ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
CONSIGLIO NAZIONALE DELLE RICERCHE

Anna Galluccio, Paolo Nobili

A $(1 - \frac{1}{2Q})$-APPROXIMATION ALGORITHM FOR MAX VERTEX COVER IN CLIQUE-COVERABLE GRAPHS

R. 611 Giugno 2004

Anna Galluccio – Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti"-CNR, Viale Manzoni 30, 00185 Roma, Italy. E-mail: `galluccio@iasi.rm.cnr.it`.

Paolo Nobili – Dipartimento di Matematica, Università di Lecce, and IASI-CNR, Italy. E-mail: `paolo.nobili@unile.it`.

This work was partially funded under the Italian Research Program PRIN 2003, n. 2003014039.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica, CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.rm.cnr.it
URL: http://www.iasi.rm.cnr.it

Abstract

The *maximum vertex cover* problem is the following problem: given a graph $G = (V, E)$ with nonnegative weights on the edges and a positive integer p, find p vertices of G such that the total weight of edges covered by these vertices is maximized. This problem is NP-hard, and the best known polynomial time approximation algorithm to solve it in the general case has a performance guarantee of $3/4$. In this paper, we show that this approximation ratio can be increased to $1 - 1/(2q)$ provided that the edge weighting of the input graph can be partitioned into cliques of size at least q.

1. Introduction

The *maximum vertex cover* problem (*MVC*, for short) is defined as follows: given a graph $G = (V, E)$ and a positive integer p, find p nodes of G such that they cover as many edges as possible. Associating with each edge e of G a nonnegative weight $w(e)$, we define a weighted version of the problem when one wants to maximize the total weight of the edges covered by at most p vertices. The MVC problem is related to the well known *vertex cover (VC)* problem where one wants to minimize the number of nodes covering all the edges of a given graph and it is NP-hard, since its polynomial solvability would clearly imply the polynomial solvability of the vertex cover problem. So, we look for approximation algorithms. An approximation algorithm for a maximization problem is said to achieve approximation ratio ρ if the solution provided by the algorithm for any instance of the problem is always at least ρ times the optimum for the same instance.

Several algorithmic results were presented even in the not so recent literature, the major of them concerned with the general *maximum coverage* problem where instead of edges the sets to be covered by p vertices are generic subsets of a given set. The maximum coverage problem is a useful model for a variety of applications regarding covering graphs with subgraphs, circuit layout, scheduling, and facility location (see [9] for further details) and this explains the continous interest it gave rise to.

In early 80's, Nemhauser and Wolsey [12] and Conforti and Cornuejols [4] considered the general problem of maximizing a submodular set function $F(X)$. In particular, let $F(X)$ be a nondecreasing submodular function satisfying $F(\emptyset) = 0$ and let X^* be the set maximizing $F(X)$; they showed that the greedy heuristic achieves an approximation factor of $1 - (1 - 1/|X^*|)^{|X^*|} > 1 - e^{-1}$. It is not difficult to see that the maximum vertex cover problem belongs to this class of optimization problems: in fact, the function $F(X)$ representing the sum of the weights of the sets covered by the vertices in X satisfies the required conditions [13].

In 1999 Ageev and Sviridenko [3] developed a polynomial time algorithm to find an approximate solution to the maximum coverage problem. Their algorithm has an approximation ratio of $1 - (1 - 1/r)^r$ where r is the size of the largest subset to be covered: this ratio is in general not better than the one achieved by the greedy algorithm. Feige [6] showed that, when there is no bound to the maximum size of the sets to be covered, no performance guarantee better than $1 - (1 - 1/p)^p$ can be achieved provided that $P \neq \mathsf{NP}$. A further generalization of the maximum coverage problem was introduced by Khuller, Moss and Naor [10] and consisted of a maximum coverage problem plus a knapsack constraint. For this problem they provide a $(1 - e^{-1})$-approximation algorithm.

The MVC problem can be seen as a maximum coverage problem where the sets to be covered are two-element sets, namely $r = 2$. The problem is interesting on its own also because of its tight connections with the MAX2SAT problem [7], and recent work was focused on classes of graphs where maximum vertex cover can be solved in polynomial time ([2, 1]).

For the MVC problem, Cornuejols, Nemhauser and Wolsey [5] showed that the greedy algorithm almost always finds an optimal solution. Neverthless, as far as we know the best known approximation factor for the maximum vertex cover problem is $3/4$ and it is achieved by the algorithm of Ageev and Sviridenko. Recently, Han et al. [8] describe a semidefinite programming based algorithm for the MVC problem whose approximation factor depends on $\sigma = p/|V|$. This approximation factor increases with σ and for σ sufficiently large it gives an approximation factor better than $3/4$.

In this paper we show that when the edge weighting of the input graph can be partitioned into cliques of size at least q, the approximation factor of the maximum vertex cover can be increased to $1-1/(2q)$ independently from σ. In particular, we provide a new linear relaxation of the classical quadratic formulation of MVC introduced in [3] and [7]. Then we show that the optimal solution z^* of our formulation (computable in polynomial time), with objective function value f^*, can be "rounded" in polynomial time to an integral solution $\tilde{z}$, with objective function value $\tilde{f}$, such that $\tilde{f} \geq (1-\frac{1}{2q})f^*$.

2. The pipage technique

In this section we describe a deterministic rounding method introduced by Ageev and Sviridenko and called *pipage* [3]. This method was used to derive a $1-(1-1/r)^r$-approximation algorithm for the maximum coverage problem where r is the maximum size of the sets to be covered.

The *pipage* rounding method can be described in its generality as follows: let $\mathcal{P}$ be the polyhedron

$$\mathcal{P} = \{x \in \Re^n : \sum_{i=1}^{n} x_i = p, 0 \leq x \leq 1\}, \tag{1}$$

and let $F(x_1, x_2, \ldots, x_n)$ be a function defined on the rational points of $\mathcal{P}$ and computable in polynomial time. Any nonbinary vector $x \in \mathcal{P}$ has at least two fractional components, say x_i and x_j, whose values are strictly between 0 and 1. Thus, we may define $x_{ij}(\epsilon)$ as the point $(x_1, \ldots, x_i+\epsilon, \ldots, x_j-\epsilon, \ldots, x_n)$, and, for $\epsilon \in [-min\{x_i, 1-x_j\}, min\{1-x_i, x_j\}]$, clearly $x_{ij}(\epsilon)$ belongs to $\mathcal{P}$.

If the function $F(x)$ satisfies the following property:

(*) *$F(x_{ij}(\epsilon))$ is convex with respect to $\epsilon \in [-min\{x_i, 1-x_j\}, min\{1-x_i, x_j\}]$ for each pair of indices $i \neq j$ and for each point $x \in \mathcal{P}$,*

then it is not difficult to verify that ϵ can be chosen so that the new point $x_{ij}(\epsilon)$ in $\mathcal{P}$ has fewer fractional components and $F(x_{ij}(\epsilon)) \geq F(x)$. So, we update the current solution by substituting x with $x_{ij}(\epsilon)$. Now, by iterating this rounding procedure at most $n-1$ times we obtain a binary vector $\bar{x}$ such that

$$F(\bar{x}) \geq F(x)$$

where x was the original fractional point in $\mathcal{P}$. This method is the so-called *pipage* rounding procedure.

Now, consider a graph $G=(V,E)$ with $|V|=n$ nodes and $|E|=m$ edges, with a nonnegative weight w_{ij} associated with every edge $ij \in E$, and let p be a positive integer. The MVC problem can be formulated as a nonlinear $0-1$ program as follows:

$$\begin{aligned} \max & \sum_{ij \in E} w_{ij}(x_i + x_j - x_i x_j) \\ & \sum_{i \in V} x_i \leq p \\ & x_i \in \{0,1\} \quad i \in V. \end{aligned} \tag{2}$$

The function $F(x) = \sum_{ij \in E} w_{ij}(x_i + x_j - x_i x_j)$ satisfies property $(*)$ since $F(x_{ij}(\epsilon))$ is a quadratic polynomial in ϵ such that the coefficient of the quadratic term is nonnegative for each pair of indices i, j and for each point $x \in \mathcal{P}$. Thus, the pipage rounding procedure can be applied to the function $F(x)$.

3. Linear formulations

An integer linear programming formulation of the MVC problem can be easily obtained by introducing two sets of $0-1$ variables: the variables x_i, $i = 1, \dots, n$ whose value is 1 if and only if the node i belongs to the (partial) vertex cover, and the variables z_{ij} whose value can be 1 only if the edge ij is covered. Hence, the optimal solutions of MVC are represented by the solutions of the following integer program:

$$
\begin{array}{lll}
\max w^T z & & \\
\quad \sum_{i \in V} x_i \le p & & \\
\quad x_i + x_j \ge z_{ij} & ij \in E & \\
\quad x_i \in \{0,1\} & i \in V & \\
\quad z_{ij} \in \{0,1\} & ij \in E. &
\end{array}
\tag{3}
$$

where w and z are the vectors whose components are w_{ij} and z_{ij}, respectively.

The *maximum vertex cover polytope*, denoted as $\mathcal{P}_{MVC}(G) \subseteq \Re^{n+m}$, is the convex hull of the integral vectors (x, z) which satisfy the constraints (3). Then, MVC is the problem of maximizing the linear function $w^T z$ over $\mathcal{P}_{MVC}(G)$. This polyhedron is related with the well known vertex cover polyhedron $\mathcal{P}_{VC}(G)$, which is the convex hull of the incidence vectors of vertex covers of a graph. In fact, when $p = n$, the polyhedron $\mathcal{P}_{VC}(G)$ is obtained from $\mathcal{P}_{MVC}(G)$ by projecting onto the subspace of the x variables the face $\mathcal{P}_{MVC}(G) \cap \{z \in \Re^m : z_{ij} = 1, ij \in E\}$.

Proposition 3.1. *$\mathcal{P}_{MVC}(G)$ is full-dimensional (for $p \ge 1$).*

Proof. Let Y be the set of (x, z) vectors composed by: (i) the null vector; (ii) for every node $i \in V$, the vector having $x_i = 1$ and all the other components equal to zero; (iii) for every edge $ij \in E$ and arbitrary node i belonging to it, the vector having $x_i = z_{ij} = 1$ and all the other components equal to zero. It is sufficient to observe that Y constitutes a set of $n+m+1$ affinely independent integral vectors satisfying the constraints of (3). ∎

A natural linear relaxation of MVC is obtained by removing the integrality constraints from (3) and by optimizing over the resulting polyhedron $\mathcal{P}'(G)$. Better formulations can be obtained by adding to the linear system defining $\mathcal{P}'(G)$ valid inequalities for $\mathcal{P}_{MVC}(G)$. In the following, we provide a family of such inequalities, called *clique inequalities*.

Lemma 3.2. *Let Q be a clique of G of size $q \le p+1$ and let h be an integer satisfying $1 \le h \le q-2$. Then the following inequality*

$$
\sum_{ij \in E(Q)} z_{ij} \le h \sum_{i \in V(Q)} x_i + \frac{(q-h)(q-h-1)}{2}
\tag{4}
$$

is facet defining for $\mathcal{P}_{MVC}(G)$.

Proof. We can assume that the clique Q coincides with G: the proof for the general case can be easily obtained by lifting. Hence, we have $q = n$ and $q(q-1)/2 = m$. Let (x, z) be any $0-1$ vector in $\mathcal{P}_{MVC}(G)$ and let $X \subseteq V$ and $Z \subseteq E$ be the sets of nodes and edges whose incidence vectors are x and z, respectively. Observe that Z is a set of edges (not necessarily maximal) covered by the set X of nodes. Inequality (4) then is equivalent to $|Z| - h|X| \leq m - hn + h(h+1)/2$. To show that such an inequality is valid it is sufficient to observe that, by letting $\delta = |X| - n + h$, we have $|Z| \leq m - (h-\delta)(h-\delta-1)/2$, i.e.,

$$|Z| - h|X| \leq m - hn + \frac{h(h+1)}{2} - \frac{\delta(\delta+1)}{2} \leq m - hn + \frac{h(h+1)}{2}.$$

Observe that inequality (4) is satisfied as an equality by the incidence vectors (x, z) of pairs of sets (X, Z) such that $n - h - 1 \leq |X| \leq n - h$ and $Z \equiv Z(X)$, where $Z(X)$ denotes the (maximal) set of edges covered by X. All such vectors belong to $\mathcal{P}_{MVC}(G)$ because $p \geq n-1$.

Let $\mathcal{F}$ be the facet of $\mathcal{P}_{MVC}(G)$ containing all the vectors that satisfy (4) as an equality, and let $\alpha^T x + \beta^T z \leq \gamma$ be a valid inequality which defines $\mathcal{F}$. Since such an inequality is defined modulo a multiplicative positive constant and γ must be positive, without loss of generality we may assume $\gamma = (q-h)(q-h-1)/2 = m - hn + h(h+1)/2$. We will complete the proof by showing that $\alpha^T x + \beta^T z \leq \gamma$ coincides with (4).

Let $\tilde{X} \subseteq V$ be any set of nodes satisfying $|\tilde{X}| = n - h - 1$ and let u, v, w be any three nodes such that $u \in \tilde{X}$, $v, w \notin \tilde{X}$. Observe that $|Z(\tilde{X})| = m - h(h+1)/2$. Let $(\tilde{x}, \tilde{z})$ be the incidence vector of $(\tilde{X}, Z(\tilde{X}))$, $(\tilde{x}', \tilde{z}')$ the incidence vector of $(\tilde{X} \cup \{v\}, Z(\tilde{X} \cup \{v\}))$, $(\bar{x}, \bar{z})$ the incidence vector of $(\tilde{X} - \{u\} \cup \{w\}, Z(\tilde{X} - \{u\} \cup \{w\}))$, $(\bar{x}', \bar{z}')$ the incidence vector of $(\tilde{X} - \{u\} \cup \{v, w\}, Z(\tilde{X} - \{u\} \cup \{v, w\}))$. We have $\alpha^T \tilde{x} + \beta^T \tilde{z} = \alpha^T \tilde{x}' + \beta^T \tilde{z}' = \alpha^T \bar{x} + \beta^T \bar{z} = \alpha^T \bar{x}' + \beta^T \bar{z}' = \gamma$. Hence, we obtain $\alpha^T(\tilde{x} - \tilde{x}' - \bar{x} + \bar{x}') + \beta^T(\tilde{z} - \tilde{z}' - \bar{z} + \bar{z}') = 0$ which implies $\beta_{vu} = \beta_{vw}$. By the arbitraryness of the choice of $\tilde{X}$, u, v, w, we have that all the β coefficients are equal: let b be their common value.

Now, observe that $\alpha^T(\tilde{x}' - \tilde{x}) + \beta^T(\tilde{z}' - \tilde{z}) = 0$ implies

$$\alpha_v + \sum_{i \notin \tilde{X} \cup \{v\}} \beta_{vi} = \alpha_v + hb = 0.$$

Again, by the arbitraryness of the choice of $\tilde{X}$ and v, we get that all the α coefficients are equal to $-hb$. Finally, by $\alpha^T \tilde{x} + \beta^T \tilde{z} = -hb(n-h-1) + b[m - h(h+1)/2] = \gamma = m - hn + h(h+1)/2$ we get

$$(1-b)\left(m - hn + \frac{h(h+1)}{2}\right) = 0.$$

Since $hn - h(h+1)/2 \leq (n-1)(n-2)/2 < n(n-1)/2 = m$, we must have $b = 1$ which completes the proof. ∎

Inequality (4) can be seen as a generalization of the well known clique facet-defining inequality $\sum_{i \in V(Q)} x_i \geq |Q| - 1$ for the vertex cover polyhedron $\mathcal{P}_{VC}(G)$. In fact, such inequality is obtained from (4) by letting $h = 1$ and $z_{ij} = 1$ for $ij \in E(Q)$.

Another generalization of a well known facet-defining inequality for the vertex cover polyhedron is the odd-cycle inequality described in the following.

Lemma 3.3. *Let C be a cycle of G of length $2k+1$ ($k \leq p+1$). Then the following inequality*

$$\sum_{ij \in E(C)} z_{ij} \leq \sum_{i \in V(C)} x_i + k \tag{5}$$

is facet defining for $\mathcal{P}_{MVC}(G)$.

Proof. Again, we can assume that the cycle C coincides with G: the proof for the general case can be easily obtained by lifting. Let (x, z) be any $(0,1)$ vector in $\mathcal{P}_{MVC}(G)$ and let $X \subseteq V$ and $Z \subseteq E$ be the sets of nodes and edges whose incidence vectors are x and z, respectively. Moreover, assume that the nodes of C are ordinately numbered from 0 to $n-1$ and let X be partitioned into sets X_D and X_N, where X_D is the set of "dominant" nodes (let us say that a node $i \in X$ is *dominant* if $i+1 \notin X$, sums taken modulo n). Inequality (5) is equivalent to $|Z| - |X| \leq (n-1)/2$. To show that such an inequality is valid it is sufficient to observe that $|Z| \leq 2|X_D| + |X_N|$, and that $|X_D| \leq (n-1)/2$. Now, to show that inequality (5) is facet-defining, consider the following sets (sums taken modulo n): (i) for any node $i \in V$ let $X_i = \{i, i+2, i+4, \ldots, i+n-1\}$, $Z_i = E$; (ii) for any edge $ij \in E$ (where $j \equiv i+1$ modulo n) let $X_{ij} = \{i+2, i+4, \ldots, i+n-1\}$, $Z_{ij} = E - \{ij\}$. It is easy to observe that the incidence vectors of the pairs of sets (X_i, Z_i), (X_{ij}, Z_{ij}) are $n+m$ affinely independent vectors satisfying inequality (5) as an equality. ∎

In the following section we shall consider the stronger formulation of the MVC problem which is obtained by adding the *clique* inequalities associated with a family $\mathcal{Q}$ of cliques of G to the linear relaxation of (3). We let $i \in V$ ($ij \in E$) be any node (edge) of G. With some abuse of notation, we shall write $i \in Q$ ($ij \in Q$) if the node i (the edge ij) belongs to the clique Q. Thus we have:

$$
\begin{array}{lll}
\max w^T z & & \\
\quad \sum_{i \in V} x_i \leq p & & \\
\quad x_i + x_j \geq z_{ij} & ij \in E & \\
\quad \sum_{ij \in Q} z_{ij} \leq h \sum_{i \in Q} x_i + \frac{(|Q|-h)(|Q|-h-1)}{2} & Q \in \mathcal{Q}, \quad h = 1, \ldots, |Q|-2 & (6) \\
\quad 0 \leq x_i \leq 1 & i \in V & \\
\quad 0 \leq z_{ij} \leq 1 & ij \in E. &
\end{array}
$$

4. Clique-coverable graphs

Let $G = (V, E)$ be a graph with edge weights $w_{ij} \in \Re_+$, $ij \in E$ and let $\mathcal{Q} = \{Q_1, \ldots, Q_t\}$ be a family of cliques of G. The weighted graph (G, w) is said to be covered by $\mathcal{Q}$ if there exist nonnegative numbers $c_1, \ldots, c_t$ such that:

$$w_{ij} = \sum_{Q_h \ni ij} c_h.$$

Moreover, a weighted graph (G, w) is said to be *clique-coverable* if there exists a family of cliques $\mathcal{Q}$ such that (G, w) is covered by $\mathcal{Q}$. If we denote by w and c the vectors whose components are w_{ij} and c_h, respectively, and by A the $0-1$-matrix with $|E|$ columns whose rows are the incidence vectors of the edge-sets of cliques in $\mathcal{Q}$, we can write the above property as $w^T = c^T A$.

Notice that, given a weighted graph (G, w) and a family of cliques $\mathcal{Q}$, it is easy to check whether (G, w) is covered by $\mathcal{Q}$. Moreover, if (G, w) is clique-coverable, then it is covered by a family $\mathcal{Q}$ of polynomial size. In the remaining of the section we assume that (G, w) is clique-coverable.

Consider the following linear program:

$$\begin{aligned}
&\max c^T y \\
&\quad \sum_{i\in V} x_i \le p \\
&\quad y_Q \le h\sum_{i\in Q} x_i + \frac{(|Q|-h)(|Q|-h-1)}{2} \qquad Q\in\mathcal{Q},\quad h=0,1,\dots,|Q|-1 \\
&\quad 0\le x_i\le 1 \qquad\qquad\qquad\qquad\qquad\quad i\in V
\end{aligned} \tag{7}$$

where $y = Az$ is the vector whose components are variables associated with the cliques of $\mathcal{Q}$, and y_Q is meant to represent the number of edges in Q covered by the current solution. Thus, (7) is a relaxation of (6).

For any clique Q of G of size q and any vector x satisfying $0 \le x_i \le 1$, $i \in V$, we define

$$L_Q(x) = \min_{0\le h\le q-1}\left\{ h\sum_{i\in Q} x_i + \frac{(q-h)(q-h-1)}{2}\right\}.$$

Now, if we replace $\sum_{i\in Q} x_i$ by a new variable s, and observe that the minimum in the above expression is attained when $h = q - \lceil s\rceil$, we get

$$\lambda_Q(s) = L_Q(x) = (q-\lceil s\rceil)s + \frac{\lceil s\rceil(\lceil s\rceil - 1)}{2}. \tag{8}$$

Let $L(x)$ denote the vector whose components are $L_Q(x)$, $Q \in \mathcal{Q}$. Since (G, w) is covered by $\mathcal{Q}$, we have:

Proposition 4.1. *The linear program (7) is equivalent to:*

$$\begin{aligned}
&\max c^T y \\
&\quad y = L(x) \\
&\quad \sum_{i\in V} x_i \le p \\
&\quad 0\le x_i\le 1 \quad i\in V.
\end{aligned} \tag{9}$$

Finally, for any clique Q of size q in G, define

$$F_Q(x) = (q-1)\sum_{i\in Q} x_i - \sum_{ij\in Q} x_i x_j,$$

and let $F(x)$ be the vector whose components are $F_Q(x)$, $Q \in \mathcal{Q}$.

To prove our result we need the following theorem of Motzkin and Straus [11] on the maximum of a square-free quadratic form on a simplex.

Theorem 4.2. *Let ω be the size of a maximum clique of a graph $G = (V, E)$ with $|V| = n$ and let $\mathcal{P}_1$ be the polyhedron $\{x \in \Re^n : \sum_{i=1}^n x_i = 1, x_i \ge 0\}$. Then*

$$\max_{x\in\mathcal{P}_1} \sum_{ij\in E} x_i x_j = \frac{1}{2}\frac{(\omega-1)}{\omega}.$$

Let Q be any maximum clique of G. Motzkin and Straus also showed that the maximum of the quadratic form is attained by letting $x_i = 1/\omega$, for $i \in Q$, and $x_i = 0$, otherwise.

Now, let s be a nonnegative real smaller than or equal to ω and let $\mathcal{P}_s$ be the polyhedron $\{x \in \Re^n : \sum_{i=1}^n x_i = s, 0 \leq x_i \leq 1\}$. We generalize the result of Motzkin and Straus as follows:

Lemma 4.3. *Let ω be the size of a maximum clique of a graph $G = (V, E)$ with $|V| = n$ and let $s \leq \omega$. Then*

$$\max_{x \in \mathcal{P}_s} \sum_{ij \in E} x_i x_j = \frac{s^2}{2} \frac{(\omega - 1)}{\omega}.$$

Proof. If $s = 0$ the result is obvious. Otherwise we set $y_i = x_i/s$ and $\mathcal{P}' = \{y \in \Re^n : \sum_{i=1}^n y_i = 1, 0 \leq y_i \leq 1/s\}$. Observe that $y \in \mathcal{P}'$ if and only if $x \in \mathcal{P}_s$ and

$$\max_{x \in \mathcal{P}_s} \sum_{ij \in E} x_i x_j = s^2 \left(\max_{y \in \mathcal{P}'} \sum_{ij \in E} y_i y_j \right).$$

Also, we have $\mathcal{P}' \subseteq \mathcal{P}_1 = \{y \in \Re^n : \sum_{i=1}^n y_i = 1, y_i \geq 0\}$. Thus, by Theorem 4.2, we have

$$\max_{y \in \mathcal{P}'} \sum_{ij \in E} y_i y_j \leq \max_{y \in \mathcal{P}_1} \sum_{ij \in E} y_i y_j = \frac{1}{2} \frac{(\omega - 1)}{\omega}.$$

Moreover, let Q be a clique of G of maximum size. Now, the maximum in $\mathcal{P}_1$ of $\sum_{ij \in E} y_i y_j$ is attained at the point $\bar{y}_i = 1/\omega, i \in Q$ and $\bar{y}_i = 0$ otherwise. Since $\bar{y}$ belongs to $\mathcal{P}'$ the result follows. ∎

Theorem 4.4. *For any clique Q of size $q \geq 3$, $s \leq q$, and any vector $x \in \mathcal{P}_s$, we have*

$$F_Q(x) \geq \frac{2q - 1}{2q} L_Q(x).$$

Proof. To prove the theorem it suffices to show that

$$\min_{x \in \mathcal{P}_s} \frac{F_Q(x)}{L_Q(x)} \geq \frac{2q - 1}{2q}.$$

By (8), we replace $L_Q(x)$ by $\lambda_Q(s)$. Then, since $\min_{x \in \mathcal{P}_s} F_Q(x) = (q-1)s - max_{x \in \mathcal{P}_s} \sum_{ij \in Q} x_i x_j$, we apply Lemma 4.3 and rewrite $\min_{x \in \mathcal{P}_s} F(x)$ as a function of s as well:

$$\min_{x \in \mathcal{P}_s} F_Q(x) = \phi_Q(s) = (q - 1)s - \frac{s^2}{2} \frac{(q - 1)}{q}.$$

We now show that $\phi_Q(s)/\lambda_Q(s) \geq (2q - 1)/2q$, namely that

$$\frac{2q(q - 1)s - s^2(q - 1)}{2q(q - \lceil s \rceil)s + q\lceil s \rceil(\lceil s \rceil - 1)} \geq \frac{2q - 1}{2q}.$$

After some algebraic manipulations, the above inequality can be reduced to the following one:

$$(\lceil s \rceil - 1)(2q - 1)(2s - \lceil s \rceil) \geq 2s(q - 1)(s - 1) \tag{10}$$

If $s \in [0,1]$ then the left hand side of inequality (10) becomes zero while the right hand side is nonpositive, and so the inequality holds.

If $s > 1$ then we replace $\lceil s \rceil$ by $s + fract(s)$ in (10) and we get

$$s(s-1) + fract(s)(2q-1)(1 - fract(s)) \geq 0$$

which clearly holds. Thus the theorem follows. ∎

The following is an immediate consequence of the above theorem.

Theorem 4.5. *Let $G = (V, E)$ be a graph, let $\mathcal{Q}$ be a family of cliques of G and let $F(x)$ and $L(x)$ be functions defined as above; moreover, let $\bar{q}$ be the minimum size of a clique in $\mathcal{Q}$. For any nonnegative vector c and for any vector $\bar{x}$ with $0 \leq x_i \leq 1$, we have*

$$c^T F(\bar{x}) \geq \frac{2\bar{q}-1}{2\bar{q}} c^T L(\bar{x}).$$

Proof. The result follows by noticing that, by Theorem 4.4, for any clique $Q \in \mathcal{Q}$ of size q, by letting $s = \sum_{i \in Q} \bar{x}_i$, the following inequalities hold:

$$F_Q(\bar{x}) \geq \frac{2q-1}{2q} L_Q(\bar{x}) \geq \frac{2\bar{q}-1}{2\bar{q}} L_Q(\bar{x}).$$

Thus, since the vector c is nonnegative, the thesis follows. ∎

We can now state the main result of the paper.

Theorem 4.6. *Let (G, w) be a clique-coverable graph, where $\mathcal{Q}$ is the associated family of cliques; moreover, let $\bar{q}$ be the minimum size of a clique $Q \in \mathcal{Q}$. Let OPT be the optimal value of (3). Then, there exists a polynomial-time algorithm which computes a solution $(\tilde{x}, \tilde{z})$ of (3) such that*

$$w^T \tilde{z} \geq \frac{2\bar{q}-1}{2\bar{q}} OPT.$$

Proof. Let $\bar{x}, \bar{y}$ be an optimal solution of (7), with value $c^T \bar{y}$. Since (7) is a relaxation of (3), its optimal value is not smaller than the optimal value of (3) and can be computed in polynomial time, since $|\mathcal{Q}|$ is polynomial. Moreover, by Proposition 4.1 we have $\bar{y} = L(\bar{x})$ and, by Theorem 4.5, we have

$$c^T F(\bar{x}) \geq \frac{2\bar{q}-1}{2\bar{q}} c^T L(\bar{x}).$$

Note that without loss of generality we may assume that $\bar{x}$ satisfies $\sum_{i \in V} \bar{x}_i = p$. By applying the pipage rounding procedure described in Section 2, we can transform $\bar{x}$ into a $(0,1)$-vector $\tilde{x}$ such that $\sum_{i \in V} \tilde{x}_i = \sum_{i \in V} \bar{x}_i = p$ and $c^T F(\tilde{x}) \geq c^T F(\bar{x})$. Let $\tilde{z}$ be the vector whose components $\tilde{z}_{ij}$ are defined by $\tilde{z}_{ij} = max\{\tilde{x}_i, \tilde{x}_j\}$; then $(\tilde{x}, \tilde{z})$ is a solution of (3). Note that $F(\tilde{x}) = A\tilde{z}$ and so $c^T F(\tilde{x}) = w^T \tilde{z}$. Hence we have

$$w^T \tilde{z} = c^T F(\tilde{x}) \geq c^T F(\bar{x}) \geq \frac{2\bar{q}-1}{2\bar{q}} c^T L(\bar{x}) = \frac{2\bar{q}-1}{2\bar{q}} c^T \bar{y} \geq \frac{2\bar{q}-1}{2\bar{q}} OPT$$

and the theorem follows. ∎

Observe that the approximation factor proved by the above theorem is not worse than 3/4 and for triangle coverable weighted graphs it is already 5/6.

References

[1] N. Apollonio, L. Caccetta and B. Simeone, *Cardinality constrained path covering problems*, Networks. To appear.

[2] N. Apollonio and A. Sëbo, *Minsquare factors and Maxfix cover*, (IPCO 2004), LNCS 3064 (2004), pp. 388-400.

[3] A. A. Ageev and M. I. Sviridenko, *Approximation algorithms for maximum coverage and max cut with given sizes of parts*, (IPCO 1999) LNCS 1610 (1999), pp. 17–30.

[4] M. Conforti and G. Cornuejols, *Submodular set functions, matroids and the greedy algorithm: tight worst-case bounds and some generalizations of the Rado-Edmonds theorem*, Discr. App. Math **7** (1984), pp. 251–274.

[5] G. Cornuejols, G.L. Nemhauser and L. Wolsey, *Worst-case and probabilistic analysis of algorithms for a location problem*, Oper. Research **28** (1980), pp. 847–858.

[6] U. Feige, *A threshold of* $\ln n$ *for approximating set cover*, J. ACM **45** (1998), pp.634–652.

[7] M.X. Goemans and D.P. Williamson, *New* $3/4$*-approximation algorithms for the maximum satisfiability problem*, SIAM J. Disc. Math. **7** (1994), pp.656–666.

[8] Q. Han, Y. Ye, H. Zhang, J. Zhang, *On approximation of max-vertex cover*, Europ. J. Operations Research **143** (2002), pp.342-355.

[9] D. S. Hochbaum, *Approximating covering and packing problems: set cover, vertex cover, independent, and related problems*, in: "Approximation Algorithms for NP-Hard Problems", Dorit S. Hochbaum Eds., PWS Publishing Company, Boston, MA, 1997.

[10] S. Khuller, A. Moss and J. (Seffi) Naor, *The budgeted maximum coverage problem*, Inf. Proc. Letters **70** n. 1 (1999), pp. 39–45.

[11] T.S. Motzkin and E.G. Straus, *Maxima for graphs and a new proof of a theorem of Turán*, Canad. J. Math. **17** (1965), pp. 533–540.

[12] G.L. Nemhauser and L. Wolsey, *Maximizing submodular set functions: formulations and analysis of algorithms*, in: "Studies of Graphs and Discrete Programming", North Holland, Amsterdam (1981), pp.279–301.

[13] R. V. Vohra and N. G. Hall, *A probabilistic analysis of the maximal covering location problem*, Discr. Appl. Math. **43** (1993), pp. 175–183.