Vietnam Journal of Mathematics **38**:2(2010) 157-168

Vietnam Journal
of
MATHEMATICS

Minimum Connected Dominating Sets in Finite Graphs ⋆

Le Cong Thanh

Institute of Mathematics, 18 Hoang Quoc Viet road, 10307 Hanoi, Vietnam

Received July 02, 2009

Abstract. The minimum connected dominating set problem asks for a minimum size subset of vertices with the following property: each vertex is required to be either in the subset, or adjacent to some vertex in the subset, and the subgraph induced by the subset is connected. This problem is known to be *NP*-hard and, for any small $\epsilon > 0$, it cannot be solved by a polynomial time approximation algorithm with the performance ratio less than $(1-\epsilon)\ln|V|$ for any graph $G = (V, E)$ unless $P = NP$.

The present work deals with almost-every-case analysis of a simple greedy algorithm for this problem. We show that for almost every graph instance $G = (V, E)$ of the problem, the greedy algorithm produces a connected dominating set with at most $\log|V|$ vertices and achieves the performance ratio less than $1 + \frac{3\log\log|V|}{\log|V|}$. Thus in almost every-case, the algorithm finds in polynomial time a solution that is extremely close to optimal.

2000 Mathematics Subject Classification: 68Q17.

Key words: Dominating set, greedy algorithm, performance ratio.

1. Introduction

A *dominating set* of a graph is a subset of vertices such that every vertex of the graph is either in the subset, or adjacent to some vertex in the subset, and a *connected* dominating set has an additional condition that the subgraph induced by the dominating set is connected.

⋆ This research was partially supported by the National Foundation for Science and Technology Development (NAFOSTED), Vietnam.

The *minimum connected dominating set* (*Min-CDS*) *problem* is defined as follows: Given a graph $G = (V, E)$, find a minimum size subset D of vertices, such that D forms a dominating set and the subgraph induced by D is connected. This problem is closely related to the *maximum leaf spanning tree problem* [4], and recently it is received much attention in study of wireless networks [1-3]. The terminology is that of [4].

The *Min-CDS* problem is known to be a fundamential *NP*-hard problem in graph theory [4] and moreover, for any small $\epsilon > 0$, it cannot be solved by a polynomial time approximation algorithm with the performance ratio less than $(1-\epsilon)\ln|V|$ for any graph $G = (V, E)$ unless $P = NP$ [6]. Nevertheless, this problem can be approximated by some greedy algorithms with performance ratios $\ln \Delta + 3$ [5] and $\ln \Delta + 2$ [7], where Δ is the maximum degree of a vertex in the graph instance of the problem.

Notice that approximation algorithms are usually evaluated by analysing the performance on some particular instances, and so performance guarantees are in their nature worst-case bounds. Hence the algorithms often behave significantly better in practice than their guarantees would suggest, whereas performance analysis from "almost every-case" point of view gives us information about performance guarantees for approximation algorithms on almost all instances one expects to encounter in practice. The performance analysis of greedy algorithms for some basis problems in graph theory, it has been discussed in [8], tends to confirm our observation that the behaviour of algorithms in almost every-case is generally much better than worst-case behaviour.

The present work deals with almost every-case analysis of a polynomial time natural greedy algorithm for the minimum connected dominating set problem. Namely, we are able to show that its performance ratio is at most $1 + \frac{3\log\log|V|}{\log|V|}$ for almost every graph instance $G = (V, E)$. Thus the greedy algorithm for the *Min-CDS* problem finds in almost every-case a solution that is extremely close to optimal.

In Sec. 2 we shall established a lower bound of the domination number, namely the optimal value for the dominating set problem. The performance of the greedy algorithm for this problem will be analysed in Sec. 3.

2. Dominating Sets and Domination Number

Let $G = (V, E)$ be a graph with vertex set V and edge set E. The *domination number* of G, denoted by $\gamma(G)$, is defined as the minimum cardimality of a dominating set of a graph G.

Let V be a set of n distinguishable vertices. Denote by $\mathcal{G}_n$ the set of all graphs with vertex set V. Clearly the set $\mathcal{G}_n$ has $p = 2^{\binom{n}{2}}$ graphs, and for it we write $\mathcal{G}_n = \{G_i / i = 1, 2, ..., p\}$. Denote also by $\partial_k(G)$ the number of dominating sets with k vertices of G and by $\partial_k(n)$ the mean value of $\partial_k(G)$ over $\mathcal{G}_n$, $1 \le k \le n$, i.e.,

$$\partial_k(n) = \frac{1}{p}\sum_{i=1}^{p}\partial_k(G_i).$$

Lemma 2.1.

$$\partial_k(n) = \binom{n}{k}\left(1-\frac{1}{2^k}\right)^{n-k}.$$

Proof. Denote by $\mathcal{D}_k(n)$ the collection of subsets with k vertices in V. For every graph $G_i \in \mathcal{G}_n$, $1 \le i \le p$, and every set $D_j \in \mathcal{D}_k(n)$, $1 \le j \le q = \binom{n}{k}$, we define the variable $x(G_i, D_j)$ as

$$x(G_i, D_j) = \begin{cases} 1 & \text{if } D_j \text{ is a dominating set of } G_i, \\ 0 & \text{if otherwise.} \end{cases}$$

Then by the definition of $\partial_k(n)$ we have

$$\partial_k(n) = \frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{q} x(G_i, D_j) = \frac{1}{p}\sum_{j=1}^{q}\sum_{i=1}^{p} x(G_i, D_j) = \frac{1}{p}\sum_{j=1}^{q} g(D_j),$$

where $g(D_j)$ is the number of graphs in $\mathcal{G}_n$ such that D_j is their dominating set.

Clearly, for any D_j $(j = 1, 2, \ldots, q)$,

$$g(D_j) = (2^k - 1)^{n-k}.2^{\binom{n}{2}-k(n-k)}.$$

Consequently

$$\partial_k(n) = \frac{q}{p}(2^k-1)^{n-k}.2^{\binom{n}{2}-k(n-k)} = \binom{n}{k}\left(1-\frac{1}{2^k}\right)^{n-k}$$

and so the proof is complete. ■

We now give an upper bound for the mean value $\partial_k(n)$ with a fixed k. As customary for short, $\log x$ denotes the logarithm to the base 2.

Lemma 2.2. *Let* $k_0 = \lfloor \log n - 2\log\log n \rfloor$. *Then*

$$\partial_{k_0}(n) < \frac{1}{n^{\log\log n}},$$

when n *is sufficiently large.*

Proof. Clearly $\binom{n}{k_0} < n^{k_0}$, and $(1-\frac{1}{a})^a < \frac{1}{e} < \frac{1}{2}$ for any positive integer a. Then we have

$$\begin{aligned}\partial_{k_0}(n) &= \binom{n}{k_0}\left(1-\frac{1}{2^{k_0}}\right)^{n-k_0}\\ &< n^{k_0}.2^{-\frac{n-k_0}{2^{k_0}}}\\ &\le n^{k_0}.2^{-(n-k_0)\frac{\log^2 n}{n}}\\ &= n^{k_0-(n-k_0)\frac{\log n}{n}}\\ &\le n^{-2\log\log n+\frac{\log^2 n}{n}-\frac{2\log n.\log\log n}{n}}\\ &\le n^{-\log\log n},\end{aligned}$$

showing the assertion of the lemma. ■

Theorem 2.3. *For almost every graph G any dominating set contains more than $\log n - 2\log\log n$ vertices, where n is the number of vertices of G.*

Proof. Let $\zeta_{n,k}$ be a random variable taking the value d with the probability $H_{n,k}(d)/|\mathcal{G}_n|$, where $H_{n,k}(d)$ is the number of graphs $G \in \mathcal{G}_n$ such that $\partial_k(G) = d$. Denote by $E\zeta_{n,k}$ the expectation of the variable $\zeta_{n,k}$. Then we have

$$E\zeta_{n,k} = \partial_k(n)$$

and so for $t > 0$

$$\text{Prob}(\zeta_{n,k} < t.E\zeta_{n,k}) = \text{Prob}(\partial_k(G) < t.\partial_k(n)).$$

Therefore, applying Markov's inequality $\text{Prob}(\zeta_{n,k} < t.E\zeta_{n,k}) > 1 - \frac{1}{t}$ for the random variable $\zeta_{n,k}$ with $k = k_0$ (where $k_0 = \lfloor \log n - 2\log\log n \rfloor$) and also by Lemma 2.2, we find that

$$\text{Prob}(\partial_{k_0}(G) < t.n^{-\log\log n}) > \text{Prob}(\partial_{k_0}(G) < t.\partial_{k_0}(n)) > 1 - \frac{1}{t}.$$

Let us choose $t = n$. Then $\partial_{k_0}(G) < n^{1-\log\log n} < 1$ when n is sufficiently large. Hence we have

$$\text{Prob}(\partial_{k_0}(G) = 0) > 1 - \frac{1}{n} \longrightarrow 1 \text{ as } n \longrightarrow \infty.$$

This means that almost every graph G has no dominating set with k_0 vertices ($k_0 = \lfloor \log n - 2\log\log n \rfloor$) and so G has no one with less than $\log n - 2\log\log n$ vertices, where n is the number of vertices of the graph G.

Thus the proof is complete. ■

Corollary 2.4. *For almost every graph G with n vertices the domination number $\gamma(G)$ satisfies*

$$\log n - 2\log\log n \le \gamma(G) \le \log n.$$

The lower bound on $\gamma(G)$ is easily obtained from the theorem above. And the upper bound will be obtained by constructing a dominating set with at most

$\log n$ vertices using a greedy algorithm, which will be considered in the next section.

3. Greedy Algorithm for Min-CDS

The *Min-CDS* problem has been introduced in the previous section. This problem is known to be *NP*-hard and can be solved by a simple greedy algorithm in polynomial time. However, as other approximation algorithms for the *Min-CDS* problem, this algorithm in worst-case has the performance ratio greater than $(1-\epsilon)\ln|V|$ (for any small $\epsilon > 0$ and on some graph instance $G = (V, E)$). In this section we investigate almost every-case performance of the greedy algorithm. We will show that the algorithm in almost-case has the performance ratio arbitrarily close to 1.

3.1. Algorithm $Gr|CDS$

To describe the greedy algorithm for the *Min-CDS* problem, it is appropriate to say a few words about the terminology. A *star* is a graph in which some vertex is incident to each of the edges of the graph. Equivalently, a star consists of a vertex designated *center* along with a set of *leaves* adjacent to it, and also consists of all edges that join the center and the leaves. We say that the center of a star *covers* its leaf vertices

The natural idea behind the *greedy algorithm for Min-CDS*, donoted $Gr|CDS$, is the following: Grow a tree T, starting from the star with the maximum number of leaves in a given graph. At each step (after the first) we will pick a star that consists of a center in T along with the maximum number of leaves outside T. In the end we will find a spanning tree (if the graph is connected), and will pick the non-leaf vertices, or the star centers, as the connected dominating set.

Let T_i be the tree constructed by the algorithm at Step i, $i = 1, 2, \ldots$. Then vertices of T_i are *dominated* and vertices outside T_i are *undominated* at the time. Thus the algorithm $Gr|CDS$ consists in finding a connected dominating set by starting with some vertex that covers the maximum number of vertices and iteratively adding dominated vertices that covers the maximum number of undominated vertices until it is no longer possible.

3.2. Preliminaries

Given a graph $G = (V, E)$ with $|V| = n$ and a non-empty subset $U \subseteq V$ with $|U| = m$, $1 \le m \le n$. Let $1 \le k \le m-1$ and $1 \le h \le n-m$. Throughout the section we use the following notations:

$Z^{k,h}(u)$ - the star that has a center $u \in U$, and that consists of k leaves in U and h leaves outside U;

$\mathcal{S}_U^{k,h}(G)$ - the number of stars of $G = (V, E)$ that have centers in U and that consist of k leaves in U and h leaves outside U;

$\mathcal{S}_U^{\mathcal{B}}(G)$ - the number of stars of G that have centers in U, and that contain at least $\frac{m}{2}$ leaves in U and at least $\frac{n-m}{2}$ leaves outside U (the number of *"bright"* stars of G);

$\mathcal{S}_U^{\mathcal{B}}(n)$ - the mean value of $\mathcal{S}_U^{\mathcal{B}}(G)$ over $\mathcal{G}_n$;

$\mathcal{C}_U^{\mathcal{B}}(n)$ - the collection of stars that have centers in U, and that contain at least $\frac{m}{2}$ leaves in U and at least $\frac{n-m}{2}$ leaves outside U.

Put $k^* = \lceil \frac{m}{2} \rceil$ and $h^* = \lceil \frac{n-m}{2} \rceil$. Then it is easy to see that the collection $\mathcal{C}_U^{\mathcal{B}}(n)$ consists of s such stars, where

$$s = m \sum_{k=k^*}^{m-1} \sum_{h=h^*}^{n-m} \binom{m-1}{k} \binom{n-m}{h}.$$

We calculate now the mean value $\mathcal{S}_U^{\mathcal{B}}(n)$.

Lemma 3.1.

$$\mathcal{S}_U^{\mathcal{B}}(n) = \frac{m}{2^{n-1}} \sum_{k=k^*}^{m-1} \sum_{h=h^*}^{n-m} \binom{m-1}{k} \binom{n-m}{h}.$$

Proof. As in the proof of Lemma 2.1, we have

$$\mathcal{S}_U^{\mathcal{B}}(n) = \frac{1}{p} \sum_{i=1}^{p} \mathcal{S}_U^{\mathcal{B}}(G_i) = \frac{1}{p} \sum_{j=1}^{s} g^*(Z_j^{k,h}(u)),$$

where $g^*(Z_j^{k,h}(u))$ is the number of graphs in $\mathcal{G}_n$ that contain $Z_j^{k,h}(u)$ as a maximal star, i.e. in the graphs no orther star properly contains it. Clearly, for each star $Z_j^{k,h}(u)$ with $j = 1, 2, ..., s$,

$$g^*(Z_j^{k,h}(u)) = 2^{\binom{n}{2} - (n-1)}.$$

Therefore

$$\mathcal{S}_U^{\mathcal{B}}(n) = \frac{1}{p} \sum_{j=1}^{s} g^*(Z_j^{k,h}(u)) = \frac{s}{p} 2^{\binom{n}{2} - (n-1)} = \frac{s}{2^{n-1}},$$

implying the assertion of the lemma. ■

Let $\xi_{U,n}$ be a random variable taking the value r with the probability $\mathcal{H}_{U,n}(r)/|\mathcal{G}_n|$, where $\mathcal{H}_{U,n}(r)$ is the number of graphs $G \in \mathcal{G}_n$ such that $\mathcal{S}_U^{\mathcal{B}}(G) = r$. Then we find the expectation of $\xi_{U,n}$ that

$$E\xi_{U,n} = \mathcal{S}_U^{\mathcal{B}}(n) = \frac{m}{2^{n-1}} \sum_{k=k^*}^{m-1} \sum_{h=h^*}^{n-m} \binom{m-1}{k} \binom{n-m}{h}.$$

Lemma 3.2. *Let U be a subset of m vertices of V such that $\frac{n}{2} \leq m \leq n$. Then*

$$\operatorname{Var}\xi_{U,n} \leq \frac{2}{\log^2 n}(E\xi_{U,n})^2.$$

Proof. By definition $\operatorname{Var}\xi_{U,n} = E(\xi_{U,n})^2 - (E\xi_{U,n})^2$. In order to find an upper bound for the variance $\operatorname{Var}\xi_{U,n}$, we now estimate the expectation $E(\xi_{U,n})^2$.

Similarly to the proof of Lemma 2.1, it is easily shown that

$$E(\xi_{U,n})^2 = \frac{1}{p}\sum_{i=1}^{p}(\mathcal{S}_U^{\mathcal{B}}(G_i))^2 = \frac{1}{p}\sum_{j,l=1}^{s} g^*(Z_j^{k_1,h_1}(u_1), Z_l^{k_2,h_2}(u_2)),$$

where the summation is over all ordered star pairs $(Z_j^{k_1,h_1}(u_1), Z_l^{k_2,h_2}(u_2))$ in $\mathcal{C}_U^{\mathcal{B}}(n) \times \mathcal{C}_U^{\mathcal{B}}(n)$, and $g^*(Z_j^{k_1,h_1}(u_1), Z_l^{k_2,h_2}(u_2))$ is the number of graphs in $\mathcal{G}_n$ that contain both stars $Z_j^{k_1,h_1}(u_1)$ and $Z_l^{k_2,h_2}(u_2)$ as maximal.

Now let us examine the three possibilities of the pairs of the stars $Z_j^{k_1,h_1}(u_1)$ and $Z_l^{k_2,h_2}(u_2)$:
(a) *The stars have the common center $u = u_1 = u_2$.*
(b) *The stars have no centers in common, i.e., $u_1 \neq u_2$, but have the common edge, namely the edge with endpoints u_1 and u_2.*
(c) *The stars have no centers and also edges in common.*

Put $F_{(x)} = \frac{1}{p}\sum_{(x)} g^*(Z_j^{k_1,h_1}(u_1), Z_l^{k_2,h_2}(u_2))$, where $\sum_{(x)}$ means that the summation is over all pairs $(Z_j^{k_1,h_1}(u_1), Z_l^{k_2,h_2}(u_2))$ of stars that satisfy the above conditions (x) for $x = a, b, c$. Then, according to the above formula for $E(\xi_{U,n})^2$, we have

$$E(\xi_{U,n})^2 = F_{(a)} + F_{(b)} + F_{(c)}. \tag{1}$$

We now calculate and estimate the $F_{(a)}$, $F_{(b)}$ and $F_{(c)}$ on the assumption that $\frac{n}{2} \leq m \leq n$.

(A) By the definition of $g^*(Z_j^{k_1,h_1}(u_1), Z_l^{k_2,h_2}(u_2))$ we have

$$F_{(a)} = \begin{cases} \frac{m}{2^{n-1}} \sum\limits_{k=k^*}^{m-1} \sum\limits_{h=h^*}^{n-m} \binom{m-1}{k}\binom{n-m}{h} & \text{if } Z_j^{k_1,h_1}(u_1) = Z_l^{k_2,h_2}(u_2), \\ 0 & \text{if otherwise.} \end{cases}$$

Clearly $\sum\limits_{k=k^*}^{m-1}\binom{m-1}{k} > 2^{m-3}$ with $k^* = \lceil \frac{m}{2} \rceil$ and $\sum\limits_{h=h^*}^{n-m}\binom{n-m}{h} > 2^{n-m-2}$ with $h^* = \lceil \frac{n-m}{2} \rceil$. By the assumption $m \geq \frac{n}{2}$, it is easily shown that

$$F_{(a)} = E\xi_{U,n} \geq \frac{n}{32},$$

and so

$$F_(a) \leq \frac{32}{n}(E\xi_{U,n})^2.$$

(B) Now let us estimate the $F_{(b)}$. By the definition of $F_{(b)}$ and since $\binom{m-2}{k_2-1} = \binom{m-1}{k_2}\frac{k_2}{m-1}$ we obtain that

$$F_{(b)} = \frac{1}{p}\sum_{k_1,k_2=k^*}^{m-1}\sum_{h_1,h_2=h^*}^{n-m} m\binom{m-1}{k_1}\binom{n-m}{h_1}k_1\binom{m-2}{k_2-1}\binom{n-m}{h_2}.2^{\binom{n}{2}-2(n-1)+1}$$
$$= \frac{2m}{(m-1).2^{2(n-1)}}\sum_{k_1,k_2=k^*}^{m-1}\sum_{h_1,h_2=h^*}^{n-m}\binom{m-1}{k_1}k_1\binom{n-m}{h_1}\binom{m-1}{k_2}k_2\binom{n-m}{h_2}.$$

Let us consider the summation $\sum_{k=k^*}^{m-1}\binom{m-1}{k}k$. Clearly

$$\sum_{k=k^*}^{m-1}\binom{m-1}{k}k = \sum_{i=0}^{m-k^*-1}\binom{m-1}{k^*+i}(k^*+i)$$
$$= k^*\sum_{k=k^*}^{m-1}\binom{m-1}{k} + \sum_{i=1}^{m-k^*-1}\binom{m-1}{k^*+i}i.$$

We shall show that

$$\sum_{i=1}^{m-k^*-1}\binom{m-1}{k^*+i}i \le \frac{m}{\log^2 m}\sum_{k=k^*}^{m-1}\binom{m-1}{k}, \tag{2}$$

and so

$$\sum_{k=k^*}^{m-1}\binom{m-1}{k}k \le \left(\frac{m+1}{2} + \frac{m}{\log^2 m}\right)\sum_{k=k^*}^{m-1}\binom{m-1}{k}, \tag{3}$$

since $k^* \le \frac{m+1}{2}$. Indeed putting $A_i = \binom{m-1}{k^*+i}i$ for $i = 1, 2, \ldots, m-k^*-1$. It is easy to see that $A_{i-1} < A_i$ for $1 < i \le \frac{\sqrt{m}-1}{2}$ and $A_i > A_{i+1}$ for $\frac{\sqrt{m}-1}{2} \le i \le m-k^*$, and so

$$\max\left\{\binom{m-1}{k^*+i}i : i = 1, 2, \ldots, m-k^*-1\right\} = \binom{m-1}{k^*+i_0}i_0,$$

where $i_0 = \left\lceil\frac{\sqrt{m}-1}{2}\right\rceil$. Now put $B_i = \binom{m-1}{k^*+i}\frac{m}{\log^2 m}$ for $i = 0, 1, 2, \ldots, m-k^*-1$. Clearly $B_i \ge B_{i+1}$ and $B_0 \ge B_1 \ge B_2 \ge \ldots \ge B_{i_0} \ge A_{i_0}\left\lceil\frac{2\sqrt{m}}{\log^2 m}\right\rceil$. Hence

$$\sum_{i=0}^{i_0} B_i \ge \sum_{i=1}^{i_1} A_i,$$

where $i_1 = \left\lceil \frac{m}{\log^2 m} \right\rceil$ and it is less than $(i_0+1)\left\lceil \frac{2\sqrt{m}}{\log^2 m} \right\rceil$. Furthermore, it is obvious that $B_{i_0+j} \geq A_{i_1+j}$ for $j = 1, 2, \ldots, m - i_1$. Then, combining the last two inequalities, we have (2).

Thus, according to the above formula for $F_{(b)}$, applying inequality (3) with $k = k_1, k_2$, and by the assumption $m \geq \frac{n}{2}$, we find that

$$F_{(b)} \leq \frac{2}{m(m-1)} \left(\frac{m+1}{2} + \frac{m}{\log^2 n} \right)^2 (E\xi_{U,n})^2 \leq \left(\frac{1}{2} + \frac{3}{\log^2 n} \right) (E\xi_{U,n})^2.$$

(C) By the definition of $F_{(c)}$, since the equality $\binom{m-2}{k} = \binom{m-1}{k} \frac{m-k-1}{m-1}$ and the inequality $m - k - 1 \leq \frac{m}{2}$ hold for each k such that $\left\lceil \frac{m}{2} \right\rceil = k^* \leq k \leq m-1$, and since $m \geq \frac{n}{2}$, we have

$$\begin{aligned}
F_{(c)} &= \frac{1}{p} \sum_{k_1,k_2=k^*}^{m-1} \sum_{h_1,h_2=h^*}^{n-m} m \binom{m-1}{k_1} \binom{n-m}{h_1} (m-k_1-1) \binom{m-2}{k_2} \binom{n-m}{h_2} \times \\
&\quad \times 2^{\binom{n}{2}-2(n-1)+1} \\
&\leq \frac{m}{2(m-1)} . \frac{m^2}{2^{2(n-1)}} \sum_{k_1,k_2=k^*}^{m-1} \sum_{h_1,h_2=h^*}^{n-m} \binom{m-1}{k_1} \binom{n-m}{h_1} \binom{m-1}{k_2} \binom{n-m}{h_2} \\
&= \frac{m}{2(m-1)} (E\xi_{U,n})^2 \\
&\leq \left(\frac{1}{2} + \frac{1}{n} \right) (E\xi_{U,n})^2.
\end{aligned}$$

Finally, according to (1) and combining (A)-(C), when n is sufficiently large we obtain that

$$\begin{aligned}
E(\xi_{U,n})^2 &\leq \left(\frac{32}{n} + 1 + \frac{3}{\log^2 n} + \frac{1}{n} \right) . (E\xi_{U,n})^2 \\
&\leq \left(1 + \frac{4}{\log^2 n} \right) . (E\xi_{U,n})^2,
\end{aligned}$$

so

$$\operatorname{Var} \xi_{U,n} = E(\xi_{U,n})^2 - (E\xi_{U,n})^2 \leq \frac{4}{\log^2 n} (E\xi_{U,n})^2.$$

Thus the proof is complete. ■

Theorem 3.3. *Let U be a subset of vertices of a graph $G = (V, E)$ with $|V| = n$ such that $|U| \geq \frac{n}{2}$. Then*

$$\operatorname{Prob}(\mathcal{S}_U^{\mathcal{B}}(G) \geq 1) \geq 1 - \left(\frac{2 \log\log n}{\log n} \right)^2.$$

Proof. Applying Chebyshev's inequality for the variable $\xi_{U,n}$ and t such that $0 < t < E\xi_{U,n}$, we have

$$\mathrm{Prob}(\xi_{U,n} = 0) \leq \mathrm{Prob}(|\xi_{U,n} - E\xi_{U,n}| \geq t) \leq \frac{\mathrm{Var}\,\xi_{U,n}}{t^2}.$$

Then, by choosing $t = \frac{E\xi_{U,n}}{\log\log n}$ and also by Lemma 3.2, we obtain that

$$\mathrm{Prob}(\xi_{U,n} = 0) \leq \left(\frac{2\log\log n}{\log n}\right)^2.$$

Therefore, by the definitions of $\xi_{U,n}$ and $\mathcal{S}_U^{\mathcal{B}}(G)$, we have

$$\mathrm{Prob}(\mathcal{S}_U^{\mathcal{B}}(G) = 0) \leq \left(\frac{2\log\log n}{\log n}\right)^2,$$

implying the assertion of the theorem. ■

Note that for $U = V$, $S_V^{\mathcal{B}}(G)$ is the number of stars of $G = (V, E)$ with $|V| = n$ that contain at least $\frac{n}{2}$ leaves in V. (In other words, $S_V^{\mathcal{B}}(G)$ is the number of vertices of the maxinum degree at least $\frac{n}{2}$ in G). Now, for each subset $U \subset V$ such that $|U| = m$, let us denote by $S_U^{\mathcal{B}.o}(G)$ *the number of stars of* G *that have centers in* U *and contain at least* $\frac{n-m}{2}$ *leaves outside* U. It is obvious that the assertions $S_V^{\mathcal{B}}(G) \geq 1$ and $S_U^{\mathcal{B}.o}(G) \geq 1$ concern the performance of the greedy algorithm $Gr|CDS$. Since $S_U^{\mathcal{B}.o}(G) \geq \mathcal{S}_U^{\mathcal{B}}(G)$, Theorem 3.3 has the following immediate consequence.

Corollary 3.4. *Let* U *be a subset of vertices of a graph* $G = (V, E)$ *with* $|V| = n$ *such that* $\frac{n}{2} \leq |U| < n$. *Then*

(i) $\mathrm{Prob}(S_V^{\mathcal{B}}(G) \geq 1) \geq 1 - (\frac{2\log\log n}{\log n})^2$,

(ii) $\mathrm{Prob}(S_U^{\mathcal{B}.o}(G) \geq 1) \geq 1 - (\frac{2\log\log n}{\log n})^2$.

3.3. Performance of Algorithm $Gr|CDS$

The object of this subsection is to analyse the algorithm $Gr|CDS$ in almost every-case. The analysis will be based on Corollary 3.4.

Let us return to the algorithm $Gr|CDS$ described in Subsec. 3.1. As we noted earlier, the algorithm $Gr|CDS$ finds at the first step a star with the maximum number of leaves in a given graph, namely it finds a vertex that covers the maximum number of undominated vertices, and at each step (after the first) it finds a dominated vertex that covers the maximum number of undominated vertices. Clearly, the last value closely relates with the performance of the algorithm. As a matter of fact, the algorithm $Gr|CDS$ will be quickly terminated if it can find at each step a vertex that covers as much as possible the undominated vertices. This motivates us to do the following definition.

The algorithm $Gr|CDS$ is said to be *well-performance* on an instance G if at each step of the algorithm for G it finds a vertex that covers at least half the undominated vertices.

Theorem 3.5. *For almost every graph instance $G = (V, E)$ of the Min-CDS problem, the algorithm $Gr|CDS$ is well-performance and so terminated after no more than $\log|V|$ steps.*

Proof. Given a graph G with vertex set V of n elements. Let us consider a family $\mathcal{F}_n$ of subsets U_i for $i = 1, 2, \ldots$, such that $U_1 \subset U_2 \subset \ldots \subset U_i \subset \ldots \subset V$, and $U_1 \geq \frac{n}{2}$ and $|U_{i+1}| \geq |U_i| + \frac{n-|U_i|}{2}$.

Since $n > |U_i| \geq n(1 - \frac{1}{2^i})$ for each $i = 1, 2, \ldots$, the family $\mathcal{F}_n$ consists of at most $\log n$ subsets and so every U_i in $\mathcal{F}_n$ satisfies the condition of Corollary 3.4, namely $\frac{n}{2} \leq |U_i| < n$. Then applying (ii) of Corollary 3.4 to U_i for $i = 1, 2, \ldots$, we obtain that

$$\text{Prob}(\mathcal{S}_{U_i}^{\mathcal{B},o}(G) \geq 1) \geq 1 - \left(\frac{2\log\log n}{\log n}\right)^2.$$

Now to complete the proof it suffices to show that asssertion (i) of Corollary 3.4 and the above assertions are all true at the same time. Indeed, since the family $\mathcal{F}_n$ has at most $\log n$ subsets, we obtain that

$$\text{Prob}(\mathcal{S}_V^{\mathcal{B}}(G) \geq 1 \ \&\ \mathcal{S}_{U_i}^{\mathcal{B},o}(G) \geq 1, \forall U_i \in \mathcal{F}_n) \geq 1 - \frac{(2\log\log n)^2}{\log n}.$$

Clearly the last probability tends to 1 as $n \longrightarrow \infty$. This shows that, for almost every problem instance, vertex sets of the trees constructed by the algorithm $Gr|CDS$ at each step form a family such as $\mathcal{F}_n$, and so the theorem is proved. ∎

Let now $R_{Gr|CDS}(G)$ be the performance ratio of the greedy algorithm $Gr|CDS$ on a graph instance G. It is defined by

$$R_{Gr|CDS}(G) = \frac{Gr|CDS(G)}{OPT_{CDS}(G)},$$

where $Gr|CDS(G)$ is the value of the solution found by $Gr|CDS$ when applied to G, and $OPT_{CDS}(G)$ is the optimal value of the *Min-CDS* problem for G.

In order to estimate the ratio $R_{Gr|CDS}(G)$, we note that for almost every graph $G = (V, E)$, the value $Gr|CDS(G)$ is smaller than $\log|V|$, since the dominating set found by the algorithm $Gr|CDS$ consists of all centers of the stars which are picked at each step and also the algorithm is terminated after no more than $\log|V|$ steps. Note now that $OPT_{CDS}(G) \geq \gamma(G)$, the domination number of G. Hence the lower bound on $\gamma(G)$, which was presented in Corollary 2.4, follows that the value $OPT_{CDS}(G)$ is more than $\log|V| - 2\log\log|V|$. Then the upper bound on $R_{Gr|CDS}(G)$ is easily obtained for almost every G.

In conclusion, we have the following result.

Theorem 3.6. *For almost every graph instance* $G = (V, E)$ *of the Min-CDS problem, the following assertions hold:*

(i) *The algorithm* $Gr|CDS$ *finds a connected dominating set with at most* $\log |V|$ *vertices, and*

(ii) *The performance ratio* $R_{Gr|CDS}(G)$ *of the algorithm on an instance* G *is less than* $1 + \frac{3 \log \log |V|}{\log |V|}$.

Thus, for almost every instance of the problem, the algorithm $Gr|CDS$ finds a solution that is extremely close to optimal. In particular, from assertion (i) above we easily get the upper bound on the domination number $\gamma(G)$, which was formulated in Corollary 2.4.

References

1. K. M. Alzoubi, P.-J. Wan, and O. Frieder, New distributd algorithm for connected dominating set in wireless ad hoc networks, in: *Proc. HICSS, Hawaii*, 2002, pp. 3881-3887.
2. X. Cheng, X. Huang, D. Li, W. Wu, and D.-Z. Du, Polynomial-time approximation scheme for connected dominating set in ad hoc wireless networks, *Network*, **42** (4) (2003), 202–208.
3. B. Das and V. Bharghavan, *Routing in ad hoc Networks Using Minimum Connected Dominating Sets*, ICC' 97, Montreal, Canada, 1997.
4. M. R. Garey and D. S. Johnson, *Computers and Intractability - A Guide to Theory of NP-completeness*, W. H. Freeman, San Fransico, 1979.
5. S. Guha and S. Khuller, Approximation algorithms for connected dominating sets, *Algorithmica* **20** (1998), 374–387.
6. C. Lund and M. Yannakakis, On the hardness of approximating minimization problems, *J. ACM*, **41** (5) (1994), 960–981.
7. L. Ruan, H. Du, X. Jia, W. Wu, Y. Li, and K.-I Ko, A greedy approximation for minimum connected dominating sets, *Theoretical Computer Science* **329** (2004), 325–330.
8. L. C. Thanh, Performance analysis of greedy algorithms for Max-IS and Min-Maxl-Match, *Vietnam J. Math.* **36** (3) (2008), 327–336.