Discrete Mathematics 118 (1993) 23–31
North-Holland

Coloring a graph optimally with two colors

H.J. Broersma and F. Göbel

Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, Netherlands

Received 15 November 1989
Revised 31 October 1991

Abstract

Broersma, H.J. and F. Göbel, Coloring a graph optimally with two colors, Discrete Mathematics 118 (1993) 23–31.

Let G be a graph with point set V. A (2-)*coloring* of G is a map of V to {red, white}. An *error* occurs whenever the two endpoints of a line have the same color. An *optimal coloring* of G is a coloring of G for which the number of errors is minimum. The minimum number of errors is denoted by $\gamma(G)$, we derive upper and lower bounds for $\gamma(G)$ and prove that if G is a graph with n points and m lines, then $\max\{0, m-\lfloor \frac{1}{4}n^2 \rfloor\} \leqslant \gamma(G) \leqslant \lfloor \frac{1}{2}m - \frac{1}{4}(h(m)-1) \rfloor$, where $h(m)=\min\{n \mid m \leqslant \binom{n}{2}\}$. The lower bound is sharp, and for infinitely many values of m the upper bound is attained for all sufficiently large n.

1. Introduction

Suppose the specification of an electrical network has the following property: the components can be partitioned into two classes A and B such that all connections are between a component of class A and a component of class B. In other words, the underlying graph is bipartite. In such a case the network can be built as follows. Take a thin rectangular plate of insulating material and put all components of class A on one face, in a row near the edge, and all components of class B on the other face, in a row near a perpendicular edge; see Fig. 1.

If wires are laid as indicated in Fig. 1, a connection between two components can now be made by drilling a hole at the crossing-point of the corresponding wires and filling it with conductive material.

This way of realising a network is attractive. The problem is the fact that so few underlying graphs are bipartite. Now any graph can be made a bipartite graph by

Correspondence to: H.J. Broersma, Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500, AE Enschede, Netherlands.

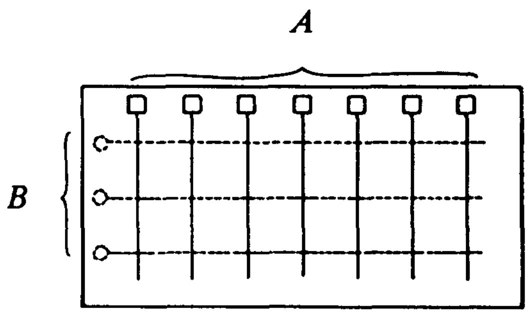

Fig. 1.

subdividing a number of suitable lines (e.g. *all* lines). Obviously one likes to have this number minimum, since each subdivision requires extra material and extra space.

So we have a purely graph theoretic problem: given a graph $G=(V,E)$ color the points with two colors such that the number of 'errors' (lines with equally colored endpoints) is minimum; call this minimum $\gamma(G)$.

It has been shown that this problem is NP-complete (see [2, p. 196]). Hence, in cases where the graph is large, it is reasonable to consider heuristic algorithms for solving the optimisation problem. In this paper, we only consider upper and lower bounds for the minimum number of errors, in terms of simple parameters of the graph.

An (n,m)-graph is a graph with $n=n(G)$ points and $m=m(G)$ lines. All graph theoretical terms not defined here can be found in [3]. The largest integer not exceeding the real number x is denoted by $\lfloor x \rfloor$; the smallest integer not smaller than x is denoted by $\lceil x \rceil$. The set of natural numbers is the set $\{1,2,3,\dots\}$.

2. Simple properties and the lower bound

In some cases the determination of $\gamma(G)$ can be reduced to determining γ for some smaller graphs.

Proposition 2.1. *If $G_1,\dots,G_k$ are the blocks of G, then $\gamma(G)=\sum_{i=1}^{k}\gamma(G_i)$.*

Proof. First determine an optimal coloring for each of the blocks separately. Then put the blocks together, changing if necessary the coloring of a block to the complementary coloring. □

Our next result is a convexity result.

Proposition 2.2. *If n and m are fixed, and G runs through all (n,m)-graphs, then $\gamma(G)$ takes all values between its minimum and its maximum.*

Proof. Let G^+ and G^- be (n,m)-graphs for which γ is maximum and minimum, respectively. In both graphs, label the points $1, 2, \ldots, n$ (arbitrarily). If $G^+ \neq G^-$, there is a pair i, j such that $\{i, j\}$ is a line of G^+, not of G^-. Similarly, G^- has a line $\{p, q\}$ not in G^+. Consider the graph $G^0 = G^+ - \{i, j\} + \{p, q\}$, and compare $\gamma(G^+)$ and $\gamma(G^0)$. It is easy to see that $\gamma(G^0) = \gamma(G^+)$ or $\gamma(G^0) = \gamma(G^+) - 1$. Now we repeat this process which gradually changes G^+ into G^-. Since γ never drops by 2 or more (although it can increase by 1), all intermediate values have been assumed when we reach G^-. □

The following result gives a lower bound for $\gamma(G)$.

Proposition 2.3. *$\gamma(G) \geq m - \lfloor \frac{1}{4}n^2 \rfloor$ and for all n, m with $m - \lfloor \frac{1}{4}n^2 \rfloor \geq 0$, there is an (n, m)-graph G_0 with $\gamma(G_0) = m - \lfloor \frac{1}{4}n^2 \rfloor$.*

Proof. The largest number of lines in a bipartite graph on n points is $\lfloor \frac{1}{2}n \rfloor \lceil \frac{1}{2}n \rceil = \lfloor \frac{1}{4}n^2 \rfloor$ and so $\gamma(G) \geq m - \lfloor \frac{1}{4}n^2 \rfloor$. If now $m - \lfloor \frac{1}{4}n^2 \rfloor \geq 0$, let G_0 be obtained from $K_{\lfloor \frac{1}{2}n \rfloor, \lceil \frac{1}{2}n \rceil}$ by adding $m - \lfloor \frac{1}{4}n^2 \rfloor$ arbitrary lines. Then $\gamma(G_0) = m - \lfloor \frac{1}{4}n^2 \rfloor$. □

Remark. The above result implies that $\max\{0, m - \lfloor \frac{1}{4}n^2 \rfloor\}$ is the true minimum of $\gamma(G)$ over all (n, m)-graphs G, so the lower bound is sharp in that (strong) sense. We shall see that the situation is not so simple for upper bounds.

3. An upper bound in terms of spanning trees

If G is a connected graph and T is a spanning tree of G, then $\xi(G, T)$ denotes the number of odd fundamental cycles in the cycle basis with respect to T.

Proposition 3.1. *Let G be a connected graph and T be a spanning tree of G. Then $\gamma(G) \leq \xi(G, T)$.*

Proof. Since T is bipartite, the points of T (i.e. the points of G) can be colored such that there are no errors among the lines of T. Using such a coloring, the chords leading to an even cycle are also without error. Hence the result. □

The next result is not a bound, but it is closely related to the previous result, and it is of interest since it leads to some heuristics.

Proposition 3.2. *If G is connected, then G has a spanning tree T_0 such that $\gamma(G) = \xi(G, T_0)$.*

Proof. Color $V(G)$ optimally with 2 colors. Consider the lines of G with differently colored end-points. These lines, together with possible isolated points of G induce a spanning subgraph H of G without errors. The graph H contains a spanning forest. Now suppose this forest has 2 or more components. All lines between the components are errors! This means that a change to the complementary coloring in one component decreases the number of errors, contradicting the optimality of the original coloring. Hence the spanning forest is in fact a spanning tree. It is easy to see that this spanning tree can serve as T_0. □

4. An upper bound in terms of the chromatic number

As Erdős [1] pointed out, it is easy to see that $\gamma(G) \leqslant \lfloor \frac{1}{2}m \rfloor$ for every (n, m)-graph G. Indeed, in a coloring of G with fewest errors, each point is adjacent to at least as many points of the opposite color as of the same color, or else we would reduce the number of errors by changing its color. We now obtain an improvement to Erdős's bound in terms of the chromatic number χ of G; we shall obtain a different improvement in the next section.

Proposition 4.1. *Let* $k = \lfloor \frac{1}{2}(\chi - 1) \rfloor$. *Then* $\gamma(G) \leqslant km/(2k+1)$.

Proof. Suppose χ is even, so $\chi = 2k+2$. Color $V(G)$ properly with χ colors $1, 2, \ldots, \chi$. Let the number of points in color class i be n_i $(i = 1, \ldots, \chi)$ and let the number of lines between classes i and j be m_{ij}. Now merge the color classes in two groups R and W, each being the union of $k+1$ original color classes, and color the points of R red and the points of W white. The number of errors is now

$$\sum_{i,\, j \in R} m_{ij} + \sum_{i,\, j \in W} m_{ij}.$$

There are $\frac{1}{2}\binom{2k+2}{k+1}$ ways to form the classes R and W. The total number of errors over all possible ways of forming R and W is $\sum_{i,j} m_{ij}\binom{2k}{k-1}$, since there are $\binom{2k}{k-1}$ cases where colors 1 and 2 are in the same class, etc.

Now $\sum_{i,j} m_{ij} = m$, hence the average number of errors after forming R and W is

$$\frac{m\binom{2k}{k-1}}{\frac{1}{2}\binom{2k+2}{k+1}},$$

which reduces to $km/(2k+1)$. Hence $\gamma(G) \leqslant km/(2k+1)$.

The proof for odd χ is similar. □

As an application, consider a network in which 3 types of components can be distinguished such that only connections between components of unequal types occur. For the underlying graph this means $\chi \leqslant 3$, and in the nontrivial case where $\chi = 3$ we have $\gamma(G) \leqslant \lfloor \frac{1}{3}m \rfloor$.

5. An upper bound in terms of *n*, *m* and *c*

This section is devoted to a proof of the following proposition.

Proposition 5.1. *Let G be an (n, m)-graph with c components and $\gamma(G)=\gamma$. Then* $\gamma \leqslant \lfloor \frac{1}{2}m - \frac{1}{4}(n-c) \rfloor$.

Proof. Note that $\lfloor \;\; \rfloor$ is a superadditive function: $\lfloor a+b \rfloor \geqslant \lfloor a \rfloor + \lfloor b \rfloor$. Also, γ, m and $n-c$ are all additive over disjoint graphs. Thus it suffices to prove the result $\gamma \leqslant \lfloor \frac{1}{2}m - \frac{1}{4}(n-1) \rfloor$ for a connected graph G.

Also, by Proposition 2.1, γ, m and $n-1$ are additive over blocks. Thus it suffices to prove the result for a block. Since the result obviously holds for K_2 (the only possible block that is not 2-connected), it suffices to prove the result for a 2-connected graph G.

So suppose G is 2-connected, let f be the number of errors in a coloring of G, and let g be the number of 'good' lines, so that $f+g=m$. We will show by induction on n that $g \geqslant f + \lfloor \frac{1}{2}n \rfloor$ in each optimal coloring of G. For $n \leqslant 3$, the statement $g \geqslant f + \lfloor \frac{1}{2}n \rfloor$ is easily verified.

Now let $n \geqslant 4$. We claim that G has a minimal cut set L (of lines) such that at least one of the resulting components is even. If G is 3-connected, the existence of such a cut set is trivial: take all lines between an arbitrary subgraph isomorphic to K_2 and the rest of the graph. If $\kappa(G)=2$, let $\{p,q\}$ be a point-cut that is chosen in such a way that one of the components of $G-\{p,q\}$ is as small as possible, and let r be a neighbour of p in that component. Then the set of all lines from $\{p,r\}$ to the rest of G is a cut set with the required property.

Since L is minimal, $G-L$ has exactly two connected components, G_1 and G_2, say. Because of the induction hypothesis we now have, in an obvious notation, $g_1 \geqslant f_1 + \lfloor \frac{1}{2}n_1 \rfloor$ in G_1 and $g_2 \geqslant f_2 + \lfloor \frac{1}{2}n_2 \rfloor$ in G_2. Let g_3 be the number of good lines in L, and f_3 the number of errors. Then $g_3 \geqslant f_3$ or else we change all colors in G. Adding the three inequalities we obtain

$$\sum_1^3 g_i \geqslant \sum_1^3 f_i + \lfloor \tfrac{1}{2}n_1 \rfloor + \lfloor \tfrac{1}{2}n_2 \rfloor.$$

In each optimal coloring of G, we have $g \geqslant \sum_1^3 g_i$ and $f \leqslant \sum_1^3 f_i$. Furthermore, since n_1 or n_2 is even, we have $\lfloor \frac{1}{2}n_1 \rfloor + \lfloor \frac{1}{2}n_2 \rfloor = \lfloor \frac{1}{2}n \rfloor$. Hence

$$g \geqslant f + \lfloor \tfrac{1}{2}n \rfloor.$$

Since $f+g=m$, this is equivalent to $f \leqslant \frac{1}{2}m - \frac{1}{2}\lfloor \frac{1}{2}n \rfloor$, and hence

$$\gamma \leqslant \lfloor \tfrac{1}{2}m - \tfrac{1}{2}\lfloor \tfrac{1}{2}n \rfloor \; \rfloor = \lfloor \tfrac{1}{2}m - \tfrac{1}{4}(n-1) \rfloor. \qquad \square$$

Although the bound of Proposition 5.1 is certainly an improvement over the bound $\lfloor \frac{1}{2}m \rfloor$, it is still not sharp. For example, let n be of the form $2k+1$ with $k \geqslant 1$, and let

$m=\binom{n}{2}-k=2k^2$. All graphs with these parameters are connected. In the following section we will see that Proposition 6.2 can be applied here; the result is $\gamma(G)=k^2-k$. However, the upper bound of Proposition 5.1. is equal to $\lfloor k^2-\frac{k}{2}\rfloor$ in this case.

6. Dense graphs

Proposition 6.1.

$$\gamma(K_n)=\lfloor \tfrac{1}{4}(n-1)^2\rfloor=\binom{\lfloor\frac{1}{2}n\rfloor}{2}+\binom{\lceil\frac{1}{2}n\rceil}{2}.$$

Proof. Suppose k points of K_n are colored white, and $n-k$ are colored red. Then the number of errors is $\binom{k}{2}+\binom{n-k}{2}$. Choose k such that this quantity is minimized. This yields the result. □

Proposition 6.2. *Let $n>1$, $m=\binom{n}{2}-d$ where $0\leqslant d<\frac{1}{2}n$. Then $\gamma(G)=\gamma(K_n)-d$ for each (n,m)-graph G.*

Proof. Let G be an (n,m)-graph with $2d<n$. The complement H of G had d lines, where $d<\frac{1}{2}n$, hence H is not connected. Let H_j $(j=1,\ldots,k)$ be the j-th component of H, with n_j points and m_j lines. Since H_j is connected, we know that $m_j\geqslant n_j-1$. Hence $d=\sum_{j=1}^{k} m_j\geqslant n-k$, hence $k\geqslant n-d\geqslant\lfloor\frac{1}{2}n\rfloor+1$. According to Lemma 6.3 below, there exists a way of grouping the components of H into two classes containing $\lfloor\frac{1}{2}n\rfloor$ and $\lceil\frac{1}{2}n\rceil$ points. This means that there exists a partition of $V(G)$ into two parts with these cardinalities such that all lines that do not belong to G are between points in the same part. Hence the result. □

Lemma 6.3. *If $n_1,n_2,\ldots,n_k$ is a sequence of natural numbers with sum n and $k\geqslant\lfloor\frac{1}{2}n\rfloor+1$, then there exists a subsequence with sum $\lfloor\frac{1}{2}n\rfloor$.*

Remark. The following elegant proof is due to Wetterling [4].

Proof. First let n be even: $n=2h$, so that $k\geqslant h+1$. Now consider a regular $2h$-gon and partition the vertices into k groups of successive points, such that the group sizes are $n_1,n_2,\ldots,n_k$ when traversing the circumference. Now color the first point of each group red, and all other points black. The number of diameters is h, whereas the number of red points is least $h+1$. Hence there is at least one diameter with 2 red end-points, and hence a subsequence with sum h.

Let n be odd: $n=2h+1$, so $k\geqslant h+1$. Consider a regular $(2h+1)$-gon and color the vertices in the same way as above. If there is no subsequence with sum $\lfloor\frac{1}{2}n\rfloor=h$, then each point diametrically opposite a red point has 2 black neighbours. The number of red points is at least $h+1$, so the number of black points is at least $h+2$, but the total number of points is only $2h+1$. Hence there is a subsequence with sum h.

So each cyclic arrangement of the sequence $n_1, \ldots, n_k$ contains a subsequence of successive terms with sum $\lfloor \frac{1}{2}n \rfloor$. □

The bound $d < \frac{1}{2}n$ in Proposition 6.2 cannot be relaxed, as we shall see in Proposition 7.3 below.

7. The exact maximum of $\gamma(G)$

Let $\gamma^*(n,m) = \max\{\gamma(G) \mid G$ is an (n,m)-graph$\}$. Note that γ^* is nondecreasing in n and m. Let

$$\gamma^{**}(m) = \max_n \gamma^*(n,m) = \lim_{n\to\infty} \gamma^*(n,m).$$

The following result is a natural extension of Proposition 5.1.

Proposition 7.1. *Suppose* $\gamma^*(n,m) = \lfloor \frac{1}{2}m - \frac{1}{4}(n-1) \rfloor$. *Then* $\gamma^*(p,m) = \gamma^*(n,m)$ *for all* $p > n$, *and so* $\gamma^{**}(m) = \gamma^*(n,m)$.

Proof. Let G be a (p,m)-graph. If G has at least two nontrivial components G_1 and G_2, we can identify a point of G_1 with a point of G_2, and add an isolated point, so as to form a new (p,m)-graph without changing γ. By repeating this process if necessary, we find a (p,m)-graph H with only one nontrivial component such that $\gamma(H) = \gamma(G)$. Let H have c components. If $p - c \geq n$, then Proposition 5.1 gives

$$\gamma(H) \leq \lfloor \tfrac{1}{2}m - \tfrac{1}{4}(p-c) \rfloor \leq \lfloor \tfrac{1}{2}m - \tfrac{1}{4}(n-1) \rfloor = \gamma^*(n,m),$$

while if $p - c < n$ then we can remove $p - n$ isolated points from H to give an (n,m)-graph with the same value of γ, again showing that $\gamma(H) \leq \gamma^*(n,m)$. In either case, $\gamma(G) = \gamma(H) \leq \gamma^*(n,m)$, so that $\gamma^*(p,m) \leq \gamma^*(n,m)$. But the reverse inequality obviously holds, and so the proposition is proved. □

The converse of Proposition 7.1 is not true: we shall show in Proposition 7.6 that $\gamma^*(n,8) = 2$ for all $n \geq 5$, whereas $\lfloor \frac{1}{2}8 - \frac{1}{4}(5-1) \rfloor = 3$.

However, we do not know any other counterexamples to the converse of Proposition 7.1.

Now let $h(m) = \lceil \frac{1}{2}(1 + \sqrt{8m+1}) \rceil = \min\{n \mid m \leq \binom{n}{2}\}$, the smallest number of points that a graph with m lines can have. If an (n,m)-graph has c components, then evidently there is a graph with m lines and $n - c + 1$ points, and so $n - c \geq h(m) - 1$. Thus the following result follows immediately from Proposition 5.1.

Proposition 7.2. *If G has m lines, then* $\gamma(G) \leq \lfloor \frac{1}{2}m - \frac{1}{4}(h(m)-1) \rfloor$.

Proposition 7.3. *Let* $m=\binom{n}{2}-d$ *where* $0\leqslant d\leqslant n-2$, *so that* $n=h(m)$. *Then*

$$\gamma^*(n,m)=\begin{cases}\lfloor \frac{1}{4}(n-1)^2\rfloor-d & \text{if } d<\frac{1}{2}n,\\ \lfloor \frac{1}{4}(n-2)^2\rfloor & \text{if } d\geqslant\frac{1}{2}n.\end{cases}$$

Proof. The formula for $d<\frac{1}{2}n$ follows immediately from Proposition 6.2. Now let G_i be obtained from K_{n-1} by adding an extra point and i extra lines, where $0\leqslant i\leqslant\lceil\frac{1}{2}(n-1)\rceil$. Evidently $\gamma(G_i)=\gamma(K_{n-1})=\lfloor\frac{1}{4}(n-2)^2\rfloor$. But if $j=\lceil\frac{1}{2}(n-1)\rceil$, then G_j has $m=\binom{n}{2}-d$ lines where $d=\lfloor\frac{1}{2}(n-1)\rfloor<\frac{1}{2}n$, and so $\gamma(G_j)=\gamma^*(n,m)$ by the first part of the proposition. Since γ^* is nondecreasing in m, it follows that $\gamma^*(n,m)=\gamma(G_i)$ for $0\leqslant i\leqslant j-1$, and this is the second part of the proposition. □

Suppose that the value of $\gamma^*(h(m),m)$ given by Proposition 7.3 is equal to the upper bound in Proposition 7.2. Then, by Proposition 7.1, $\gamma^*(n,m)$ is constant (and known) for all $n\geqslant h(m)$. We call these values of m *easy*, and all other values *hard*. It is of some interest to know which values of m are easy.

Proposition 7.4. *m is easy if and only if m is of the form* $\binom{p}{2}$ *or* $\binom{p}{2}\pm 1$ *or* $\binom{2p+1}{2}+2$.

Proof. By writing out. □

Although easy values of m have asymptotic density 0, among 'small' numbers they are quite common. For example, in the set $\{0,1,\ldots,50\}$ there are 29 easy values, the hard ones being

$$8,13,17,18,19,24,25,26,30,\ldots,34,39,\ldots,43,47,\ldots,50.$$

For some of the hard values, a graph with $n=h(m)+1$ points can be given, such that the upper bound $\lfloor\frac{1}{2}m-\frac{1}{4}(n-1)\rfloor$ is attained.

Proposition 7.5. *If* $m=\binom{n-2}{2}+3$ $(n\geqslant 5)$, *then*

$$\gamma^*(n,m)=\lfloor\tfrac{1}{2}m-\tfrac{1}{4}(n-1)\rfloor=\lfloor\tfrac{1}{2}m-\tfrac{1}{4}(h(m)-1)\rfloor.$$

Proof. The second equality would fail if m were odd and n divisible by 4, or if m were even and n of the form $4k+2$; it is easy to check that neither of these situations can arise.

Consider the graph G with $\kappa(G)=1$ and two blocks viz. K_{n-2} and K_3. Then

$$\gamma(G)=\gamma(K_{n-2})+\gamma(K_3)=\lfloor\tfrac{1}{4}(n-3)^2\rfloor+1=\lfloor\tfrac{1}{4}(n^2-6n+13)\rfloor$$

and

$$\lfloor\tfrac{1}{2}m-\tfrac{1}{4}(n-1)\rfloor=\lfloor\tfrac{1}{2}(n^2-6n+13)\rfloor,$$

so the upper bound is attained. □

So if m is of the form $\binom{n-2}{2}+3$, $\gamma^*(n,m)$ is known for all n. In the range $0,\ldots,50$ this deals with the values $13,18,24,31,39,48$. Note that $h(13)=6$, $\gamma^*(6,13)=4$ by Proposition 7.3, and $\gamma^*(n,13)=\gamma^{**}(13)=5$ for all $n\geqslant 7$ by Proposition 7.5. 13 is the smallest value of m such that $\gamma^*(n,m)$ is not constant for all $n\geqslant h(m)$.

γ^* can be determined for certain other values of m and n by ad hoc arguments. The case $m=8$ is an exceptional one as far as we know.

Proposition 7.6. $\gamma^*(n,8)=2$ *for* $n\geqslant h(8)=5$.

Proof. The wheel W_5 on five points has $\gamma(W_5)=2$, and so $\gamma^*(n,8)\geqslant 2$ for all $n\geqslant 5$. For disconnected $(n,8)$-graphs, we have $\gamma^*(n,8)\leqslant\gamma^*(n-1,8)$, so it suffices to prove $\gamma^*(n,8)\leqslant 2$ for connected $(n,8)$-graphs.

If G is a connected $(n,8)$-graph, then $n\in\{5,6,7,8,9\}$. For $n\geqslant 7$, the cycle rank is $\leqslant 2$, hence $\gamma^*\leqslant 2$. For $n\leqslant 6$, it is readily checked, using a table, that again $\gamma^*\leqslant 2$. In fact. $\gamma(G)=2$ for both $(5,8)$-graphs. $\square$

Proposition 7.6 implies that $\gamma^{**}(8)=2$, whereas $\lfloor\frac{1}{2}m-\frac{1}{4}(h(m)-1)\rfloor=3$ when $m=8$. The question of whether $\gamma^{**}(m)=\lfloor\frac{1}{2}m-\frac{1}{4}(h(m)-1)\rfloor$ for all $m\neq 8$ is still open.

Acknowledgment

We thank the referees for their valuable suggestions to improve the presentation.

References

[1] P. Erdős, On some extremal problems in graph theory, Israel J. Math. 3 (1965) 113–116.
[2] M.R. Garey and D.S. Johnson, Computers and Intractability (Freeman, San Francisco, CA. 1979).
[3] F. Harary, Graph Theory (Addison-Wesley, Reading, MA, 1969).
[4] W.W.E. Wetterling, personal communication.