Every graph of sufficiently large average degree contains a C_4-free subgraph of large average degree

Daniela Kühn Deryk Osthus*

Abstract

We prove that for every k there exists $d = d(k)$ such that every graph of average degree at least d contains a subgraph of average degree at least k and girth at least six. This settles a special case of a conjecture of Thomassen.

1 Introduction

Thomassen [6] conjectured that for all integers k, g there exists an integer $f(k, g)$ such that every graph G of average degree at least $f(k, g)$ contains a subgraph of average degree at least k and girth at least g (where the *average degree* of a graph G is $d(G) := 2e(G)/|G|$ and the *girth* of G is the length of the shortest cycle in G). Erdős and Hajnal [2] made a conjecture analogous to that of Thomassen with both occurrences of average degree replaced by chromatic number. The case $g = 4$ of the conjecture of Erdős and Hajnal was proved by Rödl [5], while the general case is still open.

The existence of graphs of both arbitrarily high average degree and high girth follows for example from the result of Erdős that there exist graphs of high girth and high chromatic number. The case $g = 4$ of Thomassen's conjecture (which corresponds to forbidding triangles) is trivial since every graph can be made bipartite by deleting at most half of its edges. Thus $f(k, 4) \leq 2k$. The purpose of this paper is to prove the case $g = 6$ of the conjecture.

Theorem 1. *For every k there exists $d = d(k)$ such that every graph of average degree at least d contains a subgraph of average degree at least k whose girth is at least six.*

A straightforward probabilistic argument shows that Thomassen's conjecture is true for graphs G which are almost regular in the sense that their maximum degree is not much larger than their average degree (see Lemma 4 for the C_4-case). Indeed, such graphs G do not contain too many short cycles. Thus if we consider the graph G_p obtained by selecting each edge of G with probability p (for a suitable p), it is easy to show that with nonzero probability G_p contains far fewer short cycles than edges. Deleting one edge on every short cycle then yields a subgraph of G with the desired properties.

*Graduate school "Combinatorics, Geometry and Computation", supported by Deutsche Forschungsgemeinschaft grant GRK/588-1

Thus the conjecture would hold in general if every graph of sufficiently large average degree would contain an almost regular subgraph of large average degree. However, this is not the case: Pyber, Rödl and Szemerédi [4] showed that there are graphs with $cn \log\log n$ edges which do not contain a k-regular subgraph (for all $k \geq 3$). These graphs cannot even contain an almost regular subgraph of large average degree, since e.g. another result in [4] states that every graph with at least $c_k n \log(\Delta(G))$ edges contains a k-regular subgraph. On the other hand, the latter result implies that every graph G with at least $c_k n \log n$ edges contains a k-regular subgraph (which was already proved by Pyber [3]), and thus, if k is sufficiently large, G contains also a subgraph of both high average degree and high girth.

2 Proof of the theorem

We say that a graph is C_4-*free* if it does not contain a C_4 as a subgraph. We prove the following quantitative version of Theorem 1. (It implies Theorem 1 since every graph can be made bipartite by deleting at most half of its edges.) We remark that we have made no attempt to optimize the bounds given in the theorem.

Theorem 2. *Let $k \geq 2^{16}$ be an integer. Then every graph of average degree at least $64k^{3+2\cdot 11^{64k^3}}$ contains a C_4-free subgraph of average degree at least k.*

We now give a sketch of the proof of Theorem 2. As a preliminary step we find a bipartite subgraph (A, B) of the given graph G which has large average degree and where the vertices in A all have the same degree. We then inductively construct a C_4-free subgraph of (A, B) in the following way. Let $a_1, a_2, \dots$ be an enumeration of the vertices in A. At stage i we will have found a C_4-free subgraph G_i of (A, B) whose vertex classes are contained in $\{a_1, \dots, a_i\}$ and B, and such that the vertices in $V(G_i) \cap A$ all have the same degree in G_i. We then ask whether the subgraph of (A, B) consisting of G_i together with all the edges of G incident with a_{i+1} (and their endvertices) contains many C_4's. If this is the case, the vertex a_{i+1} is 'useless' for our purposes. We then let $G_{i+1} := G_i$ and consider the next vertex a_{i+2}. But if a_{i+1} is not 'useless', we add a_{i+1} together with suitable edges to G_i to obtain a new C_4-free graph G_{i+1}. We then show that either the C_4-free graph G^* consisting of the union of all the G_i has large average degree or else that there is a vertex $x \in B$ and a subgraph (A', B') of $(A, B) - x$ which has similar properties as (A, B) and such that $A' \subseteq N(x)$ (Lemma 6). In the latter case, we apply the above procedure to this new graph (A', B'). If this again does not yield a C_4-free subgraph with large average degree, there will be a vertex $x' \in B'$ and a subgraph (A'', B'') of $(A', B') - x'$ as before. So both x and x' are joined in G to all vertices in A''. Continuing this process, we will either find a C_4-free subgraph with large average degree or else a large $K_{s,s}$. But $K_{s,s}$ is regular and so, as was already mentioned in Section 1, it contains a C_4-free subgraph as required (Lemma 4).

We shall frequently use the following basic fact [1, Prop. 1.2.2.].

Proposition 3. *Every graph of average degree d contains a subgraph of minimum degree at least $d/2$.* □

The following lemma implies that Theorem 2 holds for the class of all graphs whose maximum degree is not much larger than their average degree. It can easily be generalized to longer cycles.

Lemma 4. *If G is a graph of average degree d and maximum degree αd, then G contains a C_4-free subgraph of average degree at least $d^{1/3}/(4\alpha)$.*

Proof. Let $n := |G|$ and put $k := d^{1/3}/(4\alpha)$. Let G_p denote the (random) spanning subgraph of G obtained by including each edge of G in G_p with probability $p := 2k/d$. Let X_4 denote the number of labelled C_4's in G_p and let X_e denote the number of edges in G_p. Then $\mathbb{E}[X_e] = pdn/2$. Since the number of C_4's contained in G is at most $\frac{dn}{2}(\alpha d)^2$ (indeed, every C_4 is determined by first choosing an edge $xy \in G$ and then choosing a neighbour of x and a neighbour of y so that these neighbours are joined by an edge in G), it follows that

$$\mathbb{E}[X_4] \leq \frac{dn}{2}(\alpha d)^2 p^4 \leq \frac{8\alpha^2 k^3}{d} \cdot p \cdot \frac{dn}{2} \leq \mathbb{E}[X_e]/2.$$

Let $X := X_e - X_4$. Then by the above, $\mathbb{E}[X] \geq \mathbb{E}[X_e]/2 = pdn/4 = kn/2$. Thus $\mathbb{P}[X \geq kn/2] > 0$, and so G contains a subgraph H with the property that if we delete an edge from each C_4 in H, the remaining graph H' still has at least $kn/2$ edges. Thus H' is as desired. □

Proposition 5. *Let $D > 0$, $0 \leq c_0 < 1$ and $c_1 \geq 1$. Let $G = (A, B)$ be a bipartite graph with at least $D|A|$ edges and such that $d(a) \leq c_1 D$ for every vertex $a \in A$. Then there are at least $(1 - c_0)/(c_1 - c_0)|A|$ vertices $a \in A$ with $d(a) \geq c_0 D$.*

Proof. Let t denote the number of vertices $a \in A$ with $d(a) \geq c_0 D$. Then $c_1 Dt + c_0 D(|A| - t) \geq e(G) \geq D|A|$, which implies that $t(c_1 D - c_0 D) \geq |A|(D - c_0 D)$. □

Given $c, d \geq 0$, we say that a bipartite graph (A, B) is a *(d, c)-graph* if A is non-empty, $|B| \leq c|A|$ and $d(a) = \lceil d \rceil$ for every vertex $a \in A$. Given a graph G and disjoint sets $A, B \subseteq V(G)$, we write $(A, B)_G$ for the induced bipartite subgraph of G with vertex classes A and B.

Lemma 6. *Let $c, d \in \mathbb{N}$ be such that d is divisible by c, $c \geq 2^{16}$ and $d \geq 4c^3$. Let $G = (A, B)$ be a $(d/c, c)$-graph. Then G contains either a C_4-free subgraph of average degree at least c or there exists a vertex $x \in B$ and a $(d/c^{11}, c^{11})$-graph $(A', B') \subseteq G$ such that $A' \subseteq N(x)$ and $B' \subseteq B \setminus \{x\}$.*

Proof. Given a bipartite graph (X, Y) and a set $Y' \subseteq Y$, we say that a path P of length two whose endvertices both lie in Y' is a *hat* of Y', and that the endvertices of P *span* this hat.

Let $a_1, a_2, \ldots$ be an enumeration of the vertices in A. Let us define a sequence $A_0 \subseteq A_1 \subseteq \ldots$ of subsets of A and a sequence $G_0 \subseteq G_1 \subseteq \ldots$ of subgraphs of G such that the following holds for all $i = 0, 1, \ldots$:

G_i is C_4-free and has vertex classes $A_i \subseteq \{a_1, \dots, a_i\}$ and B, and $d_{G_i}(a) = 2c^2$ for every $a \in A_i$.

To do this, we begin with $A_0 := \emptyset$ and the graph G_0 consisting of all vertices in B (and no edges). For every $i \geq 1$ in turn, we call the vertex a_i *useless* if $N_G(a_i)$ spans at least $d^2/(8c^4)$ hats contained in G_{i-1}. If a_i is useless, we put $A_i := A_{i-1}$ and $G_i := G_{i-1}$. If a_i is not useless, let us consider the auxiliary graph H on $N_G(a_i)$ in which two vertices $x, y \in N_G(a_i)$ are joined if they span a hat contained in G_{i-1}. Since a_i is not useless, we have that

$$\begin{aligned} e(\overline{H}) = & \binom{d_G(a_i)}{2} - e(H) \geq \left(\frac{d/c - 1}{d/c} - \frac{1}{4c^2}\right) \frac{d_G(a_i)^2}{2} \\ \geq & \left(1 - \frac{1}{2c^2}\right) \frac{d_G(a_i)^2}{2}, \end{aligned}$$

where the last inequality holds since $d \geq 4c^3$. Turán's theorem (see e.g. [1, Thm. 7.1.1.]) applied to $\overline{H}$ now shows that H contains an independent set of size at least $2c^2$. Hence there are $2c^2$ edges of G incident with a_i such that the graph consisting of G_{i-1} together with a_i and these edges does not contain a C_4. We then let G_i be this graph and put $A_i := A_{i-1} \cup \{a_i\}$.

Let $A^* := \bigcup_i A_i$ and $G^* := \bigcup_i G_i$. Thus the accepted graph G^* is C_4-free. Let $A^1 := A \setminus A^*$, and let $G^1 := (A^1, B)_G$. We show that either G^* has average degree at least c (which correponds to Case 1 below) or else that there are $x \in B$ and (A', B') as in the statement of the lemma (Case 2). We will distinguish these two cases according to the properties of the neighbourhoods and the second neighbourhoods of the vertices in B. For this, we need some definitions.

For every $a \in A^1$ consider the auxiliary graph H_a on $N_{G^1}(a) = N_G(a)$ in which two vertices are joined by an edge if they span a hat contained in the accepted graph G^*. Since a is useless, this graph has at least $d^2/(8c^4)$ edges (and d/c vertices), and so it has average degree at least $d/(4c^3)$. By Proposition 3, H_a contains a subgraph H'_a with minimum degree at least $d/(8c^3)$, and so with at least $1 + d/(8c^3)$ vertices. Let $B^2 := \bigcup_{a \in A^1} V(H'_a)$, and let G^2 be the subgraph of G^1 whose vertex set is $A^1 \cup B^2$ and in which every $a \in A^1$ is joined to all of $V(H'_a)$. Thus the following holds.

For every $a \in A^1$ we have that $d_{G^2}(a) \geq 1 + d/(8c^3)$, and every vertex in $N_{G^2}(a)$ spans a hat contained in G^* with at least $d/(8c^3)$ other vertices in $N_{G^2}(a)$. $(*)$

Given any vertex $x \in B^2$, let G^2_x denote the subgraph of G^2 induced by the vertices in $A^2_x := N_{G^2}(x)$ and $B^2_x := N_{G^2}(N_{G^2}(x)) \setminus \{x\}$. Let

$$u := \frac{d}{2^8 c^7},$$

and say that a vertex $b \in B^2_x$ is *x-rich* if $d_{G^2_x}(b) \geq u$.

Case 1. *For every vertex $x \in B^2$ we have that*

$$\sum_{b \in B_x^2,\ b \text{ is } x\text{-rich}} d_{G_x^2}(b) \le \frac{e(G_x^2)}{16c^2}. \tag{1}$$

We will show that in this case, every vertex $x \in B^2$ is incident with at least $8c^2 d_{G^2}(x)$ edges of the accepted graph G^* and thus that $e(G^*) \ge 8c^2 e(G^2)$. Before doing this, let us first show that the latter implies that the average degree of G^* is at least c. Indeed, since $e(G^1) = d|A^1|/c$, we have

$$e(G^2) \overset{(*)}{\ge} \frac{d}{8c^3}|A^1| = \frac{1}{8c^2} e(G^1).$$

Thus $e(G^*) \ge e(G^1)$. Also $d_{G^*}(a) = 2c^2$ for every $a \in A^*$ while $d_{G^1}(a) = d/c \ge 2c^2$ for every $a \in A^1$, and so

$$d(G^* \cup G^1) \ge \frac{2 \cdot 2c^2|A|}{|A| + |B|} \ge \frac{4c^2|A|}{(1+c)|A|} \ge 2c.$$

Recalling that $e(G^*) \ge e(G^1)$, this now shows that $d(G^*) \ge d((G^* \cup G^1) - E(G^1)) \ge d(G^* \cup G^1)/2 \ge c$.

Thus it suffices to show that $d_{G^*}(x) \ge 8c^2 d_{G^2}(x)$ for every vertex $x \in B^2$. So let $x \in B^2$, and put $t := d_{G^2}(x) = |A_x^2|$. Let B_x^3 be the subset of B_x^2 obtained by deleting all x-rich vertices, and let $G_x^3 := (A_x^2, B_x^3)_{G_x^2}$. Let $y_1, \dots, y_t$ be an enumeration of the vertices in A_x^2. For all $i = 1, \dots, t$, let N_i denote the set of all vertices in $N_{G_x^2}(y_i) = N_{G^2}(y_i) \setminus \{x\}$ spanning a hat with x which is contained in G^*. Hence by $(*)$

$$|N_i| \ge \frac{d}{8c^3}. \tag{2}$$

We now use the existence of these hats to show that x is incident with at least $8c^2 t$ edges of G^* (namely edges contained in these hats). Let $N_i' := N_i \cap B_x^3$ and $n_i := |N_i \setminus N_i'|$. Thus $n_i \le d_{G_x^2}(y_i) - d_{G_x^3}(y_i)$, and so

$$\sum_{i=1}^{t} n_i \le e(G_x^2) - e(G_x^3) \overset{(1)}{\le} \frac{e(G_x^2)}{16c^2} \le \frac{dt}{16c^3}.$$

Hence

$$\sum_{i=1}^{t} |N_i'| = \sum_{i=1}^{t} (|N_i| - n_i) \overset{(2)}{\ge} \frac{dt}{8c^3} - \frac{dt}{16c^3} = \frac{dt}{16c^3}.$$

But every vertex of G_x^3 lies in at most u of the sets $N_1', \dots, N_t'$, since $d_{G_x^3}(b) \le u$ for every $b \in B_x^3$. Thus

$$\left| \bigcup_{i=1}^{t} N_i' \right| \ge \frac{1}{u} \sum_{i=1}^{t} |N_i'| \ge 16c^4 t.$$

That means that x spans hats contained in G^* with at least $16c^4 t$ other vertices in B_x^3. But as every vertex in A^* has degree $2c^2$ in G^*, this implies that x

has at least $16c^4t/(2c^2) \geq 8c^2t$ neighbours in G^*. So we have shown that $d_{G^*}(x) \geq 8c^2 d_{G^2}(x)$ for every $x \in B^2$, as desired.

Case 2. *There exists a vertex $x \in B^2$ not satisfying (1).*

Let B_x^4 be the set of all x-rich vertices in B_x^2, let $G_x^4 := (A_x^2, B_x^4)_{G_x^2}$ and put $t := d_{G^2}(x) = |A_x^2|$. Then the choice of x implies that $t > 0$ and

$$e(G_x^4) \geq \frac{e(G_x^2)}{16c^2} \overset{(*)}{\geq} \frac{1}{16c^2} \cdot \frac{dt}{8c^3} = \frac{dt}{2^7c^5}.$$

Hence the average degree in G_x^4 of the vertices in A_x^2 is at least $D' := d/(2^7c^5)$. Proposition 5, applied with $D = D'$, $c_0 = 1/2$ and $c_1 = d/(cD') = 2^7c^4$, now implies that there are at least

$$\frac{1-c_0}{c_1-c_0} \cdot t = \frac{t}{2(c_1-\frac{1}{2})} \geq \frac{t}{2c_1} = \frac{t}{2^8c^4}$$

vertices $a \in A_x^2$ with $d_{G_x^4}(a) \geq D'/2 \geq d/c^{11}$. Let A_x^4 be the set of these vertices. Thus $|A_x^4| \geq t/(2^8c^4)$. But then the subgraph of $(A_x^4, B_x^4)_{G_x^4}$ obtained by deleting edges so that every vertex in A_x^4 has degree $\lceil d/c^{11} \rceil$ is a $(d/c^{11}, c^{11})$-graph. Indeed, the only thing that remains to be checked is that $|B_x^4| \leq c^{11}|A_x^4|$. But since

$$u|B_x^4| = \frac{d}{2^8c^7}|B_x^4| \leq e(G_x^4) \leq \frac{td}{c} \leq 2^8c^3d|A_x^4|,$$

this follows by recalling that $c \geq 2^{16}$. □

We can now put everything together.

Proof of Theorem 2. We may assume (by deleting edges if necessary) that the given graph G has average degree $d := 64k^{3+2\cdot 11^{64k^3}}$. Pick a bipartite subgraph G' of G which has average degree at least $d/2$. By Proposition 3, there is a (bipartite) subgraph G'' of G' which has minimum degree at least $d/4$. Let A and B be the vertex classes of G'', where $|A| \geq |B|$. Let G_0 be the subgraph of G'' obtained by deleting sufficiently many edges to ensure that all vertices in A have degree exactly d/k. Thus G_0 is a $(d/k, k)$-graph. We now apply Lemma 6 to G_0. If this fails to produce a C_4-free subgraph of average degree at least k, we obtain a vertex $x_1 \in B_0$ and a $(d/k^{11}, k^{11})$-graph $G_1 = (A_1, B_1)$ with $A_1 \subseteq N_{G_0}(x_1)$ and $B_1 \subseteq B_0 \setminus \{x_1\}$ to which we can apply Lemma 6 again. Continuing in this way, after $s := 64k^3$ applications of Lemma 6, we either found a C_4-free subgraph of average degree at least k, or sequences $x_1, \dots, x_s$ and $G_1 = (A_1, B_1), \dots, G_s = (A_s, B_s)$, where G_s is a $(d/k^{11^s}, k^{11^s})$-graph. But then each x_i is joined in G to every vertex in A_s. Since A_s is non-empty, we have $|B_s| \geq d/k^{11^s}$ and so in fact

$$|A_s| \geq |B_s|/k^{11^s} \geq d/k^{2\cdot 11^s} = s.$$

Thus G contains the complete bipartite graph $K_{s,s}$. The result now follows by applying Lemma 4 to this $K_{s,s}$. □

References

[1] R. Diestel, *Graph Theory*, Springer-Verlag 1997.

[2] P. Erdős, Problems and results in chromatic graph theory. In: F. Harary, editor, *Proof Techniques in Graph Theory*, Academic Press 1969, 27–35.

[3] L. Pyber, Regular subgraphs of dense graphs, *Combinatorica* **5** (1985), 347–349.

[4] L. Pyber, V. Rödl and E. Szemerédi, Dense graphs without 3-regular subgraphs, *J. Combin. Theory B* **63** (1995), 41–54.

[5] V. Rödl, On the chromatic number of subgraphs of a given graph, *Proc. Amer. Math. Soc.* **64** (1977), 370–371.

[6] C. Thomassen, Girth in graphs, *J. Combin. Theory B* **35** (1983), 129–141.

Daniela Kühn
Mathematisches Seminar
Universität Hamburg
Bundesstraße 55
D - 20146 Hamburg
Germany
E-mail address: `kuehn@math.uni-hamburg.de`

Deryk Osthus
Institut für Informatik
Humboldt-Universität zu Berlin
Unter den Linden 6
D - 10099 Berlin
Germany
E-mail address: `osthus@informatik.hu-berlin.de`