THE MAXIMUM NUMBER OF EDGES IN A MINIMAL GRAPH OF DIAMETER 2

By

Zoltán Füredi

IMA Preprint Series # 408

March 1988

THE MAXIMUM NUMBER OF EDGES IN A MINIMAL GRAPH OF DIAMETER 2

ZOLTÁN FÜREDI[1]

Abstract. A graph $\mathcal{G}$ is a minimal graph of diameter 2 if it has diameter 2 and the deletion of any edge increases its diameter. Here the following conjecture of Murty and Simon is proved for $n > n_0$. If $\mathcal{G}$ has n vertices then it has at most $\lfloor n^2/4 \rfloor$ edges. The only extremum is the complete bipartite graph.

1. PRELIMINARIES

A graph $\mathcal{G}$ is a pair $(V(\mathcal{G}), E(\mathcal{G}))$ (or shortly (V, E)) where E (the *edge-set*) is a set of pairs of V. (V is called *vertex-set.*) Let S be a subset of vertices. Then $\mathcal{G}(S)$ denotes the subgraph induced by S, and $\mathcal{G}(A, B)$ stands for the induced bipartite subgraph (for $A \cap B = \emptyset$). The *neighborhood* of a vertex v is denoted by $N_{\mathcal{G}}(v)$ (or sometimes briefly by $N(v)$), i.e., $N(v) = \{u \in V : \{u, v\} \in E\}$. Note that $v \notin N(v)$. The size of $N(v)$ is called the *degree* of v, $\deg_{\mathcal{G}}(v)$. The graph $\mathcal{G}$ has *diameter* 2 if it is not the complete graph and for each two vertices $u, v \in V$ either $\{u, v\}$ is an edge of $\mathcal{G}$, or $N(u) \cap N(v) \neq \emptyset$ (or both). $\mathcal{G}$ is called a *minimal* graph of diameter 2 if its diameter is 2, and the deletion of any of its edges spoils this property. Plesník [**P**] observed that all known minimal graphs of diameter 2 on n vertices have no more than $n^2/4$ edges, and that the complete bipartite graphs are minimal graphs of diameter 2. Independently, Simon and Murty (see in [**CH**]) stated these as the following conjecture:

CONJECTURE 1.1. *If $\mathcal{G}$ is a minimal graph of diameter 2 on n vertices, then $|E(\mathcal{G})| \leq \lfloor n^2/4 \rfloor$, with equality holding if and only if $\mathcal{G}$ is the complete bipartite graph $\mathcal{K}(\lfloor n/2 \rfloor, \lceil n/2 \rceil)$.*

Let $\mathcal{G}$ be a minimal graph of diameter 2 with n vertices. Plesník [**P**] proved that $|E(\mathcal{G})| < 3n(n-1)/8$. Caccetta and Häggkvist [**CH**] obtained $|E(\mathcal{G})| < 0.27n^2$. Fan [**F**] proved affirmatively the first part of the Conjecture 1.1 for $n \leq 24$ and for $n = 26$. For $n \geq 25$ he obtained

$$|E(\mathcal{G})| < \frac{1}{4}n^2 + \frac{n^2 - 16.2n + 56}{320} < 0.2532n^2.$$

An incorrect proof was published [**X**] in 1984.

[1]Math. Inst. Hungarian Acad. Sci. 1364 Budapest, P.O.B. 127.
This research was done while the author visited the Institute of Applied Mathematics at the University of Minnesota, Minneapolis, MN 55455

THEOREM 1.2. *Conjecture 1.1 is true for $n > n_0$.*

The value of n_0 is explicitly computable, but the proof given here yields a vastly huge number (a tower of 2's of height about 1000).

This paper is organized as follows. In Section 2 a lemma is proved about the number of disjoint neighborhoods in an arbitrary graph. In Section 3 we prove that $|E(\mathcal{G})| < (1+o(1))n^2/4$ holds for all n. The main idea of the proof is that we delete some $o(n^2)$ edges of $\mathcal{G}$ such that the remaining graph, $\mathcal{G}_0$, has only at most $n^2/4$ edges. In this step we utilize a result of Ruzsa and Szemerédi [**RSz**] about triangle-free, 3-uniform hypergraphs. In Section 4 we put back the deleted edges. Then after a lengthy argument, where we repeatedly use the structure of $\mathcal{G}_0$, we conclude that the Conjecture is true for sufficiently large n. In Section 5 we have some closing remarks on further open problems.

2. THE NUMBER OF DISJOINT NEIGHBORHOODS IN A GRAPH

Let $\mathcal{F}$ be an arbitrary graph on n vertices. Define the set of pairs with disjoint neighborhoods as follows:

$$\operatorname{disj}\mathcal{F} = \{\{u,v\} : N_{\mathcal{F}}(u) \cap N_{\mathcal{F}}(v) = \emptyset\}.$$

LEMMA 2.1. $|E(\mathcal{F})| + |\operatorname{disj}\mathcal{F}| \le \lfloor n^2/2 \rfloor$.

For the complete bipartite graph $\mathcal{K}(\lfloor n/2 \rfloor, \lceil n/2 \rceil)$ equality holds in Lemma 2.1.

Proof. We use induction on n. The cases $n = 1, 2$ are trivial. Suppose that the vertex x has maximum degree, i.e., $|N_{\mathcal{F}}(x)| = \max \deg_{\mathcal{F}}(v)$. If $N_{\mathcal{F}}(x) = \emptyset$, then the left hand side in Lemma 2.1 is $\binom{n}{2} < \lfloor n^2/2 \rfloor$. So we may suppose that there exists a $y \in N(x)$. For every $z \in N(x) \setminus \{y\}$ we have $x \in N(y) \cap N(z) \neq \emptyset$, hence

$$\deg_{\operatorname{disj}\mathcal{F}}(y) \le n - \deg_{\mathcal{F}}(x), \tag{2.1}$$

and by definition

$$\deg_{\mathcal{F}}(y) \le \deg_{\mathcal{F}}(x). \tag{2.2}$$

Summing up (2.1) and (2.2) we have

$$\deg_{\operatorname{disj}\mathcal{F}}(y) + \deg_{\mathcal{F}}(y) \le n. \tag{2.3}$$

We distiguish between two subcases.

1) Suppose first that there exists a $y_0 \in N(x)$ such that the left hand side of (2.3) *is* only at most $n-1$. Let $\mathcal{F}'$ be the graph obtained from $\mathcal{F}$ by deleting the vertex y_0 and the edges through y_0. Obviously

$$|E(\mathcal{F})| = |E(\mathcal{F}')| + \deg_{\mathcal{F}}(y_0), \tag{2.4}$$

and it is easy to see that

$$|\operatorname{disj}\mathcal{F}| \le |\operatorname{disj}\mathcal{F}'| + \deg_{\operatorname{disj}\mathcal{F}}(y_0). \tag{2.5}$$

Summing up (2.4), (2.5), then using the induction hypothesis for $\mathcal{F}'$ and the assumption for y_0, we obtain that

$$|E(\mathcal{F})| + |\operatorname{disj}\mathcal{F}| \le \lfloor (n-1)^2/2 \rfloor + (n-1) \le \lfloor n^2/2 \rfloor.$$

2) Suppose now that equality holds in (2.3) for every $y \in N(x)$. Then equality holds in (2.1) for all $y \in N(x)$, which implies that $\mathcal{K}(N(x), V(\mathcal{F}) \setminus N(x))$ is a subgraph of $\operatorname{disj}\mathcal{F}$. Consequently, there is no edge of $\mathcal{F}$ in $N(x)$. Equality holds in (2.2) as well, so $\mathcal{K}(N(x), V(\mathcal{F}) \setminus N(x))$ is a subgraph of $\mathcal{F}$, too. Hence $\mathcal{F} = \mathcal{K}(N(x), V(\mathcal{F}) \setminus N(x))$. Finally, for this graph the left hand side in Lemma 2.1 is at most $2\lfloor n^2/4 \rfloor$. □

3. The proof of $\max|E(\mathcal{G})| = \frac{1}{4}(1+o(1))n^2$

Let $\mathcal{G}$ denote a minimal graph of diameter 2 with n vertices. Define the set of *critical pairs* as follows. $\{u, v\} \in \operatorname{crit}\mathcal{G}$ if there is a *unique* path of length at most 2 with endpoints u and v. Call this unique path *critical* path and denote it by $P(u, v)$. There are two cases.

I) If $P(u, v)$ consists of only a single edge then we call it *type I.* II) If $P(u, v)$ consists of two edges then we call them *type II.* It is possible that an edge of $\mathcal{G}$ has both types. But the minimality of $\mathcal{G}$ ensures that every edge has at least one type, i.e., every edge belongs to a critical path. For an edge $E \in E(\mathcal{G})$, denote $m(E)$ the *multiplicity* of E, i.e. the number of critical paths in which the edge E appears.

LEMMA 3.1. *For any $m > 0$ the number of edges of $\mathcal{G}$ with multiplicity at least m is at most $n(n-1)/m$.*

Proof. The total sum of multiplicities is at most twice the number of critical pairs, i.e. it is at most $2\binom{n}{2}$. □

An upper bound on the number of light paths. Let m be an arbitrary positive number. A critical path is called *light* if it has two edges, and both have multiplicity less than m. We are going to give an upper bound (depending on n and m) for the number of light paths. To do this we recall some definitions and results from the extremal hypergraph theory.

A *3-graph* (or 3-uniform hypergraph) $\mathcal{H}$ is a pair $(V(\mathcal{H}), E(\mathcal{H}))$, where $V(\mathcal{H})$ is a finite set (the set of *vertices*), and $E(\mathcal{H})$ is a set of 3-element subsets of $V(\mathcal{H})$ (the set of *edges*). $\mathcal{H}$ is called *linear* if every two distinct edges intersect in at most 1 element. Three edges of a hypergraph form a *triangle* if they pairwise intersect, but no vertex is contained in all the three of them. For example, a triangle in a linear 3-graph is isomorphic to $\{\{1,2,3\},\{3,4,5\},\{5,6,1\}\}$. Denote by $\operatorname{RSz}(n)$ the largest number of edges in a triangle-free, linear 3-graph over n vertices. Ruzsa and Szemerédi proved the following theorem.

THEOREM 3.2 [**RSz**]. $\mathrm{RSz}(n) = o(n^2)$.

(Actually, they also proved that $\mathrm{RSz}(n)$ is larger than n^{2-c} for all positive c, but we need the upper bound only.)

LEMMA 3.3. *The number of light paths is less than* $27m\,\mathrm{RSz}(n)$.

Proof. Define the 3-graph $\mathcal{H}^1$ with vertex-set $V(\mathcal{G})$ as the set of 3-element sets determined by the light critical paths of $\mathcal{G}$. Consider a light critical path $P(u,v) = \{\{u,c\},\{c,v\}\}$. The critical pair $\{u,v\}$ does not appear in any other triples from $\mathcal{H}^1$, so there are at most $2(m-1)$ further triples intersecting $\{u,c,v\}$ in 2 elements. Keeping the triple $\{u,c,v\}$ and deleting those from $\mathcal{H}^1$ which intersect it in 2 elements, then continuing this process until no two triples left with intersection size 2 one obtain a linear hypergraph $\mathcal{H}^2$ such that

$$E(\mathcal{H}^2) \subset E(\mathcal{H}^1) \text{ and } |E(\mathcal{H}^2)| \geq |E(\mathcal{H}^1)|/(2m-1). \tag{3.1}$$

A 3-graph $\mathcal{H}$ is called *3-partite*, if one can partition its vertex-set $V(\mathcal{H}) = V_1 \cup V_2 \cup V_3$ such that for all edge $E \in E(\mathcal{H})$ and for all i $(1 \leq i \leq 3)$ one has $|E \cap V_i| = 1$. Erdős and Kleitman proved the following simple but important fact.

FACT 3.4 [**EK**]. *Let $\mathcal{H}$ be an arbitrary r-graph. Then one can find an r-partite subhypergraph $\mathcal{H}'$ of it such that*

$$|E(\mathcal{H}')| \geq \frac{r!}{r^r}|E(\mathcal{H})|.$$

Applying Fact 3.4 to $\mathcal{H}^2$, one obtain a 3-partite, linear hipergraph $\mathcal{H}^3$ with parts V_1, V_2, V_3, such that

$$|E(\mathcal{H}^3)| \geq \frac{2}{9}|E(\mathcal{H}^2)|. \tag{3.2}$$

Let $P(u,v)$ be a critical path with edges $\{u,c\}$ and $\{c,v\}$. The vertex c is called the *center* of the triple $\{u,c,v\}$. Without loss of generality we may suppose that at least 1/3 of the triples of $\mathcal{H}^3$ have its center in V_2. This means, that there is a subhypergraph $\mathcal{H}^4$ of $\mathcal{H}^3$ such that

$$|E(\mathcal{H}^4)| \geq \frac{1}{3}|E(\mathcal{H}^3)|, \tag{3.3}$$

and with the additional property that if $\{v_1, v_2, v_3\}$ is a triple of $\mathcal{H}^4$ with $v_i \in V_i$ then $\{v_1, v_3\}$ is its critical pair.

PROPOSITION 3.5. *$\mathcal{H}^4$ is triangle-free.*

Proof. Suppose on the contrary that three triples P_1, P_2, P_3 of $\mathcal{H}^4$ form a triangle. Then $P_1 \cup P_2 \cup P_3$ intersects V_i $(1 \leq i \leq 3)$ in at least 2 elements. As $|P_1 \cup P_2 \cup P_3| = 6$, we obtain that each V_i contains exactly two vertices from the triangle. Let $(P_1 \cup P_2 \cup P_3) \cap V_i = \{a_i, b_i\}$ and $P_1 = \{a_1, a_2, a_3\}$. Without loss of generality we may suppose that P_i intersects P_1 in a_i, i.e. $P_2 = \{b_1, a_2, b_3\}$ and $P_3 = \{b_1, b_2, a_3\}$. Then (b_1, b_2, a_3) and (b_1, a_2, a_3) are two disjoint paths from b_1 to a_3, wich contradicts the earlier constraint that $\{b_1, a_3\}$ is a critical pair. $\square$

The end of the proof of Lemma 3.3. Proposition 3.5 and Theorem 3.2 imply that $|E(\mathcal{H}^4)| \leq \mathrm{RSz}(n)$, and (3.1)–(3.3) imply that $|E(\mathcal{H}^1)| \leq 27m|E(\mathcal{H}^4)|$. $\square$

The asymptotic upper bound on $|E(\mathcal{G})|$. Let $m = \frac{1}{8}\sqrt{n^2/\mathrm{RSz}(n)}$. Note that $m = m(n)$ tends to infinity according to Theorem 3.2. Delete all edges of $\mathcal{G}$ whose multiplicity is at least m, and those edges which appear in a light critical path. Denote the obtained graph by $\mathcal{G}_0$. Lemma 3.1 and 3.3 imply the following upper bound on the number of deleted edges.

$$|E(\mathcal{G})| \leq |E(\mathcal{G}_0)| + \frac{n(n-1)}{m} + 54m\,\mathrm{RSz}(n) \leq |E(\mathcal{G}_0)| + \frac{2n^2}{m}. \tag{3.4}$$

Deleting these edges from $\mathcal{G}$ we have distroyed all critical paths of lengh 2. In other words, if (u, c, v) is a critical path in $\mathcal{G}$ (with critical pair $\{u, v\}$), then the neighborhoods of u and v in $\mathcal{G}_0$ are disjoint. This implies that $\mathrm{crit}\,\mathcal{G} \subset \mathrm{disj}\,\mathcal{G}_0$, i.e.

$$|\,\mathrm{crit}\,\mathcal{G}| \leq |\,\mathrm{disj}\,\mathcal{G}_0\,|. \tag{3.5}$$

As the edge-set of $\mathcal{G}$ is the union of critical paths, and after the deletion every $V(P(u,v))$ contains at most one edge of $\mathcal{G}_0$, we conclude that the number of edges in $\mathcal{G}_0$ is not more than the number of critical pairs in $\mathcal{G}$, i.e.

$$|E(\mathcal{G}_0)| \leq |\,\mathrm{crit}\,\mathcal{G}|. \tag{3.6}$$

The inequalities (3.5) and (3.6) imply together with Lemma 2.1 that

$$|E(\mathcal{G}_0)| \leq \frac{1}{2}(|E(\mathcal{G}_0)| + |\,\mathrm{disj}\,\mathcal{G}_0|) \leq n^2/4. \tag{3.7}$$

Finally, (3.7) and (3.4) give

COROLLARY 3.6. $|E(\mathcal{G})| \leq \dfrac{n^2}{4} + \dfrac{2n^2}{m} = (1 + o(1))\dfrac{n^2}{4}$. $\square$

4. The proof of the main theorem

We continue the proof started in the previous section. We are going to define $\varepsilon_1, \dots, \varepsilon_7$ which are all funcions of n. The notation $\varepsilon_{i+1} = \varepsilon(\varepsilon_i)$ means that $\varepsilon_{i+1} \geq \varepsilon_i$ and for every positive ε there exists a δ such that $\varepsilon_{i+1} < \varepsilon$ if $\varepsilon_i < \delta$. In other words, for every ε one can find a $n(\varepsilon)$ such that $\varepsilon_1, \dots, \varepsilon_{i+1}$ are all less than ε whenever $n > n(\varepsilon)$. We will end up with the constraint $\varepsilon_7 < 1/500$, from which all the ε's and n_0 can be explicitely calculated.

$\mathcal{G}$ and $\mathcal{G}_0$ are close to each other. We have four graph, $\mathcal{G}_0 \subset \mathcal{G}$ and $\operatorname{crit} \mathcal{G} \subset \operatorname{disj} \mathcal{G}_0$. Suppose that $|E(\mathcal{G})| > (\frac{1}{4} - \varepsilon_1)n^2$. First we formulate the fact that $\mathcal{G}_0$ and $\mathcal{G}$ are *close* to each other. Choose $\varepsilon_2 = 2\varepsilon_1$ and $m \geq 2/\varepsilon_1$, then (3.4) imply that

$$|E(\mathcal{G}_0)| > (\frac{1}{4} - \varepsilon_2)n^2. \tag{4.1}$$

Moreover $|E(\mathcal{G}) \setminus E(\mathcal{G}_0)| < \varepsilon_1 n^2$. This implies that there exists a $\varepsilon_3 = \varepsilon(\varepsilon_1)$, such that for almost all vertices x (i.e. with at most $\varepsilon_3 n$ exceptions) the folowing is true

$$\deg_{\mathcal{G}_0}(x) - \deg_{\mathcal{G}}(x) < \varepsilon_3 n. \tag{4.2}$$

Note that the left hand side here is always nonnegative. Denote the set of vertices which fail in (4.2) by A_3. (Here we can choose $\varepsilon_3 = 2\sqrt{\varepsilon_2}$.)

Similar statements are true for $\operatorname{crit} \mathcal{G}$ and $\operatorname{disj} \mathcal{G}_0$. (3.7) and (4.1) implies that

$$(\frac{1}{4} + \varepsilon_2)n^2 < |E(\operatorname{disj} \mathcal{G}_0)| < (\frac{1}{4} + \varepsilon_2)n^2. \tag{4.3}$$

Then (3.5), (3.6) imply that the same is true for $|E(\operatorname{crit} \mathcal{G})|$. Hence we have $|\operatorname{disj} \mathcal{G}_0 - \operatorname{crit} \mathcal{G}| < 2\varepsilon_2 n^2$, which immediately implies the following. There exists a $\varepsilon_4 = \varepsilon(\varepsilon_3)$ and a subset $A_4 \supset A_3$ such that $|A_4| < \varepsilon_4 n$ and

$$\deg_{\operatorname{disj} \mathcal{G}_0}(x) - \deg_{\operatorname{crit} \mathcal{G}}(x) < \varepsilon_4 n \tag{4.4}$$

holds for all $x \notin A_4$. Note that the left hand side of (4.4) is always nonnegative. (We can choose $\varepsilon_4 = 2\varepsilon_3$.)

We claim that there exists an $\varepsilon_5 = \varepsilon(\varepsilon_4)$ and a set $A_5 \supset A_4$ such that $|A_5| < \varepsilon_5 n$ and the following is true for all vertex $x \notin A_5$.

$$\deg_{\mathcal{G}_0}(x) + \deg_{\operatorname{disj} \mathcal{G}_0}(x) > (1 - \varepsilon_5)n. \tag{4.5}$$

Indeed, suppose on the contrary that $|A_5| > \varepsilon_5 n$, where $\varepsilon_5 = 2\varepsilon_4$. Let $B \subset A_5 \setminus A_4$ be an arbitrary subset of size $b \sim \varepsilon_4 n$. Delete B. Then for the obtained graph $\mathcal{G}_0 \setminus B$ we can apply (2.4) and (2.5). We have

$$\begin{aligned} |E(\mathcal{G}_0)| + |\operatorname{disj} \mathcal{G}_0| &\leq |E(\mathcal{G}_0 \setminus B)| + |\operatorname{disj}(\mathcal{G}_0 \setminus B)| + \sum_{x \in B} \deg_{\mathcal{G}_0}(x) + \deg_{\operatorname{disj} \mathcal{G}_0}(x) \\ &\leq (n-b)^2/2 + b(1 - \varepsilon_5)n. \end{aligned}$$

Here the right hand side is less than $\frac{1}{2}(1 - \varepsilon_4^2)n^2$ and the left hand side is greater than $2|E(\mathcal{G}_0)|$ (by (3.7)). This contradicts to (4.1) if $\varepsilon_4 > 2\sqrt{\varepsilon_2}$, say $\varepsilon_4 = 3\sqrt{\varepsilon_2}$. □

$\operatorname{crit}\mathcal{G}$ **has a giant bipartite subgraph.** Let v be a vertex with maximum degree in $\mathcal{G}_0$, i.e. $d = \deg_{\mathcal{G}_0}(v)$, and for all x we have $\deg_{\mathcal{G}_0}(x) \le d$. Denote $N_{\mathcal{G}_0}(v)$ by D, and its complement $V(\mathcal{G}) \setminus D$ by C. By (4.1) we have

$$d > \frac{1}{2}n - \varepsilon_3 n. \tag{4.6}$$

No edge of $\operatorname{disj}\mathcal{G}_0$ is contained in D, hence

$$\deg_{\operatorname{disj}\mathcal{G}_0}(y) \le n - d \tag{4.7}$$

holds for all $y \in D$. Moreover (4.5) (and the choice of D) imply that

$$\deg_{\operatorname{disj}\mathcal{G}_0}(y) \ge n - d - \varepsilon_5 n \tag{4.8}$$

holds for all $y \in D \setminus A_5$. (4.1) and the above inequality (4.7) imply

$$(\frac{1}{4} - \varepsilon_2)n^2 < |E(\mathcal{G}_0)| \le |\operatorname{disj}\mathcal{G}_0| \le \frac{1}{2}(\sum \deg_{\operatorname{disj}\mathcal{G}_0}(x)) \le \frac{1}{2}((n-d)(n-1) + d(n-d)).$$

Hence we have that

$$d < 0.8n, \tag{4.9}$$

holds if ε_2 is sufficiently small. ($\varepsilon_2 < 0.08$.) Consider the bipartite subgraph of $\operatorname{disj}\mathcal{G}_0$, induced by C and D, i.e. $E(\operatorname{disj}\mathcal{G}_0(C,D)) = \{\{x,y\} \in E(\operatorname{disj}\mathcal{G}_0) : x \in C, y \in D\}$. By (4.8) there is a $\varepsilon_6 = \varepsilon(\varepsilon_5)$ and a $A_6 \supset A_5$, $|A_6| < \varepsilon_6 n$ such that for all $x \in C \setminus A_6$ we have

$$\deg_{\operatorname{disj}\mathcal{G}_0(C,D)}(x) > d - \varepsilon_6 n. \tag{4.10}$$

In other words $\operatorname{disj}\mathcal{G}_0(C,D)$ is almost a complete bipartite graph. Then (4.4) implies the following is true for all $x \in C \setminus A_6$

$$|N_{\operatorname{crit}\mathcal{G}}(x) \cap D| \ge d - 2\varepsilon_6 n, \tag{4.11}$$

and for all $y \in D \setminus A_6$

$$|N_{\operatorname{crit}\mathcal{G}}(y) \cap C| \ge n - d - 2\varepsilon_6 n. \tag{4.12}$$

So $\operatorname{crit}\mathcal{G}(C,D)$ is almost a complete bipartite graph , as well.

It is impossible that for some vertex u both

$$|N_{\mathcal{G}_0}(u) \cap C| > 2\varepsilon_6 n \text{ and } |N_{\mathcal{G}_0}(u) \cap D| > 2\varepsilon_6 n \tag{4.13}$$

hold. Suppose on the contrary, and let x be a vertex from $N_{\mathcal{G}_0}(u) \cap C \setminus A_6$. Then, by (4.11), at least $d - 2\varepsilon_6 n$ edges of $\operatorname{crit} \mathcal{G}$ adjacent to y go into D. So there exists an edge $\{x, y\}$ of $\operatorname{crit} \mathcal{G}$ with $y \in N_{\mathcal{G}_0}(u) \cap D$. But then (x, u, y) is the critical path belonging to the critical pair $\{x, y\}$ which contradicts the definition of $\mathcal{G}_0$. □

Suppose that $\varepsilon_6 < 0.01$. Call a vertex u of *type C* (or *D*) if it has more than $2\varepsilon_6 n$ $\mathcal{G}_0$ neighbors in C (in D, rep.). Eventually, a vertex with a small $\mathcal{G}_0$ degree has no type. But as every $\mathcal{G}_0$ degree in $D \setminus A_6$ is between d and $d - \varepsilon_6 n$ we obtain that they have types. Now we distinguish between two subcases. More exactly, we eliminate the first one and then show that the second one leads to the fact that $\mathcal{G}$ is a complete bipartite graph.

1) Suppose first that there are at least $4\varepsilon_6 n$ vertices from $D \setminus A_6$ with type D. Denote the set of these vertices in $D \setminus A_6$ by D_0. If $y \in D \setminus D_0 \setminus A_6$ (i.e. it has type C) then $|N_{G_0}(y) \cap D_0| < 2\varepsilon_6 n$. So the number of $\mathcal{G}_0$ edges between $D \setminus D_0$ and D_0 is at most $|D \setminus D_0| 2\varepsilon_6 n + |A_6||D_0|$. On the other hand every point in $y \in D_0$ has at least

$$|D| - 3\varepsilon_6 n \tag{4.14}$$

$\mathcal{G}_0$) neighbors in D (by (4.5), (4.7) and (4.13)). So y has at least $|D \setminus D_0| - 3\varepsilon_6 n$ neighbors in $D \setminus D_0$. This way we have obtained that

$$|D_0|(|D| - |D_0| - 3\varepsilon_6 n) < (|D| - |D_0|)2\varepsilon_6 n + \varepsilon_6 n|D_0|,$$

which implies that

$$|D_0| > d - 5\varepsilon_6 n. \tag{4.15}$$

So in this case D induce almost a complete graph (in $\mathcal{G}_0$). Consequently, D_0 does not contain an edge from $\operatorname{disj} \mathcal{G}_0$ (and from $\operatorname{crit} \mathcal{G}$) and there is no edge of $\mathcal{G}$ in D_0 of type I.

Consider the induced bipartite subgraph $\mathcal{G}(C \setminus A_6, D_0)$.

PROPOSITION 4.1. *Let $x \in C \setminus D_0$. Then $\deg_{\mathcal{G}(x, D_0)} \leq 1$, i.e there is at most 1 edge of $\mathcal{G}$ from x to D_0.*

Proof. Suppose that there are 2 such edges of $\mathcal{G}$ $\{x, y_1\}$ and $\{x, y_2\}$ with $y_1, y_2 \in D_0$. We have that $|D_O \setminus N_{\mathcal{G}}(y_1) \cup N_{\mathcal{G}}(y_2)|$ is at most $6\varepsilon_6 n$ (by (4.14)). So there exists a critical edge $\{x, z\}$ with $z \in D_0$ and (x, y_1, z) and (x, y_2, z) are two distinct paths in $\mathcal{G}$. (Here we used that

$$|D_0| - 6\varepsilon_6 n > d - 11\varepsilon_6 n > \frac{1}{2}n - 12\varepsilon_6 n > 0,$$

according to (4.14) and (4.6) if we suppose that $\varepsilon_6 < 0.04$.) But this contradicts to the criticality of $\{x, z\}$. □

Let F be the set of those vertices in D_0, which are not connected to $C \setminus A_6$ in $\mathcal{G}$. We claim that

$$|F| < 15\varepsilon_6 n. \tag{4.16}$$

Proof. Let $|F| = f$ and suppose that $f > 15\varepsilon_6 n$. The number of $\mathcal{G}_0$ edges in F is at least

$$\frac{1}{2}f(f - 3\varepsilon_6 n). \tag{4.17}$$

Each of these edges has type II. Now we can easily give an upper bound on the number of critical pairs $\{x, y\}$, such that $x \notin D_0 \cup (C \setminus A_6)$ and $y \in F$. This is less than $6\varepsilon_6 nf$, which is less than (4.17). Hence there exists an edge $\{y, z\}$ of $\mathcal{G}_0$ in F such that it is a part of the critical path (x, y, z), with $x \in C \setminus A_6$. However, in this case $\{x, z\}$ is an edge of $\mathcal{G}$ connecting $C \setminus A_6$ and F, a contradiction. □

Denote $D_0 \setminus F$ by D_1. For every $y \in D_1$, let $x = x(y)$ be a vertex of $C \setminus A_6$ such that $\{x, y\} \in E(\mathcal{G})$. By Proposition 4.1 these second endpoints are all distinct. Let $C_1 = \{x(y) : y \in D_1\}$. Then (4.16) and (4.15) imply that

$$\frac{n}{2} + \varepsilon_6 n > n - d \geq |C_1| = |D_1| > d - 20\varepsilon_6 n > \frac{n}{2} - 21\varepsilon_6 n. \tag{4.18}$$

Let $|D_1| = d_1$. Consider an arbitrary edge $\{x, y\}$ of $\mathcal{G}$ between C_1 and D_1. By (4.12) we have that $y \in D_1$ has at least $d_1 - 2\varepsilon_6 n$ crit $\mathcal{G}$ neighbors in C_1. If $\{x', y'\} \in E(\mathcal{G})$ and $\{x', y\}$ is a critical pair, then either $\{x, x'\}$ or $\{y, y'\}$ is not in $E(\mathcal{G})$. This implies, that

$$\deg_{\mathcal{G}(C_1)}(x) + \deg_{\mathcal{G}(D_1)}(y) \leq d_1 + 2\varepsilon_6 n. \tag{4.19}$$

Summing up (4.19) for all edge of $\mathcal{G}$ connecting C_1 and D_1 we obtain that

$$|E(\mathcal{G}(C_1))| + |E(\mathcal{G}(D_1))| < \frac{1}{2}d_1(d_1 + 2\varepsilon_6 n) < \frac{1}{8}n^2 + \frac{1}{2}\varepsilon_6 n^2. \tag{4.20}$$

It is obvious that the number of edges of $\mathcal{G}$ not included in $C_1 \cup D_1$ is not more than

$$(n - 2d_1)d < 42\varepsilon_6 n^2 \tag{4.21}$$

Finally the sum of (4.20) and (4.21) is less than $\frac{1}{2}n^2 - \varepsilon_6 n^2$ if we suppose that $\varepsilon_6 < 0.002$. This contradicts to (4.1).

2) We may suppose that the number of vertices in $D \setminus A_6$ with type D is less than $4\varepsilon_6 n$. Then for almost all but at most $4\varepsilon_6 n$ vertices y of $D \setminus A_6$ the following holds. y has at least $n - d - 3\varepsilon_6 n$ crit $\mathcal{G}$ neighbors in C by (4.12) and it has at least $n - d - \varepsilon_5 n - 2\varepsilon_6 n$ $\mathcal{G}_0$ neighbors as well, by (4.5) and (4.13). So it has at least $n - d - 6\varepsilon_6 n$ $\mathcal{G}$ neighbors in C of type I. Then it is easy to see that a similar statement is true for the points of C, too. That is , there exists a $\varepsilon_7 = \varepsilon(\varepsilon_6)$ and a set $A_7 \supset A_6$, $|A_7| < \varepsilon_7 n$ such that every vertex $y \in D \setminus A_7$ has $|C| - \varepsilon_7 n$ $\mathcal{G}$ neighbors of type I in C and every vertex $x \in C \setminus A_7$ has at least $|D| - \varepsilon_7 n$ $\mathcal{G}$ neighbors of type I in D. (Here we used that $0.4n < d < 0.8n$. We can define ε_7 as $10\sqrt{\varepsilon_6}$.)

Consequently, $D \setminus A_7$ does not contain any edge from $\mathcal{G}$, neither from crit $\mathcal{G}$. (Neither from $\mathcal{G}_0$ and disj $\mathcal{G}_0$, but from now on we won't deal with $\mathcal{G}_0$, we return to investigate directly $\mathcal{G}$.) Split A_7 into three parts. Let A_8 consist of those vertices of A_7 whose degree is at most $0.4n$. Let C_8 (D_8) consists of those vertces of $A_7 \setminus A_8$ which have more than $0.2n$ $\mathcal{G}$ neighbors in $D \setminus A_7$ (in $C \setminus A_7$, resp.). As a type I edge never appears in a triangle we have that $C_8 \cap D_8 = \emptyset$. Furthermore, there is no edge of $\mathcal{G}$ neither from crit $\mathcal{G}$ connecting C_8 to $C \setminus A_7$ (D_8 to $D \setminus A_7$, resp.). Hence if $v \in C_8$ then it has at least $0.4n$ neighbors in D. Obviously, we have that the number of edges adjacent to A_8 is

$$\leq 0.4n|A_8|. \tag{4.22}$$

For brevity use the notations $C' = C \setminus A_7 \cup C_8$, $c' = |C'|$ and $D' = D \setminus A_7 \cup D_8$, $d' = |D'|$. Now we classify the edges of $\mathcal{G}$ in $C' \cup D'$.

(i) First of all we have the edges connecting C' and D'.

(ii/C) In this class we have those edges $\{u, v\}$ which are contained in C' (so in C_8) and are part of a critical path (u, v, w) with $w \in D'$.

(ii/D) The definition is analogous to (ii/C), i.e. $\{u, v\} \in E(\mathcal{G})$ belongs to this class if $u, v \in D_8$ and there is a critical path (u, v, w) with $w \in C'$.

(iii) The rest of the edges in $C' \cup D'$.

Consider an edge $\{u, v\}$ of type (iii). Say, it is included in C_8. As u and v have a lot of common neighbors in D' the type of $\{u, v\}$ is II. Then it is a part of a critical path (u, v, w), and by definition, $w \notin D'$. But $w \notin C'$, too, otherwise we could find several common neighbors of u and w. So $\{u, v\}$ belongs to a critical pair $\{u, w\}$ with $w \in A_8$. The number of such critical pairs is bounded above by $|A_7||A_8|$, hence

$$\#(iii) \leq |A_7||A_8| \leq \varepsilon_7 n|A_8|. \tag{4.23}$$

For each edge $\{u, v\}$ from the class (ii) one can associate a critical pair $\{u, w\}$ such that one of u and w lies in C' and the other lies in D' but $\{u, w\}$ is not an edge of $\mathcal{G}$. The pair $\{u, w\}$ does not belong to another edge of type (ii) so in this way we have that the number of edges in the class (ii) is not more than the number of non-edges between C' and D'. In other words the number of edges of types (i) and (ii) is at most $|C'||D'|$. This and (4.22) and (4.23) give

$$|E(\mathcal{G})| \leq c'd' + (n - c' - d')(0.4n + \varepsilon_7 n). \tag{4.24}$$

Here the right hand side is at most $\lfloor n^2/4 \rfloor$, as desired. □

Equality can hold in (4.24) only if $A_8 = \emptyset$, and then there is no edge of type (iii) (by (4.23)). Moreover every non-edge between C' and D' must be a critical pair. There is no edge between $C'\setminus C_8$ and C_8 (and between $D'\setminus D_8$ and D_8), so there is no critical pair which is non-edge outside $C_8 \cup D_8$. So, $\mathcal{K}(C' \setminus C_8, D' \setminus D_8)$ is a subgraph of $\mathcal{G}$, and all of its edges have type I. Suppose that $|E(\mathcal{G}(C_8))| \geq |E(\mathcal{G}(D_8))|$, and let $P = \{v \in C_8 : \deg_{\mathcal{G}(C_8)} > 0\}$, $p = |P|$. Then

$$(4.25) \qquad |E(\mathcal{G}(C_8))| + |E(\mathcal{G}(D_8))| \leq p(p-1) < \varepsilon_7 np.$$

If $x \in P$ then there are at least $n/4$ edges $\{x, y\}$, $y \in D'$ of type II. Indeed, there is an edge $\{x, z\}$ contained in P, and then all the edges from x to $D' \cap N(x) \cap N(z)$ have type II. So the number of edges of type II between C_8 and D' is at least $pn/4$. Each such an edge is a part of a critical path of lenght two, with a critical pair between C' and D'. So the number of non-edges between C' and D' is much more than (4.25), if $p > 0$. Thus $|E(\mathcal{G})| \geq \lfloor n^2/4 \rfloor$ implies that $p = 0$. That is, $\mathcal{G}$ is a bipartite graph, and then a complete bipartite one. □

5. Remarks, problems

We can construct a large non-bipartite minimal graph $\mathcal{M}$ of diameter 2 as follows. Let $V(\mathcal{M}) = X \cup Y \cup \{z\}$ where $|X| = \lfloor (n-1) \rfloor$, $|Y| = \lceil (n-1) \rceil$, and let $x \in X$, $y \in Y$. The graph $\mathcal{M}$ obtained from the complete bipartite graph $\mathcal{K}(X, Y)$ by deleting the edge $\{x, y\}$ and adding the edges $\{x, z\}$ and $\{z, y\}$. With a little more effort the above proof gives the following slightly stronger statement.

Theorem 5.1. *Suppose that $\mathcal{G}$ is a minimal graph of diameter 2 over n elements, $n > n_0$. If $|E(\mathcal{G})| \geq \lfloor (n-1)^2/4 \rfloor + 1$, then either $\mathcal{G}$ is a complete bipartite graph, or it is isomorphic to $\mathcal{M}$.* □

Let now $\mathcal{G}$ be an arbitrary graph with n vertices. Let k be an integer and define $\operatorname{disj}_k \mathcal{G}$ as follows. The pair $\{x, y\}$ belongs to $\operatorname{disj}_k \mathcal{G}$ if they have at most k common neighbors, i.e. $|N(x) \cap N(y)| \leq k$. In this way $\operatorname{disj} \mathcal{G}$ defined above is just $\operatorname{disj}_0 \mathcal{G}$. If we use directly the Szemerédi lemma instead of Theorem 3.2, then we can obtain the following statement, which was the essence of the proof presented in the Section 3.

Theorem 5.2. *Let k be a fixed integer. Then from any graph $\mathcal{G}$ over n vertices one can remove $o(n^2)$ edges such that the following holds. If x and y had at most k common neighbors in $\mathcal{G}$ then in the obtained new graph $\mathcal{G}_k$ they have no common neighbor anymore. I.e., $\operatorname{disj}_k \mathcal{G} \subset \operatorname{disj} \mathcal{G}_k$.* □

The following conjecture generalizes our main theorem.

Conjecture 5.3. *Let $\mathcal{G}$ be a graph over n vertices and suppose that every two vertex is connected by at least k paths of length at most 2. Suppose further that $\mathcal{G}$ is minimal with respect this property. Then $|E(\mathcal{G})| \leq (k-1)(n-k+1) + \lfloor (n-k+1)^2/4 \rfloor$.*

Here the extremal graph would be complete 3-colored graph with parts of sizes $\lfloor (n-k+1)/2 \rfloor$, $\lceil (n-k+1)/2 \rceil$ and $k-1$. Caccetta and Häggkvist raised the following conjectures which also generalize Conjecture 1.1.

CONJECTURE 5.4 [CH]. *If $\mathcal{G}$ is a minimal graph of diameter 2, then $d' \le |V(\mathcal{G})|$, where d' denotes the average edge degree in $\mathcal{G}$, i.e.*

$$d' = \sum_{\{x,y\}\in E(\mathcal{G})} (\deg(x) + \deg(y)) = \sum_{x\in V(\mathcal{G})} (\deg(x))^2/|E(\mathcal{G})|.$$

CONJECTURE 5.5 [CH]. *If $\mathcal{G}$ is a minimal graph of diameter k, with $k > 2$, then $|E(\mathcal{G})| \le (1+o(1))n^2/2(k+1)^2$.*

The conjectured extremal graph consists of two complete bipartite graphs $\mathcal{K}(A_0, A_1)$ and $\mathcal{K}(A_{k-1}, A_k)$ where $|A_i| \sim n/(k+1)$, and $|A_1| = |A_{k-1}|$, and $|A_1|$ disjoint path of length $k-2$ connecting the points of A_1 to A_{k-1}. The method presented in this paper does not seem to be applicable in proving Conjecture 5.4, but maybe useful to attack the last one. To find further problems (and results) about diameter critical graphs one can see, e.g., **[Chu]** or **[B]**.

REFERENCES

[B] B. BOLLOBÁS, *Extremal graph theory*, Academis Press, London, 1978.

[CH] L. CACCETTA AND R. HÄGGKVIST, *On diameter critical graphs*, Discrete Mathematics, 28 (1979), pp. 223–229. MR80k:05070

[Chu] F. R. K. CHUNG, *Diameters of communication networks*, Proc. Symposia in Appl. Math. AMS, 34 (1986), pp. 1–18.

[EFR] P. ERDŐS, P. FRANKL, AND V. RÖDL, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs and Combinatorics, 2 (1986), pp. 113–121.

[EK] P. ERDŐS, AND D. J. KLEITMAN, *On coloring graphs to maximize the proportion of multicolored k-edges*, Journal of Combinatorial Theory, 5 (1968), pp. 164–169.

[F] G. FAN, *On diameter 2-critical graphs*, Discrete Mathematics, 67 (1987), pp. 235–240.

[P] J. PLESNÍK, *Critical graphs of given diameter*, Acta Fac. Rerum Natur. Univ. Comenianae Math., 30 (1975), pp. 71–93. MR53#2755

[RSz] I. Z. RUZSA, AND E. SZEMERÉDI, *Triple systems with no six points carrying three triangles*, in *Combinatorics*, (Proc. Colloq., Keszthely, 1976) *Proc. Colloq. Math. Soc. J. Bolyai* **18**, Vol. II., North-Holland, Amsterdam-New York, 1978, pp. 939–945. MR80c:05116

[X] J. M. XU, *Proof of a conjecture of Simon and Murty*, (Chinese), J. Math. Res. Exposition, 4 (1984), pp. 85–86. Corrigendum: *ibid* 5 (1985), p. 38. MR87i:05125