COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **18** (4) (1998) 493–501

HOW TO DECREASE THE DIAMETER OF TRIANGLE-FREE GRAPHS*

PAUL ERDŐS, ANDRÁS GYÁRFÁS and MIKLÓS RUSZINKÓ

Received October 12, 1997

To the memory of Paul Erdős

Assume that G is a triangle-free graph. Let $h_d(G)$ be the minimum number of edges one has to add to G to get a graph of diameter at most d which is still triangle-free. It is shown that $h_2(G)=\Theta(n\log n)$ for connected graphs of order n and of fixed maximum degree. The proof is based on relations of $h_2(G)$ and the clique-cover number of edges of graphs. It is also shown that the maximum value of $h_2(G)$ over (triangle-free) graphs of order n is $\lceil n/2-1\rceil\lfloor n/2-1\rfloor$. The behavior of $h_3(G)$ is different, its maximum value is $n-1$. We could not decide whether $h_4(G)\le(1-\epsilon)n$ for connected (triangle-free) graphs of order n with a positive ϵ.

1. Introduction

The topic of this paper grew from problems studied by the first two authors during the summer of 1995 (special cases mentioned in [5], p.229 - unfortunately with misprints). In vague form, the question is to find an optimal extension of a given triangle-free graph into a *maximal triangle-free (MTF) graph,* i.e. into a graph H which is triangle-free but adding any new edge to H destroys that property. Note that a graph is MTF if and only if it is triangle-free and of diameter at most two. Different interpretations of extensions and optimal extensions lead to various problems. This paper is focused on extensions where new vertices cannot be added.

For a given triangle-free graph G, let $h(G)$ denote the minimum number of edges one has to add to G to get a MTF extension. More generally, $h_d(G)$ is the minimum number of edges one has to add to G to get a graph of diameter at most d which is still triangle-free. In the special case $d=2$ $h_2(G)=h(G)$. In Section 4 the cases $d=3,4,5$ are investigated. The behavior of h_2 (=h) is different, it is rather related to clique coverings of graphs. Therefore – even for very simple graphs, like mK_2 (matching graph) – it seems to be difficult to determine $h(mK_2)$ precisely. Our main result states that $h(G)=\Theta(n\log n)$ (Corollary 2.7) for graphs containing no isolated vertices and having order n and fixed maximum degree d. The proof technique reveals several connections with the theory of clique coverings,

Mathematics Subject Classification (1991): 05C12, 05C35

* Research supported in part by OTKA Grants T 016414 and F 014919.

e.g., we utilize the fact that $cc(\overline{G}) \le c\log n$ for graphs with n vertices and with fixed maximum degree, where $cc(G)$ is the clique cover number, i.e., the minimum number of complete subgraphs of G needed to cover the edges of G, and $\overline{G}$ denotes the complement of G. This follows from a more general result of Alon (Theorem 2.1). Here a different proof is given, which leads to a slightly better constant c in the case of fixed maximum degree (Theorem 2.2). We note that tighter results on $cc(\overline{G})$ are known in the literature for special graphs, like paths, cycles, matchings (see [3], [8]). In fact, for our purposes a weighted version of $cc(G)$ is also needed (see Section 2). Another result (Theorem 3.3) states that for $n \ge n_0$

$$\max_{\substack{G\ is\ \triangle-free \\ |V(G)|=n}} h(G) = \lceil n/2-1\rceil \lfloor n/2-1\rfloor,$$

and Paul Erdős remarked in [5]: "The proof is not quite trivial and we hope that it will be written up with all details soon." It is very sad to do this without being urged by his frequent inquiries: "How is our paper?"

2. Extensions and clique covers

Let $cc(G)$ denote the minimum number of complete subgraphs of G needed to cover the edges of G. This parameter, the *clique-cover number* was introduced in [6] and has been studied extensively – it has many applications. For our purposes an upper bound for the clique-cover number of the complements of graphs with bounded degree is needed. The following theorem of Alon [1] is proved by a probabilistic argument.

Theorem 2.1. (Alon [1]) *Assume that a graph G has n vertices and maximum degree d. Then*

$$cc(\overline{G}) \le \frac{2e^2(d+1)^2}{\log e}\log n.$$

(Here and through the whole paper log stands for base two logarithm.) For fixed maximum degree one can use a different argument, which gives a slightly better constant.

Theorem 2.2. *Assume that G has n vertices and fixed maximum degree d. Then*

$$cc(\overline{G}) \le (2d^2-2d+1)\log n + O(\log\log n).$$

Proof. It is easy to see [7] that the strong chromatic index of G is at most $2d^2-2d+1$, i.e. the edges of G can be colored with at most $2d^2-2d+1$ colors so that each color class is a set of pairwise disjoint edges of G whose union does not have any other edges from G (strongly independent edges). The complement of the graph induced by a color class – by the well-known result of Gregory and Pullman [8] – can be covered by at most $\log n + O(\log\log n)$ cliques. Doing this for each color class we use at most $(2d^2-2d+1)\log n + O(\log\log n)$ cliques of $\overline{G}$. The proof will be finished by covering the remaining edges of $\overline{G}$ with a constant (depending only on d) number of cliques.

Let $I(x)$ denote the set of colors of edges incident to the vertex x. Notice that an edge (x,y) of $\overline{G}$ is uncovered if and only if $I(x)$ and $I(y)$ are disjoint sets. Using the notation $[n]$ for the set $\{1,2,\dots,n\}$, for $J \subseteq [2d^2-2d+1]$ let A_J denote the set of vertices x of G with $I(x)=J$. Notice, that for disjoint subsets J and K of $[2d^2-2d+1]$ the pairs $x \in A_J$ and $y \in A_K$ are all in $\overline{G}$. Since the maximum degree of G is d, the vertices A_J and A_K can be partitioned into at most $d+1$ independent sets. This gives at most $(d+1)^2$ complete subgraphs of $\overline{G}$ to cover all edges (x,y) with $x \in A_J$, $y \in A_K$. (In fact, as Kézdy [10] observed, $(d+1)^2$ can be replaced by 4.) Since the number of pairs $J, L \subseteq [2d^2-2d+1]$ depends only on d, the proof is finished. ∎

Theorem 2.2 will be used to get an upper bound on $h(G)$ for graphs with bounded maximum degree.

Theorem 2.3. *If G is a graph of order n and of fixed maximum degree d then $h(G) \le c(d) n \log n$.*

Proof. By Theorem 2.2, we can choose a clique-cover $\mathcal{C}$ of $\overline{G}$ using $m \le c(d)\log n$ cliques $S_1,\dots,S_m$, i.e. using m independent sets of G. These independent sets clearly cover at least $n-d-1$ vertices of G so one of them has at least $(n-d-1)/m$ vertices. Thus for n sufficiently large we may assume that G has an independent set S such that $|S|=m$. Let T denote the set of vertices of G at distance at most two from S. Set $R=V(G)\setminus T$. The subgraphs induced by the sets $R\cap S_i$, $i \in [m]$, form a clique-cover $\mathcal{D}$ on $\overline{G}[R]$ (although some of them can be empty). Associate with each $D \in \mathcal{D}$ a (separate) point in S. Since $|S| \ge |\mathcal{D}|$, this can be certainly done.

The first step is to extend G to G^* by adding the edges $x, f(x)$ for each $x \in D \in \mathcal{D}$. Observe that there are no triangles in G^*. Moreover, pairs of vertices of G^* such that both vertices are in R are at distance at most two in G^*. This follows from the fact that once two vertices are not connected, they are contained in some $R\cap S_i$. Therefore, both of them will be connected with the same point of S associated to $R\cap S_i$. Clearly, at most mn edges are added at this step to G.

The second step is a greedy extension of G^* . Apply repeatedly the following one-edge extension for vertices of T: if there exists $x \in T$ such that some vertex y of the current extension is at distance at least three from x then add the edge (x,y); stop otherwise. It is easy to see that after the second step we have a triangle-free graph of diameter two (MTF) and the second step adds at most $n(m+dm+d^2m)$ new edges. ∎

Next we show that Theorem 2.3 gives the right order of magnitude for $h(G)$ for graphs of bounded degree, provided that G has at least ϵn edges. We need a variant of the clique cover number, introduced by Katona (see [9]). Assume that $A_1, A_2, \ldots, A_t$ are the vertex sets of cliques in a clique-covering $\mathcal{C}$ of a graph G. Define the *size* of $\mathcal{C}$ as $\sum_{i=1}^{t} |A_i|$ and then let $cc^*(G)$ be the minimum size taken over all possible clique-coverings of G. The following lemma is a direct consequence of a result of T.G. Tarján ([12]), see also in the survey of Tuza ([14]).

Lemma 2.4. $cc^*(K_{2m} - mK_2) \geq m \log m$

The following lemma shows the connection between the functions h and cc^* ($e(G)$ denotes the number of edges in G).

Lemma 2.5. *For every triangle-free graph G, $h(G) \geq (cc^*(\overline{G}) - 2e(G))/4$.*

Proof. Assume that G can be extended to a MTF graph G^* by adding k edges to $E(G)$. Let G^+ be the graph with the added k edges and with vertex set $V(G)$. The sets $\Gamma_{G^*}(x)$ – where $\Gamma_{G^*}(x)$ is the set of neighboring vertices of x in G^* – and the edges of G^+ give a clique cover on $\overline{G}$. Therefore,

$$cc^*(\overline{G}) \leq \sum_{x \in V} |\Gamma_{G^*}(x)| + 2k = 2e(G) + 2e(G^+) + 2k = 2e(G) + 4k$$

which gives the required inequality. ∎

Theorem 2.6. *Assume that a graph G on n vertices has at least ϵn edges and maximum degree d. Then $h(G) \geq c(\epsilon, d) n \log n$.*

Proof. As it is mentioned earlier, the strong chromatic index of G is less than $2d^2$, thus there are at least $m = \lfloor \epsilon n / 2d^2 \rfloor$ strongly independent edges in G. Let H be the subgraph of G induced by the vertices of these m edges. Then, using the two lemmas above

$$h(G) \geq \frac{cc^*(\overline{G}) - 2e(G)}{4} \geq \frac{cc^*(\overline{H}) - 2e(G)}{4} \geq \frac{m \log m}{4} - \frac{e(G)}{2} \geq \frac{m \log m}{4} - \frac{dn}{4},$$

which gives the desired result. ∎

Combining Theorems 2.3 and 2.6 we get the following.

Corollary 2.7. *Assume G is a triangle-free graph without isolated vertices of order n and maximum degree d. Then*

$$c_1 n \log n \leq h(G) \leq c_2 n \log n,$$

where c_1, c_2 depend only on d.

In the special case when $d = 1$, our proofs give the following estimates for mK_2.

Corollary 2.8. $m \log m / 4 - m/2 \leq h(mK_2) \leq 8m \log m$.

3. The maximum of $h(G)$ and a Turán-type problem

It is easy to guess that the maximum of $h(G)$ over triangle-free graphs of order n is realized by the double star T_n with evenly distributed leaves, i.e. by the graph consisting of an edge and $\lceil n/2-1\rceil$, $\lfloor n/2-1\rfloor$ vertices of degree one adjacent to each of the two end-vertices of the edge, respectively. In this case no edges can be added within the chromatic classes of T_n thus the (only) optimal extension is to extend T_n to a complete bipartite graph. However, we could not find a direct proof of Theorem 3.3 which avoids results and proof techniques used for Turán type problems restricted to non-bipartite graphs. Thus, for our purposes the following theorem of Erdős [4] can be used.

Theorem 3.1. (Erdős [4]) *A non-bipartite triangle-free graph on n vertices contains at most*

$$\left\lceil \frac{n-5}{2} \right\rceil \left\lfloor \frac{n-5}{2} \right\rfloor + 2(n-5) + 5$$

edges. The maximum is attained by subdividing an edge of the evenly distributed complete bipartite graph on $n-1$ vertices with a single vertex.

From Theorem 3.1 the following corollary – which will be also used for our purposes – immediately follows. Observe that one can get the asymptotic version of this corollary from (more general) results of Simonovits [11]. (C_5 stands for the cycle of length five.)

Corollary 3.2. *A triangle-free graph on n vertices having two vertex disjoint C_5-s contains at most $\lceil n/2-5\rceil\lfloor n/2-5\rfloor+4(n-10)+20$ edges. The maximum is attained by a construction similar to the one in Theorem 3.1.*

Now we are ready to determine the maximum value of $h(G)$ taken over all triangle-free graphs on n vertices.

Theorem 3.3. *Every triangle-free graph on $n \geq n_0$ vertices can be extended into MTF by adding at most $\lceil n/2-1\rceil\lfloor n/2-1\rfloor$ edges to it. Moreover, this bound is tight, i.e.*

$$\max_{\substack{G \text{ is } \triangle-free \\ |V(G)|=n}} h(G) = \lceil n/2-1\rceil\lfloor n/2-1\rfloor,$$

for n sufficiently large.

Proof. As it is already mentioned, $h(T_n) = \lceil n/2-1\rceil\lfloor n/2-1\rfloor$, which gives the desired lower bound for the above maximum. To avoid the use of integer parts we shall prove the upper bound for $n=2k$ noting that the same proof works for graphs having odd number of vertices. The idea of the proof is to apply Turán's theorem and its variations to show that – after adding a few edges – any MTF extension will give the required number of edges.

(α) First assume that G contains at least $2k-1$ edges. By the well-known theorem of Turán [13] a triangle-free graph on $2k$ vertices may contain at most k^2 edges, thus any MTF extension of G adds at most $k^2-(2k-1)=(k-1)^2$ edges to G.
(β) If $k+3\le e(G)\le 2k-2$, then G has at least two components and one of them, say, C has at least two edges.

If C has precisely $2k-1$ vertices then C is a tree. If C is a star then by adding one edge one can extend G into MTF. If C is not a star then let x_1,x_2,x_3,x_4 be the vertices of a path in C. Adding the edges (y,x_1), (y,x_4) with $y\in V(G)\setminus V(C)$ we get a C_5 without creating a triangle.

If $|V(C)|\le 2k-2$ then let x_1,x_2,x_3 be a path in C. Select an edge $(x,y)\in E(G)$ such that $x,y\in V(G)\setminus V(C)$ if there exist one. Otherwise take two isolated vertices x and y in $V(G)\setminus V(C)$. In both cases by adding at most three edges to G we get the C_5 x_1,x_2,x_3,x,y without creating a triangle. Continuing with any MTF extension, by Theorem 3.1 we add at most $(k-2)(k-3)+2(2k-5)+5+3-(k+3)=(k-1)^2$ edges.
(γ) If $14\le e(G)\le k+2$, G has at least $k-2$ components. Assume $k\ge 12$ and take ten vertices from different components. Make two vertex disjoint C_5-s on those ten vertices by adding ten appropriate edges to G. Clearly, no triangle will occur. Then continue with any MTF extension. By Corollary 3.2 at most $(k-5)^2+4(2k-10)+20+10-14=(k-1)^2$ edges are added.
(δ) Finally, if $e(G)\le 13$, the vertex set $V(E)$ spanned by the edges is of size at most 26. First join all vertices of $V(G)\setminus V(E)$ to some vertex of $V(E)$. Then adding edges between vertices of $V(E)$, the subgraph spanned by $V(E)$ is completed to MTF. Finally, add greedily all possible edges between $V(G)\setminus V(E)$ and $V(E)$ without making a triangle. If both x and y are in $V(E)$ then their distance is at most two, because $G^*[V(E)]$ (G^* is the graph with added edges) is MTF. If x and y are in $V(G)\setminus V(E)$ their distance is two either, since they are adjacent to the same vertex from $V(E)$ (first step). If $x\in V(E)$ and $y\in V(G)\setminus V(E)$ are not adjacent then – by maximality of the third extension – some neighbor $z\in V(E)$ of x is adjacent to y, otherwise we would have added the edge (x,y) in the third step. Clearly, the extension is triangle-free, and at most $2k+13^2+2k\cdot 26=54k+13^2<(k-1)^2$ edges are added, provided that k is sufficiently large. ∎

$\overline{K_6}$ shows that in Theorem 3.3 n_0 is at least 7, however, we could not determine the smallest n_0 for which this statement holds.

4. Related problems

It seems interesting to look at $h(G)$ in terms of the order and the maximum degree of G. The proof of Theorem 2.3 shows that $h(G)\le 2mn+dmn+d^2mn$, where $m=cc(\overline{G})$ and d is the maximum degree of G. By the result of Alon (Theorem 2.1) $m\le cd^2\log n$ from which $h(G)\le cd^4\log n$. Therefore, $h(G)=o(n^2)$ is true for graphs with maximum degree $o(n^{1/4}/\log n)$. On the other hand, the (bipartite) incidence graph of a finite plane is an example of a graph of order n and of maximum degree at most $c\sqrt{n}$ for which $h(G) \geq c_1 n^2$.

Problem 4.1. Is it true that $h(G) = o(n^2)$ for graphs of order n and of maximum degree $o(\sqrt{n})$?

Since an MTF graph is of diameter 2, one can generalize the problem above by defining $h_d(G)$ as the minimum number of edges needed to extend the triangle-free graph G into a triangle-free graph of diameter at most d.

Theorem 4.2. *$h_3(G) \leq n-1$ for every triangle-free G with n vertices.*

Proof. Let p_1 be an arbitrary vertex in G and let A_1 be the set of vertices in G which are at distance at least three from p_1. Select S_1 as a maximal independent set in $G[A_1]$. Add to G all edges of the form (p_1, s) $(s \in S_1)$, and denote this extension of G by G_1. Set $T_1 = \{p_1\} \cup S_1$ and $B_1 = V(G) \setminus T_1$. In general, if T_i is defined for $1 \leq i \leq k$ select p_{k+1} from

$$B_k = V(G) \setminus \left(\cup_{j=1}^{k} T_j \right)$$

(provided that B_k is nonempty) and define A_{k+1} as the set of vertices in B_k which are at distance at least three from p_{k+1} in the graph G_k. Then G_{k+1} is the graph obtained from G_k by adding all edges (p_{k+1}, s) where s runs through a maximal independent set $S_{k+1} \subseteq A_{k+1}$. Finally, set $T_{k+1} = \{p_{k+1}\} \cup S_{k+1}$. It is easy to see that the above procedure leads to the desired extension of G. ∎

Theorem 4.2 is trivially tight if G has no edges. Restricting G to connected graphs, significant improvement is not possible, in fact for the path of length n, $h_3(P_n) \geq n-c$ with a small constant c. This follows from a stronger result [2]: for any extension of P_n to a (not necessarily triangle-free) graph of diameter three one has to add at least $n-c$ edges.

The situation for $h_4(G)$ might be different. While every connected graph can be extended to a graph of diameter at most four by adding no more than $n/2$ edges and this is about best possible [2], it is not clear whether the bound $n-1$ can be improved for the triangle-free case.

Problem 4.3. Is it true that for every connected triangle-free G, $h_4(G) \leq (1-\epsilon)n$ for some positive constant ϵ?

We finish the paper with a result for $h_5(G)$.

Theorem 4.4. *If G is a triangle-free graph of order n without isolated vertices, then $h_5(G) \leq (n-1)/2$.*

Proof. Assume that $\nu(G) = t$, where ν denotes the maximum number of independent edges of G (matching number). Clearly, this implies that G has at most $n-t$

independent vertices. Therefore, if we extend G to G^* by adding all edges from a given vertex x to a maximal independent subset of the set at distance at least three from x, no more than $n-t$ edges are added, moreover G^* is of diameter at most four and contains no triangles.

On the other hand, we shall define G^{**}, a triangle-free extension of G by adding no more than $t-1$ edges as follows. Let $M=\{e_1,e_2,\dots,e_t\}$ be a maximum matching of G. Using that M is maximal, G is triangle-free without isolated vertices, it follows easily that $V(G)$ can be covered by the vertices of t stars of G. Let C denote the set of centers of the covering stars. (In fact, C can be defined by selecting one of the vertices of e_i for each i, $1\le i\le t$.) By a slight modification of the proof of Theorem 4.2, one can add at most $t-1$ edges to $G[C]$ to get a triangle-free graph G^{**} with the property: any pair of vertices in C are at distance at most three in G^{**}. This implies that any pair of vertices of G^{**} are at distance at most five in G^{**} (using that C is the set of centers of covering stars).

The proof is finished by taking the better extension from the pair of graphs G^*, G^{**}, it extends G by at most $min\{n-t,t-1\}\le(n-1)/2$ edges. ∎

References

[1] N. Alon: Covering graphs by the minimum number of equivalence relations, *Combinatorica* **6**(3) (1986), 201–206.

[2] N. Alon, A. Gyárfás, M. Ruszinkó: Decreasing the diameter of bounded degree graphs, *submitted.*

[3] D. de Caen, D. A. Gregory, N. J. Pullman: Clique coverings of complements of paths and cycles, *Annals of Discrete Mathematics* **27** (1985), 257–268.

[4] P. Erdős: On a theorem of Rademacher–Turán, *Illinos J. Math.,* **6** (1962), 122–127.

[5] P. Erdős: Some old and new problems in various branches of combinatorics, *Discrete Mathematics,* **165/166** (1997), 227–231.

[6] P. Erdős, A. W. Goodman, L. Pósa: The representation of a graph by set intersections, *Canad. J. Math.,* **18** (1966), 106–112.

[7] R. J. Faudree, A. Gyárfás, R. H. Schelp, Zs. Tuza: The strong chromatic index of graphs, *Ars Combinatoria* **29B** (1990), 205–211.

[8] D. A. Gregory, N. J. Pullman: On a Clique Covering Problem of Orlin, *Discrete Mathematics,* **41** (1982), 97–99.

[9] E. Győri, A. V. Kostochka: On a Problem of G. O. H. Katona and T. Tarján, *Acta Math. Acad. Sci. Hungar.* **34**(3–4) (1979), 321–327.

[10] A. Kézdy: *Personal communication.*

[11] M. Simonovits: Extremal Graph Problems with Symmetrical Extremal Graphs, Additional Chromatic Conditions, *Discrete Mathematics,* **7** (1974), 349–376.

[12] T. G. Tarján: Complexity of lattice-configurations, *Studia Sci. Math. Hungar.,* **10** (1975), 203–211.

[13] P. Turán: On an extremal problem in graph theory (in Hungarian), *Mat. Fiz. Lapok,* **48** (1941), 436–452.

[14] Zs. Tuza: Applications of the Set-pair Method in Extremal Problems, II. In: *Combinatorics, Paul Erdős is Eighty, Volume 2, Bolyai Society Math. Studies 2.* 459–490.

András Gyárfás

Computer and Automation Research Institute of the Hungarian Academy of Sciences, Budapest P.O.Box 63, Hungary-1518.
gyarfas@luna.aszi.sztaki.hu

Miklós Ruszinkó

Computer and Automation Research Institute of the Hungarian Academy of Sciences, Budapest P.O.Box 63, Hungary-1518.
ruszinko@lutra.sztaki.hu

recent address

Dept. of Mathematical Sciences, Carnegie Mellon University Pittsburgh, Pennsylvania 15232 U.S.A.