}^*$ it follows, after some algebra, that the submatrix A_σ is negative definite in the subspace $\{\boldsymbol{y} \in \mathbb{R}^m : \sum_{i=1}^m y_i = 0\}$, where $m = |\sigma|$, and from Lemma 4 this happens if and only if the second condition for a set to be good is satisfied. It remains to show that $\mathrm{w}_S(i) > 0$ for all $i \in S$, but this follows directly from Lemma 3 (in fact, $\boldsymbol{x}^*$ is a KKT point) and the definition of weighted characteristic vector. □

It is simple to show that the converse of the previous result does not necessarily hold, that is a good set need not correspond to a strict local solution of (4). In order to provide such a reverse connection we need to refine the goodness notion.

Definition 4. *A good subset of vertices* S *is said to be* dominant *if, for any node* $i \notin S$*, we have* $\mathrm{w}_{S \cup \{i\}}(i) < 0$*.*

For example, in the graph of Figure 1 the subset of vertices $S = \{1, 2, 3\}$ is dominant since it is good and $\mathrm{w}_{S \cup \{4\}}(4) = -88 < 0$.

The next theorem, which is the main result of this paper, establishes a correspondence between dominant sets and (strict) local solutions of program (4). It is a generalization of a result proved independently by Pelillo and Jagota [14] and Gibbons et al. [7] which represent the "local" counterpart of the original Motzkin-Straus theorem [10], which is "global" in nature.

Theorem 1. *If* S *is a dominant subset of vertices, then its weighted characteristics vector* $\boldsymbol{x}^S$ *is a strict local solution of program (4).*

Conversely, if $\boldsymbol{x}^*$ *is a strict local solution of program (4) then its support* $\sigma = \sigma(\boldsymbol{x}^*)$ *is dominant, provided that* $\mathrm{w}_{\sigma \cup \{i\}}(i) \neq 0$ *for all* $i \notin \sigma$*.*

Proof. If S is dominant, then from Proposition 1 it follows that $\boldsymbol{x}^S$ is a KKT point for (4). Moreover, by Lemma 3, we have that the j-th nonnegative Lagrange multiplier μ_j ($j \notin S$) is positive if and only if $\mathrm{w}_{S\cup\{j\}}(j) < 0$. Therefore, the second-order sufficient conditions for local optimality [8] together with Lemma 4 imply that $\boldsymbol{x}^S$ is a strict local solution for program (4).

On the other hand, if $\boldsymbol{x}^*$ is a strict local solution of (4) then $\sigma = \sigma(\boldsymbol{x}^*)$ is good, from Lemma 5. Moreover, Proposition 1 states that $\boldsymbol{x}^* = \boldsymbol{x}^\sigma$ and $\mathrm{w}_{\sigma\cup\{j\}}(j) \leq 0$, for all $j \notin \sigma$. Therefore, the fact that σ is dominant follows trivially from the hypotheses. □

Unfortunately, the previous result does not say anything concerning the combinatorial interpretation of $\boldsymbol{x}^*$ when $\mathrm{w}_{\sigma\cup\{i\}}(i) = 0$ for some $i \notin \sigma$. As in the unweighted case [14], this has to do with the presence of "spurious" solutions in (4), namely maximizers of the weighted Lagrangian whose support do not admit a weighted characteristic vector.

4 Back to the Unweighted Case

In this section, we return to the original setting of Motzkin and Straus, namely to unweighted graphs where $w(i,j) = 1$ for all $\{i,j\} \in E$, and we show that all the concepts introduced previously do correspond to standard graph-theoretic notions related to that of a clique (due to lack of space we omit proofs, which can be found in [12]).

To this end, note that if $C \subseteq V$ is a clique of an unweighted graph $G = (V, E)$, then for any $j \notin C$ we have:

$$\mathrm{w}_{C\cup\{j\}}(j) = \deg_C(j) - |C| + 1$$

where, if S is a subset of vertices and $j \in V$, $\deg_S(j) = |\{i \in S : \{i,j\} \in E\}|$ denotes the degree of j w.r.t. S. This follows easily by induction on the cardinality of C and by the observation that for a clique C, and for any $i \in C$ and $j \notin C$, we have

$$\phi_C(i,j) = \begin{cases} \frac{1}{m} & \text{if } \{i,j\} \in E \\ \frac{1}{m} - 1 & \text{otherwise.} \end{cases}$$

Using these properties, it is simple to prove that, for any clique C of G, the following facts hold:

1. $\mathrm{w}_C(i) = 1$, for any $i \in C$, and hence $W(C) = |C|$.
2. C is a maximal clique if and only if $\mathrm{w}_{C\cup\{j\}}(j) \leq 0$ for any $j \notin C$.
3. C is a strictly maximal clique if and only if $\mathrm{w}_{C\cup\{j\}}(j) < 0$ for any $j \notin C$.

From the above properties the next proposition follows.

Proposition 2. *Let $S \subseteq V$ be a subset of vertices of an unweighted graph. Then, the following properties hold:*

1. S is a clique if and only if it is good;
2. S is a maximal clique if and only if it is maximally good;
3. S is a strictly maximal clique if and only if it is dominant.

Hence, S is a maximum clique if and only if it is a good set having largest cardinality.

Finally, note that if C is a clique then it admits a weighted characteristic vector and this coincides with its unweighted counterpart, i.e., $\boldsymbol{x}^C = \boldsymbol{x}^{C,\mathrm{u}}$.

5 Conclusions and Future Work

We have investigated the combinatorial properties of the solutions of a wide subclass of quadratic programming problems, thereby providing a first step towards a generalization of the Motzkin-Straus theorem to edge-weighted graphs. Essentially, we have shown that (strict) maximizers of the weighted Lagrangian on the standard simplex are in correspondence with certain combinatorial structures that we called dominant sets. These are the edge-weighted counterpart of strictly maximal cliques in unweighted graphs.

There are several issues which we have not addressed in this paper. These include the characterization of global maximizers, which would provide a proper generalization of the original Motzkin-Straus theorem, the study of "spurious" solutions which will probably allow us to drop the "strict" qualification in our main theorem, and the extension of the theory to graphs having also weights on the vertices. All this will be subject of future work.

References

1. D. H. Ballard and C. M. Brown. Computer vision, 1982.
2. I. M. Bomze. Evolution towards the maximum clique. *J. Global Optim.*, 10:143–164, 1997.
3. I. M. Bomze. On standard quadratic optimization problems. *J. Global Optim.*, 13:369–387, 1998.
4. I. M. Bomze, M. Budinich, P. M. Pardalos, and M. Pelillo. The maximum clique problem. In D.-Z. Du and P. M. Pardalos, editors, *Handbook of Combinatorial Optimization (Suppl. Vol. A)*, pages 1–74. Kluwer, Boston, MA, 1999.
5. I. M. Bomze, M. Budinich, M. Pelillo, and C. Rossi. Annealed replication: A new heuristic for the maximum clique problem. *Discr. Appl. Math.*, 2002. in press.
6. L. E. Gibbons, D. Hearn, and P. Pardalos. A continuous based heuristic for the maximum clique problem. In D. Johnson and M. Trick, editors, *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge*, volume 26 of *DIMACS*, pages 103–124. American Mathematical Society, 1996.
7. L. E. Gibbons, D. W. Hearn, P. M. Pardalos, and M. V. Ramana. Continuous characterizations of the maximum clique problem. *Math. Oper. Res.*, 22:754–768, 1997.
8. D. G. Luenberger. *Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA, 1984.

9. A. Massaro, M. Pelillo, and I. M. Bomze. A complementary pivoting approach to the maximum weight clique problem. *SIAM Journal on Optimization*, in press.
10. T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canad. J. Math.*, 17:533–540, 1965.
11. P. M. Pardalos and A. T. Phillips. A global optimization approach for solving the maximum clique problem. *Int. J. Computer Math.*, 33:209–216, 1990.
12. M. Pavan, M. Pelillo, and E. Jabara. On the combinatorics of continuous quadratic optimization. Technical report, Department of Computer Science, University of Venice, in preparation.
13. M. Pelillo. Relaxation labeling networks for the maximum clique problem. *J. Artif. Neural Networks*, 2:313–328, 1995.
14. M. Pelillo and A. Jagota. Feasible and infeasible maxima in a quadratic program for maximum clique. *J. Artif. Neural Networks*, 2:411–420, 1995.
15. V. T. Sós and E. G. Straus. Extremal of functions on graphs with applications to graphs and hypergraphs. *J. Combin. Theory B*, 32:246–257, 1982.
16. H. S. Wilf. Spectral bounds for the clique and independence numbers of graphs. *J. Combin. Theory B*, 40:113–117, 1986.