Filling the gap between Turán's theorem and Pósa's conjecture

Peter Allen [*] Julia Böttcher [†] Jan Hladký [‡]

June 25, 2009

Abstract

Much of extremal graph theory has concentrated either on finding very small subgraphs of a large graph (such as Turán's theorem) or on finding spanning subgraphs (such as Dirac's theorem or more recently the Pósa conjecture). Only a few results give conditions to obtain some intermediate-sized subgraph. We contend that this neglect is unjustified. In this paper we investigate minimum-degree conditions under which a graph G contains squared paths and squared cycles of arbitrary specified lengths. We determine precise thresholds, assuming that the order of G is large. This extends results of Fan and Kierstead [J. Combin. Theory Ser. B 63 (1995), 55–64] and of Komlós, Sarközy, and Szemerédi [Random Structures Algorithms 9 (1996), 193–211] concerning containment of a spanning squared paths and a spanning squared cycle, respectively.

1 Introduction

One of the main programmes of extremal graph theory is the study of conditions on the vertex degrees of a host graph G under which a target graph H

[*]DIMAP, Computer Science Building, University of Warwick, CV4 7AL, United Kingdom. Email: P.D.Allen@warwick.ac.uk.

[†]Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, D-85747 Garching bei München, Germany. Email: boettche@ma.tum.de.

[‡]Department of Applied Mathematics, Faculty of Mathematics and Physics, Charles University, Malostranské náměstí 25, 118 00, Prague, Czech Republic and Zentrum Mathematik, Technische Universität München, Boltzmannstraße 3, D-85747 Garching bei München, Germany. Email: honzahladky@googlemail.com. Partially supported by Grant Agency of Charles University under GAUK 202-10/258009, by DAAD and by BAYHOST.

appears as a subgraph of G (which we denote by $H \subseteq G$). Turán's theorem [19] is a prominent example for results of this type. It asserts that an average degree $d(G) > \frac{r-2}{r-1}n$ forces the copy of a complete graph K_r in G (and that this is best possible) where here and throughout n is the number of vertices in the host graph G. More generally, the celebrated theorem of Erdős and Stone [5] implies that for a *fixed* graph H the chromatic number $\chi(H)$ of H determines the average degree that is necessary to guarantee a copy of H: If H has chromatic number $\chi(H) = r$ and $d(G) \geq (\frac{r-2}{r-1} + o(1))n$, then H is a subgraph of G. This settles the problem for fixed target graphs, that is, graphs that are 'small' compared to the host graph.

Dirac's theorem [4], another classical result from the area, considers target graphs that are of the same order as the host graph, i.e., so-called *spanning* target graphs. Clearly, any average degree condition on the host graph that enforces a connected spanning subgraph must be trivial, and hence the average degree needs a suitable replacement in this setting. Here, the minimum degree is a natural candidate and indeed, Dirac's theorem asserts that every graph G with minimum degree $\delta(G) > \frac{1}{2}n$ has a Hamilton cycle. This implies in particular that G has a matching covering $2\lfloor n/2 \rfloor$ vertices.

A 3-chromatic version of this result follows from a theorem by Corrádi and Hajnal [3]: the minimum degree condition $\delta(G) \geq 2\lfloor n/3 \rfloor$ implies the existence of a so-called *spanning triangle factor* in G, that is, a collection of $\lfloor n/3 \rfloor$ vertex disjoint triangles. A well-known conjecture of Pósa (see, e.g., [6]) asserts that roughly the same minimum degree actually guarantees the existence of a connected super-graph of a spanning triangle factor. It states that any graph G with $\delta(G) \geq \frac{2}{3}n$ contains a spanning *squared cycle* C_n^2, which is the square of a cycle C_n on n vertices (where the square of a graph F is obtained from F by adding edges between all pairs of vertices with distance 2). This can be seen as a 3-chromatic analogue of Dirac's theorem which turned out to be much more difficult than its 2-chromatic cousin (for n divisible by 3 and 2, respectively).

Fan and Kierstead [7] proved an approximate version of Pósa's conjecture for large n. In addition they determined a sufficient and best possible minimum degree condition for the case that the squared cycle in Pósa's conjecture is replaced by a *squared path* P_n^2, i.e., the square of a spanning path P_n.

Theorem 1 (Fan & Kierstead [8]). *If G is an n-vertex graph with minimum degree $\delta(G) \geq (2n-1)/3$, then G contains a spanning squared path P_n^2.*

The Pósa Conjecture was verified for large values of n by Komlós, Sarközy, and Szemerédi [10]. The proof in [10] actually asserts the following stronger result which guarantees not only spanning squared cycles but additionally squared cycles of all lengths between 3 and n that are divisible by 3.

Theorem 2 (Komlós, Sarközy & Szemerédi [10])**.** *There exists an integer n_0 such that for all integers $n > n_0$ any graph G of order n and minimum degree $\delta(G) \geq \frac{2}{3}n$ contains all squared cycles $C_{3\ell}^2 \subseteq G$ with $3 \leq 3\ell \leq n$. If furthermore $K_4 \subseteq G$, then $C_\ell^2 \subseteq G$ for any $3 \leq \ell \leq n$ with $\ell \neq 5$.*

For squared cycles C_ℓ^2 with ℓ not divisible by 3 the additional condition $K_4 \subseteq G$ is necessary because these target graphs are not 3-colourable and hence a complete 3-partite graph shows that one cannot hope to force C_ℓ^2 unless $\delta(G) \geq (2n+1)/3$. If $\delta(G) \geq (2n+1)/3$, on the other hand, then Turán's Theorem asserts that G contains a copy of K_4 and hence Theorem 2 implies $C_\ell^2 \subseteq G$ for any $3 \leq \ell \leq n$ with $l \neq 5$. The case $\ell = 5$ has to be excluded because C_ℓ^2 is the 5-chromatic K_5.

In this paper we address the question what happens between these two extrema of target graphs with constant order and spanning target graphs. We are interested in essentially best possible minimum degree conditions that enforce subgraphs covering a certain percentage of the host graph.

Let us start with a simple example. It is easy to see, that every graph G with minimum degree $\delta(G) \geq \delta$ for $0 \leq \delta \leq \frac{1}{2}n$ has a matching covering at least 2δ vertices (see Proposition 11). This gives a linear dependence between the forced size of a matching in the host graph and its minimum degree. The result of Corrádi and Hajnal [3] mentioned earlier is a pendant of this observation for triangle factors.

Theorem 3 (Corrádi & Hajnal [3])**.** *Let G be a graph on n vertices with minimum degree $\delta(G) = \delta \in [\frac{1}{2}n, \frac{2}{3}n]$. Then G contains $2\delta - n$ vertex disjoint triangles.*

The main theorem of this paper is a corresponding result mediating between Turán's theorem and Pósa's conjecture. More precisely, our aim is to provide exact minimum-degree thresholds for the appearance of a squared path P_ℓ^2 and a squared cycle C_ℓ^2.

There are at least two reasonable guesses one might make as to what minimum degree $\delta(G) = \delta$ will guarantee which length $\ell = \ell(n, \delta)$ of squared path (or longest squared cycle). On the one hand, the degree threshold for a *spanning* squared path or cycle and for a spanning triangle factor are approximately the same. So perhaps this remains true for smaller ℓ: in light of Theorem 3 one might expect that $\ell(n, \delta)$ were roughly $3(2\delta(G) - n)$. This turns out to be far too optimistic.

On the other hand, proofs of preceding results dealing with spanning subgraphs essentially combine greedy techniques with local changes. They simply start to construct the desired subgraph in (almost) any location, and in the event of getting stuck change only a few of the vertices embedded so far; at no time do they scrap an entire half-constructed object and start anew. It would not be unreasonable to believe that this technique also leads to best possible minimum degree conditions for large but not spanning subgraphs. Clearly, in the case of (unsquared) paths such a greedy strategy provides a path of length $\delta(G)+1$. As G might be disconnected, however, it cannot guarantee longer paths if $\delta(G) < n/2$. For squared paths the following construction shows that with an arbitrary starting location one cannot hope for squared paths on more than $\frac{3}{2}(2\delta(G)-n)$ vertices: If G contains a clique C of order $2\delta - n$ and an independent set I of order $n-\delta$ such that all vertices of C are connected to all vertices of I but no other vertices of G, then it is not difficult to see that the longest squared path in G starting in an edge of C has length $\frac{3}{2}(2\delta(G)-n)$. This could lead to the idea that $\ell(n,\delta)$ were approximately $\frac{3}{2}(2\delta(G)-n)$. It is true that there are squared paths of this length in G—but this lower bound is almost always excessively pessimistic. In other words, it turns out that one has to carefully choose the 'region' of G to look for the desired squared path. Since spanning squared paths use all vertices of G this problem does not occur for these subgraphs.

For fixed n both guesses propose a linear dependence between δ and the length $\ell(n,\delta)$ of a forced squared path (or cycle). As we will see below $\ell(n,\delta)$ as a function of δ behaves very differently: it is piece-wise linear but jumps at certain points. To make this precise we introduce the following functions. Given two positive integers n and δ with $\delta \in (\frac{1}{2}n, n-1]$, we define $r_p(n,\delta)$ to be the largest integer r such that $n-\delta+\lfloor \delta/r \rfloor > \delta$ and $r_c(n,\delta)$ to be the largest integer r such that $n-\delta+\lceil \delta/r \rceil > \delta$. We then define

$$\mathrm{sp}(n,\delta) := \min\left\{ \left\lceil \tfrac{3}{2}\lceil \delta/r_p(n,\delta)\rceil + \tfrac{1}{2} \right\rceil, \, n \right\} \tag{1}$$

and

$$\mathrm{sc}(n,\delta) := \min\left\{ \left\lfloor \tfrac{3}{2}\lceil \delta/r_c(n,\delta)\rceil \right\rfloor, \, n \right\}.$$

Observe that $\mathrm{sc}(n,\delta) \le \mathrm{sp}(n,\delta)$ and that for almost every δ, $\lim_{n\to\infty} \mathrm{sc}(n,\delta)/n = \lim_{n\to\infty} \mathrm{sp}(n,\delta)/n$. The dependence between $\mathrm{sp}(n,\delta)$ and δ is illustrated in Figure 1.

With this we are ready to formulate our main theorem, which states that $\mathrm{sp}(n,\delta)$ and $\mathrm{sc}(n,\delta)$ determine maximal lengths of squared paths and cycles, respectively, forced in an n-vertex graph G with minimum degree δ. More generally, and in accordance with Theorem 2, we show that G contains any squared cycle of length $3\ell \le \mathrm{sc}(n,\delta)$ with length divisible by 3. We shall show below that these results are tight by explicitly constructing extremal

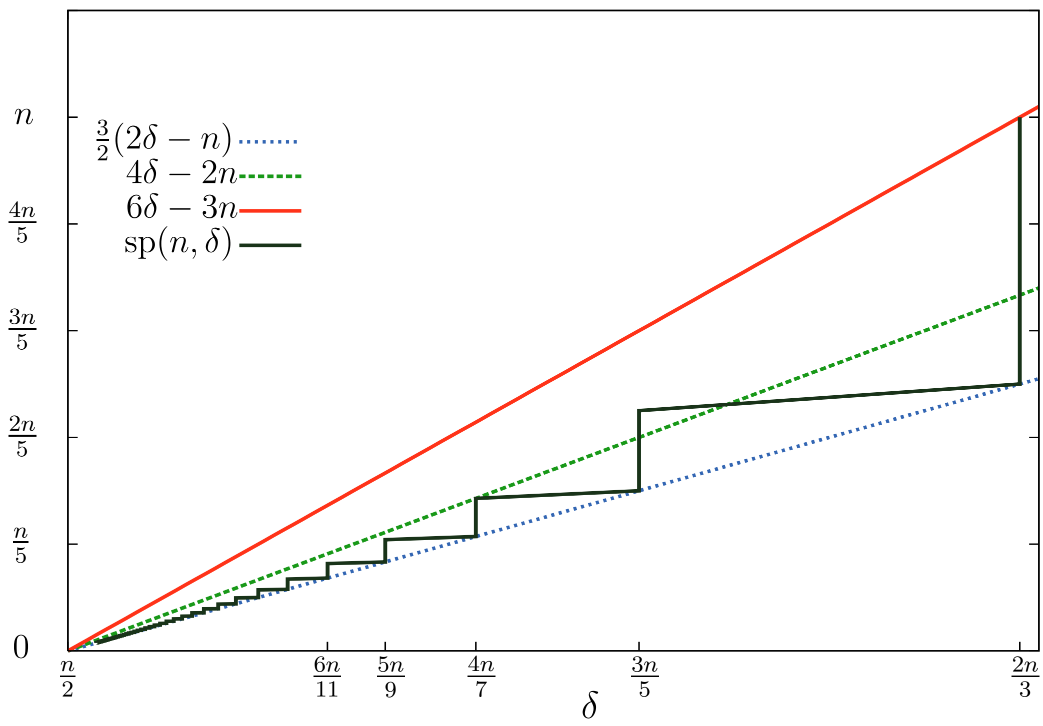

Figure 1: The behaviour of $\mathrm{sp}(n, \delta)$.

graphs $G_p(n, \delta)$ and $G_c(n, \delta)$ for squared paths and cycles. While the extremal graphs of all previously discussed results are Turán graphs (complete r-partite graphs where $r = 3$ in the case of squared paths and cycles) the graphs $G_p(n, \delta)$ and $G_c(n, \delta)$ have a rather different structure. In fact they do contain squared cycles C_ℓ^2 for all $3 \leq \ell \leq \mathrm{sc}(n, \delta)$ with $\ell \neq 5$. If any one of these 'extra' squared cycles with chromatic number 4 is not present in the host graph G, then Theorem 4 guarantees even much longer squared cycles $C_{3\ell}^2$ in G.

Theorem 4. *For any $\nu > 0$ there exists an integer n_0 such that for all integers $n > n_0$ and $\delta \in [(\frac{1}{2} + \nu)n, \frac{2}{3}n]$ the following holds for all n-vertex graphs G with minimum degree $\delta(G) \geq \delta$.*

(i) $P^2_{\mathrm{sp}(n,\delta)} \subseteq G$ and $C^2_{3\ell} \subseteq G$ for every $\ell \in \mathbb{N}$ with $\ell \leq \mathrm{sc}(n, \delta)/3$.

(ii) Either $C_\ell^2 \subseteq G$ for every $3 \leq \ell \leq \mathrm{sc}(n, \delta)$ with $\ell \neq 5$, or $C^2_{3\ell} \subseteq G$ for every $1 \leq \ell \leq 2\delta - n - \nu n$.

The proof of this result relies on Szemerédi's Regularity Lemma and is presented together with the main lemmas in Section 2. Theorem 4 cannot be extended to all values of $\delta(G)$ with $\delta(G) - \frac{1}{2}n = o(n)$ because for infinitely many values of m there are C_4-free graphs F on m vertices with $\delta(F) \geq \frac{1}{2}\sqrt{m}$ (see [16]). Indeed, let G be the n-vertex graph obtained from F by adding an independent set I on $m - \lfloor \frac{1}{2}\sqrt{m} \rfloor$ vertices and inserting all edges between F and I. It is easy to see that then $\delta(G) > \frac{1}{2}n + \frac{1}{5}\sqrt{n}$ but G does not contain a copy of C_6^2.

The following *extremal graphs* show that the bounds in (i) and (ii) of Theorem 4 are tight (see also Figure 2). For (ii) consider the complete tripartite graph $K_{n-\delta,n-\delta,2\delta-n}$. Clearly, this graph has minimum degree δ and does not contain C_ℓ^2 for any $\ell \geq 3$ not divisible by 3 or $\ell \geq 3(2\delta - n)$. For the *first part* of (i) let $G_p(n, \delta)$ be the n-vertex graph obtained from the disjoint union of an independent Y set on $n - \delta$ vertices, and $r := r_p(n, \delta)$ cliques $X_1, \dots, X_r$ with $|X_1| \leq \dots \leq |X_r| \leq |X_1| + 1$ on a total of δ vertices, by inserting all edges between Y and X_i for each $i \in [r]$. It is easy to check that $\delta(G_p(n, \delta)) = \delta$. Moreover any squared path $P_m^2 \subseteq G_p(n, \delta)$ contains vertices from at most one clique X_i. As Y is independent and P_m^2 has independence number $\lceil m/3 \rceil$ we have $\lfloor 2m/3 \rfloor \leq \lceil \delta / r_p(n, \delta) \rceil$ and thus $m \leq \lfloor \frac{1}{2}(3\lceil \delta/r_p(n, \delta) \rceil + 1) \rfloor = \mathrm{sp}(n, \delta)$. For the *second part* of (i) construct the graph $G'_c(n, \delta)$ in the same way as $G_p(n, \delta)$ but with $r := r_c(n, \delta)$ and with $|X_i| = \lceil \delta/r \rceil$ for *all* $i \in [r]$. To obtain an n-vertex graph $G_c(n, \delta)$ from $G'_c(n, \delta)$ choose v_i in X_i arbitrarily for each $i \in [r]$ and identify all v_i with $i \leq r\lceil \delta/r \rceil - \delta$. Again $G_c(n, \delta)$ has minimum degree δ, any squared cycle C_m^2 in $G_c(n, \delta)$ touches only one of the X_i, and hence $m \leq \mathrm{sc}(n, \delta)$.

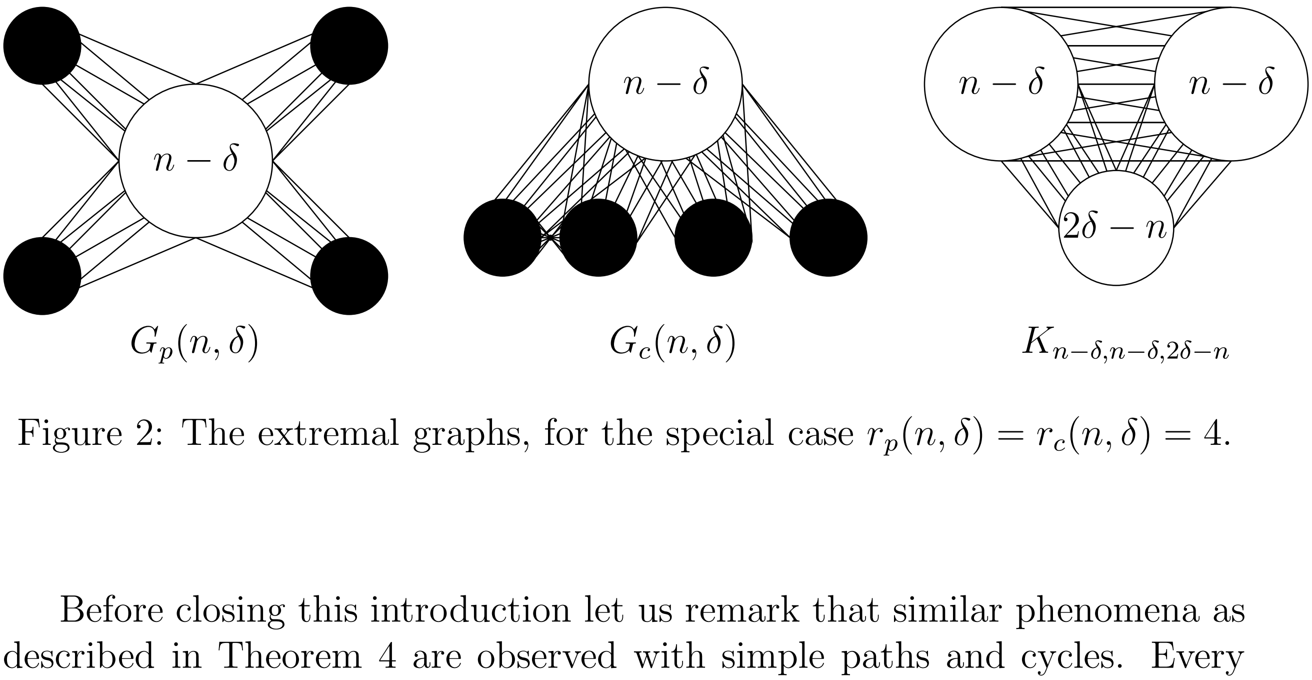

Figure 2: The extremal graphs, for the special case $r_p(n, \delta) = r_c(n, \delta) = 4$.

Before closing this introduction let us remark that similar phenomena as described in Theorem 4 are observed with simple paths and cycles. Every graph with minimum degree δ contains a path of length $\lceil n / \lfloor n/(\delta + 1) \rfloor \rceil$ and this is attained by a vertex disjoint union of cliques. This follows from an easy adjustment of the proof of Dirac's theorem. Improving on results of Nikiforov and Schelp [15] the first author proved the following theorem in [1]. The methods used for obtaining this result are quite different from those applied in this paper. In particular they do not rely on the Regularity Lemma.

Theorem 5 (Allen [1]). *Given an integer $k \geq 2$ there is n_0 such that whenever $n \geq n_0$ and G is an n-vertex graph with minimum degree $\delta \geq n/k$, the following are true.*

(*i*) *G contains C_t for every even $4 \leq t \leq \lceil n/(k-1) \rceil$,*

(*ii*) *if G does not contain a cycle of every length from $\lfloor 2n/\delta \rfloor - 1$ to $\lceil n(k-1) \rceil$ inclusive then G does contain C_t for every even $4 \leq t \leq 2\delta$.*

2 Main lemmas and proof of Theorem 4

Our proof of Theorem 4 combines the Stability Method pioneered by Simonovits [17] and the Regularity Method which pivots around the joint application of Szemerédi's celebrated Regularity Lemma [18] and the so-called Blow-up Lemma by Komlós, Sárközy and Szemerédi [11]. The combination of these two methods has proved useful for a variety of exact embedding results and was applied for example in [10]. However, this well-established technique provides only a rather loose framework for proofs of this kind. For our application we will embellish this framework with a new concept, the so-called connected triangle components of a graph.

In this section we explain how we use connected triangle components, the Regularity Method, and the Stability Method. We first provide the necessary definitions, formulate our main lemmas (whose proofs are provided in the remaining sections of this paper), and sketch how they work together in the proof of Theorem 4. The details of this proof are then presented at the end of the section.

Notation. For a graph G we write $V(G)$ and $E(G)$ to denote its vertex set and edge set, respectively, and set $v(G) = |V(G)|$ and $e(G) = |E(G)|$ and $e(X,Y) = |E(G) \cap (X \times Y)|$ for sets $X, Y \subseteq V(G)$. The graph $G[X]$ is the subgraph of G *induced* by X. The *neighbourhood* of a vertex v in G is denoted by $\Gamma(v)$ and $\Gamma(u,v)$ is the *common neighbourhood* of $u, v \in V(G)$. For an edge $uv = e \in E(G)$ we also write $\Gamma(e) = \Gamma(u,v)$. The minimum degree of G is denoted by $\delta(G)$ and for two sets $X, Y \subseteq V(G)$ we define $\delta_Y(X) = \min_{x \in X} |\Gamma(x) \cap Y|$ and $\delta_G(X) = \delta_{V(G)}(X)$.

When we write $\varepsilon \gg \varepsilon'$ for two positive real numbers ε and ε' then we mean that $\varepsilon \geq \varepsilon'$ and that we can make ε arbitrarily small by choosing ε' sufficiently small.

Connected triangle components and triangle factors. Connected triangle components and connected triangle factors are the main protagonists in the proof of Theorem 4. Roughly speaking, in a connected triangle component we can start in an arbitrary triangle and reach each other triangle by "walking" through a sequence of triangles, and a connected triangle factor is a collection of vertex disjoint triangles each pair of which is connected in this way.

To make this precise, let $G = (V, E)$ be a graph. A *triangle walk* in G is a sequence of edges $e_1, \dots, e_p$ in G such that e_i and e_{i+1} share a triangle of G for all $i \in [p-1]$. We say that e_1 and e_p are *triangle connected* in G. A *triangle component* of G is a maximal set of edges $C \subseteq E$ such that every pair of edges in C is triangle connected. Observe that this induces an equivalence relation on the edges of G, but a vertex may be part of many triangle components. In addition a triangle component does not need to form an induced subgraph of G in general. The *size* $|C|$ *of a triangle component* C is the number of vertices that are contained in some edge of C.

A *triangle factor* T in a graph G is a collection of vertex disjoint triangles in G. T is a *connected triangle factor* if all edges of T are in the same triangle component of G. The *size* of T is the number of vertices covered by T. We let $\mathrm{CTF}(G)$ denote the maximum size of a connected triangle factor in G. It is not difficult to check for example that any connected triangle factor in $G_p(n, \delta)$ contains only vertices of at most one of the cliques X_i (cf. the definition of $G_p(n, \delta)$ below Theorem 4) and of the independent set Y. Hence

$$\mathrm{CTF}\left(G_p(n,\delta)\right) = 3\left\lfloor \frac{\mathrm{sp}(n,\delta)}{3} \right\rfloor .$$

and the graph $G_p(n, \delta)$ is also extremal with respect to the size of a connected triangle factor for a given minimum degree.

We will usually find that the number of vertices in a triangle component and the size of a maximum connected triangle factor in that component are quite different. As we will explain next, for the purposes of embedding squared paths and squared cycles, however, it is the size of a connected triangle factor that is important.

The Regularity Method. The Regularity Lemma provides a partition of a dense graph that is suitable for an application of the Blow-up Lemma, which is an embedding result for large host graphs. In order to formulate the versions of these two lemmas that we will use, we first introduce some terminology.

Let $G = (V, E)$ be a graph and $\varepsilon, d \in [0, 1]$. For disjoint nonempty $U, W \subseteq V$ the *density* of the pair (U, W) is $d(U, W) = e(U, V)/|U||W|$. A pair (U, W)

with density at least d is (ε, d)*-regular* if $|d(U', W') - d(U, W)| \le \varepsilon$ for all $U' \subseteq U$ and $W' \subseteq W$ with $|U'| \ge \varepsilon|U|$ and $|W'| \ge \varepsilon|W|$. An (ε, d)*-regular partition* of G with *reduced graph* $R = (V(R), E(R))$ is a partition $V_0 \dot\cup V_1 \dot\cup \dots \dot\cup V_k$ of V with $|V_0| \le \varepsilon|V|$, $|V_i| = |V_j|$ for all $i, j \in [k]$, $V(R) = \{V_1, \dots, V_k\}$, such that (V_i, V_j) is an (ε, d)-regular pair in G whenever $V_iV_j \in E(R)$, and for each $v \in V_i$ with $i \in [k]$ there are at most $(\varepsilon + d)n$ edges incident to v that are not contained in an (ε, d)-regular pair corresponding to an edge of R (this additional requirement is not standard). In this case we also say that G has (ε, d)*-reduced graph* R and call the partition classes V_i with $i \in [k]$ *clusters* of G. Observe that our definition of the reduced graph implies that for $T \subseteq V(R)$ we can for example refer to the set $\bigcup T$, which is a subset $V(G)$.

In this paper we will use the following version of the Regularity Lemma which is an easy corollary of the so-called degree version of this lemma (see, e.g., [13, Theorem 1.10]).

Lemma 6 (Regularity Lemma). *For all $\varepsilon > 0$ and m_0 there is m_1 such that every graph G on $n > m_1$ vertices with $\delta(G) \ge \gamma n$ has an (ε, d)-reduced graph R on m vertices with $m_0 \le m \le m_1$ and $\delta(R) \ge (\gamma - d - \varepsilon)m$.*

This lemma asserts that the reduced graph R of G "inherits" the high minimum degree of G. We shall use this property in order to reduce the original problem problem of finding a squared path (or cycle) in an n-vertex graph with minimum degree γn to the problem of finding an *arbitrary* connected triangle factor of a certain size in an m-vertex graph R with minimum degree $(\gamma - d - \varepsilon)m$. The new problem is much less particular about the required subgraph than the original one and hence easier to attack (see Lemma 8).

This kind of reduction is made possible by the Blow-up Lemma. Roughly, this lemma asserts that a bounded degree graph H can be embedded into a graph G with reduced graph R if there is a homomorphism from H to a small subgraph S of R which does not "overfill" any of the clusters in S. In our setting we apply this lemma with $S = K_3$ and conclude that for *each* triangle t of a connected triangle factor T in R we find a squared path in G that almost fills the clusters of G corresponding to t. By using that T is triangle *connected* it is then possible to connect these squared paths into squared paths or cycles of the desired overall length. In addition, the Blow-up Lemma allows for some control about the start- and end-vertices of the path that is constructed in this way (cf. Lemma 7(*iii*)).

The following lemma summarises this embedding technique which is also implicit in [10]. For completeness we provide a proof of this lemma in the appendix.

Lemma 7 (Embedding Lemma). *For all $d > 0$ and $m_{\mathrm{BL}} \in \mathbb{N}$ there exist $n_{\mathrm{BL}} \in \mathbb{N}$ and $\varepsilon > 0$ such that the following hold for any graph G on $n \geq n_{\mathrm{BL}}$ vertices with (ε, d)-reduced graph R' on $m \geq m_{\mathrm{BL}}$ vertices.*

- *(i)* $C^2_{3\ell} \subseteq G$ *for every* $\ell \in \mathbb{N}$ *with* $3\ell \leq (1-d)\,\mathrm{CTF}(R')\frac{n}{m}$.
- *(ii)* *If* $K_4 \subseteq C$ *for each triangle component* C *of* R', *then* $C^2_\ell \subseteq G$ *for every* $\ell \in [3, (1-d)\,\mathrm{CTF}(R')\frac{n}{m}] \setminus \{5\}$.

Furthermore, let T be a connected triangle factor in a triangle component C of R with $K_4 \subseteq C$, let $u_1v_1, u_2v_2 \in E(G)$ be disjoint edges, and suppose that there are (not necessarily disjoint) edges $X_1Y_1, X_2Y_2 \in C$ such that the edge u_iv_i has at least $2d\frac{n}{m}$ common neighbours in each cluster X_i and Y_i for $i = 1, 2$. Then

- *(iii)* $P^2_\ell \subseteq G$ *for every* $\ell \in \mathbb{N}$ *with* $(m+2)^2 < \ell < (1-d)|T|\frac{n}{m}$, *such that* P^2_ℓ *starts in* u_1, v_1 *and ends in* u_2, v_2 *(in those orders) and at most* $(\varepsilon + d)n$ *vertices of* P^2_ℓ *are not in* $\bigcup T$.

The copies of K_4 that are required in this lemma play a crucial rôle when embedding squared cycles which are not 3-chromatic.

The Stability Method. The strategy we just described leaves us with the task of finding a big connected triangle factor T in the reduced graph R of G. However, there is one problem with this approach: The proportion τ of R covered by T is roughly equal to the proportion of G covered by the squared path P that we obtain from the Embedding Lemma (Lemma 7). However, as explained above, the relative minimum degree $\gamma_R = \delta(R)/|V(R)|$ of R is in general slightly smaller than $\gamma_G = \delta(G)/|V(G)|$, but the extremal graphs for squared paths and connected triangle factors are the same. It follows that we cannot expect that τ is larger than the proportion a maximum squared path covers in a graph with relative minimum degree γ_R, and hence smaller than the proportion we would like to cover for relative minimum degree γ_G.

Consequently we need to be more ambitious and shoot for a bigger connected triangle factor in R than we can expect for this minimum degree (cf. Lemma 8 (S1) and (S2)). This will of course not always be possible, but it will only fail if R (and hence G) is 'very close' to the extremal graph $G_p(|V(R)|, \delta(R))$ (and hence also to $G_c(|V(R)|, \delta(R))$) in which case we will say that R is *near-extremal* (cf. Lemma 8 (S3)).

This approach is called Stability Method and the following lemma states that it is feasible for our purposes. It additionally guarantees copies of K_4 as required by the Embedding Lemma. We formulate this lemma for graphs G, but use it on the reduced graph R later. Its proof does not rely on the Regularity Lemma and is given in Section 3.

Lemma 8 (Stability Lemma). *Given $\mu > 0$, for any sufficiently small $\eta > 0$ there exists n_0 such that if G has $n > n_0$ vertices and $\delta(G) = \delta \in ((\frac{1}{2} + \mu)n, \frac{2n-1}{3})$, then either*
(S1) $\mathrm{CTF}(G) \geq 3(2\delta - n)$, *or*
(S2) $\mathrm{CTF}(G) \geq \min(\mathrm{sp}(n, \delta + \eta n), \frac{11n}{20})$, *or*
(S3) *G has an independent set of size at least $n - \delta - 11\eta n$ whose removal disconnects G into components of size at most $\frac{19}{10}(2\delta - n)$.*
Moreover, in cases (S2) *and* (S3) *each triangle component of G contains a K_4.*

It remains to handle the graphs with near-extremal reduced graph. For these graphs we have a lot of structural information which enables us to show directly that they contain the squared paths and squared cycles we desire, as the following lemma documents. The proof of this lemma is provided in Section 4.

Lemma 9 (Extremal Lemma). *For every $\nu > 0$, whenever $\nu \gg \mu \gg d \gg \varepsilon > 0$ there exists N such that the following holds. Suppose G is a graph of order $n > N$ and minimum degree $\delta(G) = \delta > \frac{n}{2} + \nu n$ with (ε, d)-reduced graph R of order $m > m_{\mathrm{BL}}$. Suppose that $V(R)$ is decomposed into non-empty sets $I, B_1, B_2, \dots, B_k$, where $k \geq 2$, I is an independent set of size at least $(n - \delta - \mu n)m/n$, where each set B_i has at most $19m(2\delta - n)/(10n)$ vertices, where for any $i \neq j$ there are no edges between B_i and B_j in R, and each triangle-component of R contains a copy of K_4.*

Then G contains $P^2_{\mathrm{sp}(n,\delta)}$ and C^2_ℓ for each $\ell \in [3, \mathrm{sc}(n, \delta)] \setminus \{5\}$.

It is interesting to notice that, although the two functions $\mathrm{sp}(n, \delta)$ and $\mathrm{sc}(n, \delta)$ are different—their jumps as δ increases occur at slightly different values—they are similar enough that the Stability Lemma covers them both. We will only need to distinguish between squared paths and squared cycles when we examine the near-extremal graphs.

Proof of Theorem 4. With this we have all ingredients for the proof of our main theorem which uses the Regularity Lemma (Lemma 6) to construct a regular partition with reduced graph R of the host graph G, the Stability Lemma (Lemma 8) to conclude that R either contains a big connected triangle factor or is near-extremal, the Embedding Lemma (Lemma 7) to find long squared paths and cycles in G in the first case, and the Extremal Lemma (Lemma 9) in the second case.

Proof of Theorem 4. We require our constants to satisfy

$$\nu \gg \mu \gg \eta \gg d \gg \varepsilon > 0$$

which we choose, given ν, in that order; and we choose $n_0 > n_{\text{BL}}$ sufficiently large that we may apply Lemma 6 to any n-vertex graph, $n > n_0$, to obtain an (ε, d)-regular partition with more than m_{BL} parts (where m_{BL} and n_{BL} are as in Lemma 7).

Let $n > n_0$ and $\delta \in (n/2 + \nu n, n-1]$. Let G be any n-vertex graph with $\delta(G) \geq \delta$. We first apply Lemma 6 to G to obtain an (ε, d)-reduced graph R on $m > m_{\text{BL}}$ vertices. Let $\delta' = \delta(R) \geq (\delta/n - d - \varepsilon)m > m/2 + \mu n$. Then we apply Lemma 8 to R. There are three possibilities.

First, we could find that $\mathrm{CTF}(R) \geq 3(2\delta' - m)$. In this case by Lemma 7 we are guaranteed that for every integer ℓ with $3\ell < (1-d)\,\mathrm{CTF}(R)n/m$ we have $C^2_{3\ell} \subseteq G$. By choice of d and ε we have $3(2\delta' - m)(1-d)n/m > 3(2\delta - n - \nu n)$. Noting that $P^2_\ell \subseteq C^2_\ell$ we have $P^2_{\mathrm{sp}(n,\delta)} \subseteq G$ and $C^2_{3\ell} \subseteq G$ for each integer $\ell \leq 2\delta - n - \nu n$ as required.

Second, we could find that $\mathrm{CTF}(R) \geq \min(\mathrm{sp}(n, \delta + \eta n), \frac{11n}{20})$ and that every triangle component of R contains a copy of K_4. By Lemma 7 we are guaranteed that for every $\ell \in \{3,4\} \cup [6, (1-d)\,\mathrm{CTF}(R)n/m]$ we have $C^2_\ell \subseteq G$. By choice of d and ε we have $(1-d)\,\mathrm{CTF}(R)n/m > \mathrm{sp}(n,\delta)$, so we have $P^2_{\mathrm{sp}(n,\delta)} \subseteq G$ and for each integer $\ell \in [3, \mathrm{sc}(n,\delta)] \setminus \{5\}$ we have $C^2_{3\ell} \subseteq G$ as required.

Third, we could find that R is near-extremal: R contains an independent set on at least $m - \delta' - 11\eta n$ vertices whose removal disconnects R into components of size at most $\frac{19}{10}(2\delta' - m)$, and each triangle-component of R contains a copy of K_4. But now G satisfies the conditions of Lemma 9. It follows that G contains $P^2_{\mathrm{sp}(n,\delta)}$ and for each $\ell \in [3, \mathrm{sc}(n,\delta)] \setminus \{5\}$ the graph G contains C^2_ℓ. □

3 Triangle Components and the proof of Lemma 8

In this section we provide a proof of our stability result for connected triangle factors, Lemma 8. As indicated, our goal is to show that we either find a connected triangle factor which is at least a bit bigger than we can expect (cases (S1) and (S2) of Lemma 8) or the graph G under study has a very special, so-called extremal, configuration (case (S3)). All arguments in this proof are of combinatorial nature. Distinguishing different cases, we analyse the sizes and the structure of the triangle components in G. Before we give more details about our strategy and a sketch of the proof, we introduce some additional definitions and provide a preparatory lemma (Lemma 10).

Let G be a graph with triangle components $C_1, \ldots, C_r$. The *vertices* *of a triangle component* C_i are all vertices v such that some edge uv of G is contained in C_i. The *interior* $\mathrm{int}(G)$ of G is the set of vertices of G which are in more than one of the triangle components. For a component C_i the interior of C_i, $\mathrm{int}(C_i)$, is the the set of vertices of C_i which are in $\mathrm{int}(G)$. The remaining vertices of C_i are called the *exterior* $\partial(C_i)$. That is, $\partial(C_i)$ is formed by the set of vertices of C_i which are in no other triangle component of G. To give an example, it is easy to check that $G_p(n, \delta)$ has $r_p(n, \delta)$ triangle components; its interior is the independent set Y, with the component exteriors being the cliques $X_1, \dots, X_r$. Similarly, $G_c(n, \delta)$ has $r_c(n, \delta)$ triangle components. The following lemma collects some observations about triangle components.

Lemma 10. *Let G be an n-vertex graph with minimum degree $\delta > n/2$. Then*

(a) *each triangle component C of G satisfies $|C| > \delta$,*

(b) *for distinct triangle components C, C' we have $e(\partial(C), \partial(C')) = 0$,*

(c) *for each triangle component C, each vertex u of C, and $U := \{v : uv \in C\}$ the minimum degree in $G[U]$ is at least $2\delta - n$ and hence $|G[U]| \geq 2\delta - n + 1$.*

Proof. To see (a) let M be the vertices of a maximal clique in C (clearly $|M| \geq 3$). If u and v are in M, and x is a common neighbour of u and v, then x is also in C. Thus vertices of $G \setminus C$ are adjacent to at most 1 vertex of M and vertices of C are adjacent to at most $|M|-1$ (by maximality) vertices of M. This gives the inequality

$$|M|\delta \leq \sum_{m \in M} d(m) \leq \sum_{x \in C} (|M| - 1) + \sum_{x \notin C} 1$$

and hence $|M|\delta - n \leq (|M|-2)|C|$. Since $n < 2\delta$ we have $|C| > \delta$ as required.

The assertion (b) follows from the fact that $\Gamma(u, u') \neq \emptyset$ for $u \in \partial(C)$ and $u' \in \partial(C')$ and thus, if uu' was an edge, u and u' would be in some triangle component C'' contradicting the fact that they are in the exterior. Moreover, for an edge uv of C we have $\Gamma(u, v) \subseteq C$ as C is a triangle component. Since $|\Gamma(u, v)| \geq 2\delta - n$ we get (c). □

Let us sketch the proof of Lemma 8. Lemma 10(a) states that triangle components cannot be too small. However, it is not solely the size of the triangle components we are interested in: We want to find a triangle component that contains many vertex disjoint triangles. At this point, Lemma 10(b) comes into play. It asserts that certain spots in a triangle component induce a graph with minimum degree $2\delta - n$. In the proof of Lemma 8 we shall usually use this fact in order to find a big matching in such spots (Proposition 11 below asserts that this is possible). Clearly all edges in such a matching are triangle connected and hence it will remain to extend this matching to a set of vertex disjoint triangles. For this purpose we will analyse the size of the common neighbourhood $\Gamma(u,v)$ of an edge uv in this matching. We will usually find that $\Gamma(u,v)$ is so big that a simple greedy strategy allows us to construct the triangles. For showing that u and v have a big common neighbourhood $\Gamma(u,v)$ we will often use the following technique: We find a large set X such that neither u nor v has neighbours in X. This implies $|\Gamma(u,v)| \geq 2\delta - (n - |X|)$. Observe that Lemma 10($b$) implies that $\partial(C)$ can serve as X if both u, $v \in \partial(C')$ for some triangle components C and C'.

The strategy we just described works for most values of δ below $\frac{3}{5}n$ (we describe the exceptions below). For $\delta \geq \frac{3}{5}n$ however, the greedy type argument fails. The reason being that we usually bound the common neighbourhood of an edge used in the argument above by $4\delta - 2n$. But for $\delta \geq \frac{3}{5}n$ we might have $\mathrm{sp}(n,\delta) > 4\delta - 2n$. We solve this problem by using a different strategy in this range of δ. We will still start with a big connected matching M but use a Hall-type argument to extend the matching edges to a triangle factor T. More precisely, we find M in the exterior of some triangle component and then consider for each edge uv of M all common neighbours of uv in $\mathrm{int}(G)$. The Hall-type argument then implies that we can find distinct extensions for the edges of M. To make this argument work we use the fact that in this range of δ the set $\mathrm{int}(G)$ is an independent set.

We indicated earlier that there are some exceptional values of δ that require special treatment: values of δ around $\frac{3}{5}n$ and $\frac{4}{7}n$. Observe that in both ranges the number of triangle components of $G_p(\delta,n)$ increases (from 2 to 3 for $\frac{3}{5}n$, and from 3 to 4 for $\frac{4}{7}n$) and thus the value $\mathrm{sp}(\delta,n)$ as a function in δ jumps. Roughly speaking, the reason that these two ranges need to be treated separately is that again $\mathrm{sp}(\delta,n)$ is not substantially smaller than $4\delta - 2n$ here, but we also do not know now that $\mathrm{int}(G)$ is an independent set. For dealing with these values of δ we will use a somewhat technical case analysis which we provide at the end of this section.

As explained above we will need the following simple observation about matchings in graphs of given minimum degree.

Proposition 11. *Each graph $G = (X,E)$ with minimum degree δ has a matching covering $2\min(\delta, \lfloor |X|/2 \rfloor)$ vertices.*

Proof. Let M be a maximum matching in G and assume that M contains less than $\min(\delta, \lfloor |X|/2 \rfloor)$ edges and that there are two vertices $x, y \in X$ not covered by M. Clearly, all neighbours of x and y are covered by M and thus there is an edge uv in M with $xu, yv \in E$. But then x,u,v,y is an M-augmenting path, a contradiction. □

Before turning to the proof of Lemma 8 let us quickly collect some analytical data about the sizes $\mathrm{sp}(n,\delta)$ of connected triangle factors in $G_p(n,\delta)$ and their number of triangle components $r(n,\delta)$. It is not difficult to check that the definition of $r_p(n,\delta) =: r$ implies

$$\frac{(r+1)n-r}{2(r+1)-1} \le \delta < \frac{rn-r+1}{2r-1} \quad \text{and} \quad \frac{n-\delta}{2\delta-n+1} \le r < \frac{\delta+1}{2\delta-n+1}. \tag{2}$$

For the proof of Lemma 8 it will be useful to note in addition that for all $\delta, \delta' > \frac{n}{2} + \mu n$ with $\mu > 0$ fixed and δ' such that we have $r_p(n,\delta') \ge 3$ and either $r_p(n,\delta') \ge 5$ or $r_p(n,\delta') = r_p(n,\delta'+\eta n)$, for $\eta \le \eta_0 = \eta_0(\mu)$ sufficiently small and for $n \ge n_0 = n_0(\eta)$ sufficiently large we have, if $\mathrm{sp}(n,\delta+\eta n) \le \frac{11}{20}n$ then

$$\begin{aligned} \mathrm{sp}(n,\delta+\eta n) &\le \tfrac{3}{2}\min\left(\frac{\delta}{r_p(n,\delta+\eta n)}, \frac{\delta+3\eta n}{r_p(n,\delta+\eta n)} - 2\right), \\ \mathrm{sp}(n,\delta+\eta n) &\le 6\delta - 3n, \qquad \text{and} \qquad \mathrm{sp}(n,\delta'+\eta n) \le 4\delta' - 2n, \end{aligned} \tag{3}$$

which follows immediately from the definition of $\mathrm{sp}(n,\delta)$ in (1) (see also Figure 1).

Proof of Lemma 8. Given μ choose $\eta_0 \le \frac{1}{100}$ small enough such that (3) holds for all $\delta \ge \frac{n}{2} + \mu n$. For $\eta \le \eta_0$ let n_0 be large enough for (3) and such that $\eta n_0 \ge 2$. Define $r := r_p(n,\delta)$ and $r' := r_p(n,\delta+\eta n)$.

If G has only one triangle component then Theorem 3 guarantees that $\mathrm{CTF}(G) \ge 6\delta - 3n$ and so we are in Case (S1). Thus we may assume in the following that G has at least two triangle components. Then Lemma 10((a)) implies that $\mathrm{int}(C) \ne \emptyset$ for any triangle component C.

Suppose that C is a triangle-component of G which does not contain a copy of K_4. Let $u \in \mathrm{int}(C)$, and $U := \{v\colon uv \in C\}$. By Lemma 10 the vertex u does exist, and $\delta(G[U]) \ge 2\delta - n$. Because C contains no copy of K_4, U contains no triangle. It follows that $|U| \ge 2(2\delta - n)$, and so by Proposition 11 the set U contains a matching M with $2\delta - n$ edges. Finally we choose greedily for each $e \in M$ a vertex $v \in V(G)$ such that ev is a triangle. Since U is triangle free all these vertices must lie outside U, and since $|\Gamma(e)| \ge 2\delta - n$ we cannot fail to find distinct vertices for each edge. This yields a set T of $2\delta - n$ vertex-disjoint triangles which are all in C; so $\mathrm{CTF}(G) \ge 6\delta - 3n$ and we are in case (S1). Henceforth we assume that every triangle-component contains a copy of K_4.

We continue by considering the case $\frac{3n-2}{5} \le \delta < \frac{2n-1}{3}$ which will be treated with a Hall-type argument. The following observation readily implies the lemma in this range as we will see in Fact 2.

Fact 1. *If G has exactly 2 triangle components, $(\frac{3}{5} - 2\eta)n \leq \delta(G)$, $\mathrm{int}(G)$ is independent, and either $|\mathrm{int}(G)| < n - \delta - 11\eta n$ or the exterior X of the larger triangle component satisfies $|X| \geq \frac{19}{10}(2\delta - n)$, then* $\mathrm{CTF}(G) \geq \min(\mathrm{sp}(n, \delta + \eta n), \frac{11}{20}n)$.

To see this, note that by Lemma 10(b) a vertex $x \in X$ cannot have neighbours in the exterior of the other component and so $\Gamma(x) \subseteq X \cup \mathrm{int}(G)$ which implies $\delta(G[X]) \geq \delta - |\mathrm{int}(G)|$. By Proposition 11 there is a matching M in $G[X]$ covering $2\min(\delta - |\mathrm{int}(G)|, \lfloor |X|/2 \rfloor)$ vertices. Further, every edge of M has at least $a := 2(\delta - |X|) - |\mathrm{int}(G)|$ common neighbours in $\mathrm{int}(G)$ and every vertex $u \in \mathrm{int}(G)$ sends at least $t := \delta - (n - |X| - |\mathrm{int}(G)|)$ edges into X. Therefore u is a common neighbour of at least $b := t - |M| \geq t - \lfloor |X|/2 \rfloor + 5\eta n$ edges of M. We have $a + b \geq 3\delta - n - 2|X| + \lfloor |X|/2 \rfloor + 5\eta n \geq \lfloor |X|/2 \rfloor - 5\eta n \geq |M|$ because $(\frac{3}{5} - 2\eta)n \leq \delta$ and $|X| < n - \delta$ by Lemma 10(a). Using Hall's theorem it is easy to verify that any bipartite graph H with partition classes A and B and such that vertices in A and B have degree at least a and b, respectively, satisfies the following. If $a + b \geq \min\{|A|, |B|\}$ then H has a matching of size at least $\min\{|A|, |B|\}$. Applying this observation to the bipartite graph with $A := M$ and $B := \mathrm{int}(G)$ and edges ev for all common neighbours v of e we conclude that there is a set of vertex disjoint triangles in $\mathrm{int}(G) \cup X$ which either covers $\mathrm{int}(G)$ or includes all edges of M. This gives $\mathrm{CTF}(G) \geq 3\min(\delta - |\mathrm{int}(G)|, \lfloor |X|/2 \rfloor - 5\eta n, |\mathrm{int}(G)|) \geq 3\min(2\delta - n, \lfloor |X|/2 \rfloor - 5\eta n)$ since $|\mathrm{int}(G)| \geq 2\delta - n$ by Lemma 10(a) and $|\mathrm{int}(G)| < n - \delta$. Hence (3) implies $\mathrm{CTF}(G) \geq \min(\mathrm{sp}(n, \delta + \eta n), \frac{11}{20}n)$ unless

$$3\left(\lfloor \tfrac{|X|}{2} \rfloor - 5\eta n\right) < \min\left(\mathrm{sp}(n, \delta + \eta n), \tfrac{11}{20}n\right). \tag{4}$$

If $\mathrm{sp}(n, \delta + \eta n) < \frac{11}{20}n$ then (4) implied that $|X| - 10\eta n$ was less than the largest exterior set of $G_p(n, \delta + \eta n)$ which is of size at most $\frac{1}{2}\delta + 1$. Consequently $|X| \leq \frac{1}{2}\delta + 1 + 10\eta n < \frac{19}{10}(2\delta - n)$ because $\delta \geq (\frac{3}{5} - 2\eta)n$, and $|\mathrm{int}(G)| > |\mathrm{int}(G_p(n, \delta + \eta n))| - 10\eta n = n - \delta - 11\eta n$, a contradiction. On the other hand, $\mathrm{sp}(n, \delta + \eta n) \geq \frac{11}{20}n$ can only be the case if $\delta > (\frac{2}{3} - 2\eta)n$. But then $|X| < n - \delta < \frac{19}{10}(2\delta - n)$, and $|\mathrm{int}(G)| \geq n - 2|X|$ together with (4) is not consistent with $|\mathrm{int}(G)| < n - \delta - 11\eta n$, a contradiction.

Fact 2. *Lemma 8 is true for $\frac{3n-2}{5} \leq \delta < \frac{2n-1}{3}$.*

Observe that in this range $r = 2$. Assume G has an edge uv in $\mathrm{int}(G)$, let x be a common neighbour of u and v and C be the triangle component containing ux and vx. Then there are edges uy and vz of G outside C. The sets $\Gamma(u, y)$, $\Gamma(v, z)$ and $\{u, v, x, y, z\}$ are pairwise disjoint, and x is not adjacent to $\Gamma(u, y) \cup \Gamma(v, z) \cup \{y, z\}$. So $\delta \leq d(x) \leq (n-1) - 2(2\delta - n) - 2$ which is only possible when $\delta \leq (3n-3)/5$, a contradiction. Thus $\mathrm{int}(G)$ is an independent set, which implies $|\operatorname{int}(G)| \le n-\delta$. Hence G cannot have three triangle components by Lemma 10(a). In particular, all vertices in $\operatorname{int}(G)$ lie in both triangle components of G. So if $|\operatorname{int}(G)| \ge n-\delta-11\eta n$ then $\operatorname{int}(G)$ is the desired large independent set. If moreover all triangle component exteriors are of size $\frac{19}{10}(2\delta-n)$ at most we are in Case (S3). Otherwise (if $\operatorname{int}(G)$ is small or an exterior is large) Fact 1 gives $\operatorname{CTF}(G) \ge \min(\operatorname{sp}(n,\delta+\eta n), \frac{11}{20}n)$ which is Case (S2).

Now suppose $\delta < \frac{3n-2}{5}$ and accordingly $r \ge 3$ and $r' \ge 2$. For dealing with this case we first establish three auxiliary facts. The first one captures the greedy technique for finding a large connected triangle factor that we sketched in the beginning of this section. We will use this technique throughout the rest of the proof.

Fact 3. *If there are two sets $U_1, U_2 \subseteq V(G)$ such that no vertex in U_1 has a neighbour in U_2, all edges in $G[U_1]$ are triangle connected and $\delta(G[U_1]) \ge \delta_1$ then* $\operatorname{CTF}(G) \ge \min(3\lfloor |U_1|/2\rfloor, 3\delta_1, 2\delta-n+|U_2|)$.

By Proposition 11 we can find a matching M in U_1 covering $\min(2\lfloor |U_1|/2\rfloor, 2\delta_1)$ vertices. Now for each edge $e \in M$ in turn we pick greedily a common neighbour of e outside both M and the previously chosen common neighbours to obtain a set T of disjoint triangles. For any $x, y \in U_1$ we have $|\Gamma(x,y)| \ge 2\delta-(n-|U_2|)$. Hence T covers at least $\min(3\lfloor |U_1|/2\rfloor, 3\delta_1, 2\delta-n+|U_2|)$ vertices. Note further that T is a connected triangle factor because all edges in $G[U_1]$ are triangle connected.

Fact 4. *Let uv be an edge in* $\operatorname{int}(G)$. *Unless $r'=2$ at least one vertex, u or v, is contained in at most $r'-1$ triangle components.*

Indeed, let C_1 be the triangle component containing $uv \in \operatorname{int}(G)$ along with the (non-empty) common neighbourhood $\Gamma(u,v)$ (and perhaps some other neighbours of u or v separately). Assume that both u and v live in $r'-1$ other triangle components. Together with Lemma 10(c) this implies that $(\Gamma(u)\cup\Gamma(v))\setminus C_1$ contains in total at least $2(r'-1)$ mutually disjoint sets (since common neighbours of u and v are in C_1) of at least $2\delta-n+1$ vertices each. Let their union be U. Given any $x \in \Gamma(u,v)$, since ux and vx are both in C_1, x cannot be adjacent to any vertex of U. But then $\delta \le d(x) < n-(2r'-2)(2\delta-n+1)$ which is equivalent to $2r'-2 < (n-\delta)/(2\delta-n+1)$. By (2) the right-hand side is at most r and thus we get $2r'-2 < r$. Since $r \le r'+1$ however this is a contradiction unless $r' \le 2$.

We are ready to gather some structural information about the interior of G in the case that we do not get a triangle factor of the desired size. Our aim is to conclude that then $(*)$ $\operatorname{int}(G)$ is an independent set and that its vertices are contained in at least r' triangle components. It turns out, however, that we need to consider the cases $r = r'+1 = 2$ and $r = r'+1 = 3$ (i.e., the cases when the minimum degree δ is just a little bit below $\frac{3}{5}n$ and $\frac{4}{7}n$, respectively) separately. Unfortunately these two cases require a somewhat technical case analysis which we prefer to defer to the end of the section. We therefore state as Fact 5 $(*)$ first for these special cases.

We assume from now on, that $\mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n)$, i.e., we are not in Cases (S1) or (S2).

Fact 5. *If $r = r' + 1 = 3$ or $r = r' + 1 = 4$ then $\mathrm{int}(G)$ is an independent set all of whose vertices are contained in at least r' triangle components.*

Assuming this fact is true we can deduce $(*)$ for all values $r \geq 3$ as follows.

Fact 6. *The set $\mathrm{int}(G)$ is an independent set (and hence of size at most $n - \delta$) all of whose vertices are contained in at least r' triangle components.*

The cases $r = r' + 1 = 3$ and $r = r' + 1 = 4$ are handled by Fact 5. So we assume we are not in these cases. We will show that then each vertex of $\mathrm{int}(G)$ is contained in at least r' triangle components. Once we established this, Fact 4 implies that there are no edges in $\mathrm{int}(G)$ and so $\mathrm{int}(G)$ is an independent set as desired.

To prove that each vertex of $\mathrm{int}(G)$ is contained in at least r' triangle components we assume the contrary and show that then $\mathrm{CTF}(G) \geq \mathrm{sp}(n, \delta + \eta n)$, a contradiction. Indeed, let $w \in \mathrm{int}(G)$ and $U_1, \dots, U_k$ be the neighbours of w partitioned by the triangle-component of the edge to w, where there are $k > 1$ components containing w. By Lemma 10(c) we have $\delta(G[U_i]) \geq 2\delta - n$ and $|U_i| \geq 2\delta - n + 1$. Suppose that U_1 is the largest of the U_i's. No vertex in U_1 has a neighbour in U_2 (since the components are distinct) and all edges in $G[U_1]$ are triangle connected (because $U_1 \subseteq \Gamma(w)$). Therefore Fact 3 implies that there is a connected triangle factor T in G covering $\min(3\lfloor |U_1|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U_2|) \geq \min(3\lfloor |U_1|/2 \rfloor, 4\delta - 2n)$ vertices. If w lies only in $r'-1$ triangle components then $|U_1| \geq \delta/(r'-1)$ and therefore T covers at least $\min(3\delta/(2r'-2) - 1, 4\delta - 2n)$ vertices. This connected triangle factor is not smaller than that found in $G_p(n, \delta + \eta n)$, since (3) and the choice of η_0 and n_0 imply $\frac{3}{2}\delta/(r'-1) - 1 \geq \mathrm{sp}(n, \delta + \eta n)$ and $4\delta - 2n \geq \mathrm{sp}(n, \delta + \eta n)$ as $r \geq 3$ provided that either $r \geq 5$ or $r = r'$. This assumption is fulfilled as the cases $r = r' + 1 = 3$ and $r = r' + 1 = 4$ are excluded.

Fact 7. *We are in Case* (S3).

To conclude this we will show that $|\mathrm{int}(G)| = \alpha \geq n - \delta - 11\eta n$ and $|X_1| \leq \frac{19}{10}(2\delta - n)$ for the biggest exterior X_1 in G. Suppose for a contradiction that this is not the case. An easy calculation shows that this forces G to have exactly r' triangle components. Indeed, assume G has at least $r' + 1$ triangle components. If $\alpha < n - \delta - 11\eta n$ then each of these components C has vertices in the exterior $\partial(C)$ and so by Lemma 10(b) the minimum degree of G implies $|\partial(C)| \geq \delta - \alpha$. Accordingly $(r'+1)(\delta - \alpha) + \alpha \leq n$ which implies $(r'+1)\delta \leq (n-\alpha) + (r'+1)\alpha < n + r'(n - \delta - 11\eta n)$. Straightforward manipulation gives $\delta + \eta n < ((r'+1)n - \eta n(9r'-1))/(2(r'+1)-1)$. Since $\eta n(9r'-1) \geq 9r'-1 \geq r'$ this contradicts (2) applied to $r' = r_p(n, \delta+\eta n)$. If $|X_1| > \frac{19}{10}(2\delta - n)$ on the other hand we use Lemma 10(a) and Lemma 10(b) to get $\delta + \frac{19}{10}(2\delta - n) + (r'-1)(2\delta - n + 1) < n$. By (2) we have $r' \geq (n - \delta - \eta n)/(2\delta + 2\eta n - n + 1)$. Combined with the last inequality this gives

$$\tfrac{9}{10}(2\delta - n) + (n - \delta - \eta n)\frac{2\delta - n + 1}{2\delta - n + 1 + 2\eta n} < n - \delta$$

which is a contradiction for $\delta \geq (\frac{1}{2} + \mu)n$ (and η small enough). Hence G has exactly r' triangle components.

Now, if $r' = 2$, and accordingly $\delta \geq (\frac{3}{5} - 2\eta)n$, then Fact 1 implies $|\operatorname{int}(G)| \geq n - \delta - 11\eta n$ and $|X_1| \leq \frac{19}{10}(2\delta - n)$ because $\mathrm{CTF}(G) < \mathrm{sp}(n, \delta+\eta n)$ and so in the remainder we assume $r' > 2$. There are at most $\delta + 11\eta n$ vertices outside $\operatorname{int}(G)$. Since X_1 is the largest exterior it follows that $|X_1| \geq (\delta + 11\eta n)/r'$. To bound $\delta(G[X_1])$ from below notice that any vertex $x \in X_1$ either has δ neighbours in X_1 or a neighbour w in $\operatorname{int}(G)$. In view of the second case the fact that $\operatorname{int}(G)$ is independent and Lemma 10(c) imply that $\delta(G[X_1]) \geq \min\{2\delta - n, \delta\} = 2\delta - n$. Moreover, there are at least $r' - 1$ component exteriors other than X_1 in G, each of size at least $2\delta - n + 1$ by Lemma 10(c), and so there is a set X_2 with $|X_2| \geq (r'-1)(2\delta - n + 1)$ such that no vertex in X_1 has a neighbour in X_2. By Fact 3 we thus get $\mathrm{CTF}(G) \geq \min(3\lfloor |X_1|/2 \rfloor, 3(2\delta - n), 2\delta - n + |X_2|)$. Note that $2\delta - n + |X_2| \geq 2\delta - (n - (r'-1)(2\delta - n + 1)) = r'(2\delta - n + 1) - 1 > 6\delta - 3n$ because $r' > 2$. Further, by (3) and the choice of η_0 and n_0 we have $3\lfloor |X_1|/2 \rfloor \geq \frac{3}{2r'}(\delta + 11\eta n) - 2 \geq \mathrm{sp}(n, n + \eta n)$ and $6\delta - 3 \geq \mathrm{sp}(n, n + \eta n)$ and so $\mathrm{CTF}(G) \geq \mathrm{sp}(n, \delta + \eta n)$ which contradicts our assumption. $\square$

To complete the proof above it remains to show Fact 5. Note that we can use all facts from the proof of Lemma 8 that precede Fact 5. We will further assume that all constants and variables are set up as in this proof.

The proof of Fact 5. Recall that we assumed that $\mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n)$ in this part of the proof of Lemma 8.

We first concentrate on the case $r = 3$ and $r' = 2$. In this case $\delta(G) \in [(\frac{3}{5} - 2\eta)n, (\frac{3}{5} + \eta)n]$. Trivially each vertex of $\operatorname{int}(G)$ is contained in at least $r' = 2$ triangle components. Assume there is an edge uv in $\operatorname{int}(G)$, let x be a common neighbour of u and v, and C be the triangle component containing the triangle uvx.

Let $U_1 = \{y : uy \in C\}$ and $V_1 = \{y : vy \in C\}$. Let U_2 be the set of vertices $y \in \Gamma(u)$ such that uy is not in C, and V_2 be the set of vertices $y \in \Gamma(v)$ such that vy is not in C. Then x has no neighbour in $U_2 \dot\cup V_2$. It follows that $|U_2 \dot\cup V_2| < n - \delta$. On the other hand, by Lemma 10(c), we have $|U_2|$, $|V_2| > 2\delta - n \geq \frac{1}{5}n - 4\eta n$. This implies $|U_2| < n - \delta - |V_2| < 2n - 3\delta$ and $|V_2| < 2n - 3\delta$. Since neither u nor x have neighbours in V_2 we have

$$|\Gamma(x, u)| \geq 2\delta - (n - |V_2|) \geq \frac{2}{5}n - 8\eta n \geq n - \delta - 10\eta n.$$

Let W_u be the set of all vertices not in $U_1 \dot\cup U_2 \dot\cup V_2$. No vertex $y \in U_2$ is adjacent to any vertex in $\Gamma(x, u)$. We conclude that $|\Gamma(x, u)| < n - \delta$ and hence $|W_u| = n - |U| - |V| - |\Gamma(x, u)| \geq \frac{1}{5}n - 14\eta n$. In addition each vertex $u \in U_2$ has at most $10\eta n$ non-neighbours in $U_2 \dot\cup V_2 \dot\cup W_u$. By symmetry the same is true for vertices in V (with W_v similarly defined). But then it is easy to cover all but at most 1 vertex of U with a matching (cf. Proposition 11). This gives a matching M_u covering at least $\frac{1}{5}n - 5\eta n$ vertices of U. Because each vertex of U has at most $10\eta n$ non-neighbours in W_u we can then extend the matching edges in M_u with vertices from W_u to obtain a set T_u of (triangle connected) vertex disjoint triangles covering at least $\frac{3}{2}(\frac{1}{5}n - 100\eta n)$ vertices. Repeating the same with a matching M_v in V and vertices from W_v not used in T_u gives a connected triangle factor T_v covering $\frac{3}{2}(\frac{1}{5}n - 100\eta n)$ vertices. Since, moreover, vertices in U have at most $10\eta n$ non-neighbours in V any two vertices of U clearly have an edge of V in their common neighbourhood. Thus T_u and T_v together form a connected triangle factor covering at least $3(\frac{1}{5}n - 100\eta n)$ which is larger than $\mathrm{sp}(n, \delta + \eta n) \leq \frac{1}{2}n$, a contradiction.

Now assume that $r = 4$ and $r' = 3$. Then $\delta(G) \in [(\frac{4}{7} - 2\eta)n, (\frac{4}{7} + \eta)n]$, and consequently $\mathrm{sp}(n, \delta + \eta n) < (\frac{2}{7} + 2\eta)n$.

Assume first there is some vertex $u \in \mathrm{int}(G)$ such that u is in exactly $r' - 1 = 2$ triangle components C and C' and let U and U' be the set of neighbours of u on edges in C and C', respectively, with $|U| \geq |U'|$. Applying exactly the same strategy as in the proof of Fact 6 in Lemma 8 we obtain a triangle factor covering at least $\min(3\lfloor |U|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U'|)$ vertices. Because $|U| \geq \frac{1}{2}\delta$ we conclude from (3) that this triangle factor covers at least $(\frac{2}{7} + 2\eta)n$ vertices if $|U'| \geq (\frac{1}{7} + 6\eta)n$. Similarly, if a vertex u has three sets of neighbours U_1, U_2, U_3 on edges in three different triangle components of G then we obtain a triangle factor covering at least $\min(3\lfloor |U_1|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U_2| + |U_3|)$ vertices. This is larger than $(\frac{2}{7} + 2\eta)n$ if $|U_1| \geq (\frac{4}{21} + 2\eta)n$. Hence we can assume from now on that the following holds.

(†) If u has sets of neighbours U, U' on edges in exactly two different triangle components with $|U| \geq |U'|$ then $(\frac{1}{7} - 4\eta)n < |U'| < (\frac{1}{7} + 6\eta)n$ and

$(\frac{3}{7}-8\eta)n < |U| < (\frac{3}{7}+2\eta)n$.

(‡) If u has sets of neighbours U_1, U_2, U_3 on edges in exactly three different triangle components then $(\frac{4}{21}+2\eta)n > |U_i| > (\frac{4}{21}-6\eta)n$ for $i \in [3]$.

Now we show that $\mathrm{int}(G)$ is an independent set. Assume for a contradiction that there is an edge $uv \in \mathrm{int}(G)$. By Fact 4 one of the vertices of this edge, say u, is only in 2 triangle components; let its neighbours be U_1 and U_2 in these two triangle components, and let the neighbours of v be partitioned into sets $V_1, \dots, V_k$ according to the triangle component containing the edge to v. Assume further that $\Gamma(u,v) \subseteq U_1 \cap V_1$. Let $x \in \Gamma(u,v)$. If we had $|U_2| > (\frac{3}{7}-8\eta)n$ then since x has neighbours in neither U_2 nor V_2, and $|V_2| > (\frac{1}{7}-4\eta)n$, we would have $d(x) < (\frac{3}{7}+12\eta)n$ which is a contradiction; so by (†) we have $(\frac{1}{7}-4\eta)n < |U_2| < (\frac{1}{7}+6\eta)n$. Similarly, if $k \geq 3$ then by (‡) $|U_2 \cup V_2 \cup \cdots \cup V_k| \geq (\frac{11}{21}-8\eta)n$ and again $d(x) < (\frac{10}{21}+8\eta)n$ is a contradiction. It follows that $k=2$ and $(\frac{1}{7}-4\eta)n < |V_2| < (\frac{1}{7}+6\eta)n$.

All vertices in U_2 (resp., V_2) have at most $10\eta n$ non-neighbours outside U_1 (resp., V_1). Accordingly we obtain a situation similar to the one we established in the case $r=3$ and $r'=2$ above. We proceed similarly as there and just sketch the argument here: U_2 and V_2 induce almost complete graphs in G (i.e., for each vertex at most $10\eta n$ edges are missing). Hence we can find matchings M_u and M_v in U_2 and V_2, respectively, that almost cover these sets. Moreover almost all edges between U_2 and $W_u = V(G)-U_1-U_2-V_2$ are present and $|W_u| \geq \frac{2}{7}n - 14\eta n$ is almost twice as big as U_2. Accordingly we can use W_u to extend M_u to a collection of vertex disjoint triangles T_u. Similarly we can use $W_v = V(G)-U_2-V_1-V_2$ to extend M_v to a collection of vertex disjoint triangles T_v, avoiding T_u. Since also almost all edges between U_2 and V_2 are present $T_u \dot\cup T_v$ forms a connected triangle factor covering at least $\frac{3}{2}|U_2| + \frac{3}{2}|V_2| - 80\eta n \geq (\frac{3}{7}-100\eta)n$ vertices. With this contradiction to the size of $\mathrm{CTF}(G)$ we have established that $\mathrm{int}(G)$ is an independent set.

It remains to show that each vertex $u \in \mathrm{int}(G)$ is contained in at least 3 triangle components. Assume for a contradiction that this is not the case and u is only contained in 2 triangle components C and C' and let U and U', respectively, be the neighbours of u on edges in C and C'.

Assume that $|U| \geq |U'|$. Because $\mathrm{int}(G)$ is an independent set U and U' are contained in the exteriors of C and C'. It follows that there are no edges between U and $\partial(C')$, and exactly the same argument that we used to show (†) above now implies that $(\frac{1}{7}-4\eta n) < |\partial(C')| < (\frac{1}{7}+6\eta)n$. Since vertices in $\partial(C')$ meet only edges in C' we conclude that $|\mathrm{int}(G)| \geq (\frac{3}{7}-8\eta)n$. It follows that all but at most $8\eta n$ vertices of G are contained in $U \dot\cup U' \dot\cup \mathrm{int}(G)$. Moreover, as $\mathrm{int}(G)$ is independent, $|\mathrm{int}(G)| \leq (\frac{3}{7}+2\eta)n$ and each vertex in $\mathrm{int}(G)$ has at most $10\eta n$ non-neighbours in U. In addition, because vertices in U have no neighbours in U' they have at most $|\operatorname{int}(G)| + 8\eta n \leq (\frac{3}{7}+10\eta)n$ neighbours outside U. It follows that $\delta(G[U]) \geq (\frac{1}{7}-12\eta)n$. Thus we can proceed as before and do the following: We use Proposition 11 to find a matching M in U covering at least $(\frac{2}{7} - 24\eta)n$ vertices of U. Then we extend this matching with vertices from $\operatorname{int}(G)$ and obtain a connected triangle factor covering at least $(\frac{3}{7} - 100\eta)n$ vertices, a contradiction. □

4 Near-extremal graphs

In this section we provide the proof of Lemma 9. To prepare this proof we start with two useful lemmas. The first will be used to construct our squared paths and squared cycles from simple paths and cycles.

Lemma 12. *Given a graph G, let $T = (t_1, t_2, \dots, t_{2l})$ be a path in G and W a set of vertices disjoint from T. Let $Q_1 = (t_1, t_2)$, $Q_i = (t_{2i-3}, t_{2i-2}, t_{2i-1}, t_{2i})$ for all $1 < i \leq l$, and $Q_{l+1} = (t_{2l-1}, t_{2l})$. If there exists an ordering σ of $[l + 1]$ such that for each i, $Q_{\sigma(i)}$ has at least i common neighbours in W, then there is a squared path $(q_1, t_1, t_2, q_2, t_3, t_4, q_3, \dots)$ in G, with $q_i \in W$ for each i, using every vertex of T.*

If T is a cycle on $2l$ vertices we let instead $Q_1 = (t_{2l-1}, t_{2l}, t_1, t_2)$, $Q_i = (t_{2i-3}, t_{2i-2}, t_{2i-1}, t_{2i})$ for all $1 < i \leq l$, and σ be an ordering on $[l]$. Then, under the same conditions, we obtain a squared cycle C^2_{3l}.

Proof. We need only ensure that for each i, q_i is a common neighbour of Q_i and the q_i are distinct. This is possible by choosing for each i in succession $q_{\sigma(i)}$ to be any so far unused common neighbour of $Q_{\sigma(i)}$. □

The second, a variant on Dirac's theorem, permits us to construct paths and cycles of desired lengths which keep some 'bad' vertices far apart.

Lemma 13. *Let H be a graph on h vertices and $B \subseteq V(H)$ be of size at most $h/100$. Suppose that every vertex in B has at least $9|B|$ neighbours in H, and every vertex outside B has at least $h/2 + 9|B| + 10$ neighbours in H. Then for any given $3 \leq \ell \leq h$ we can find a cycle T_ℓ of length ℓ in H on which no four consecutive vertices contain more than one vertex of B. Furthermore, if x and y are any two vertices not in B and $5 \leq \ell \leq h$, we can find an ℓ-vertex path T_ℓ whose endvertices are x and y on which no four consecutive vertices contain more than one vertex of $B \cup \{x, y\}$.*

Proof. If we seek a path in H from x to y then we create a 'dummy edge' between x and y. If we seek a cycle, let X be any edge of $H - B$.

First we construct a path P in H covering B with the desired property. Let $B = \{b_1, b_2, \ldots, b_{|B|}\}$. For each $1 \leq i \leq |B| - 1$, choose a vertex $u_i \in H - B$ adjacent to b_i and a vertex $v_i \in H - B$ adjacent to b_{i+1}. Because both u_i and v_i have $h/2 + 9|B| + 10$ neighbours in H, they have at least $18|B| + 20$ common neighbours. At most $3|B|$ of these are either in B or amongst the chosen u_j, v_j, and so we can find a so far unused vertex w_i adjacent to u_i and v_i; since we require only $|B| - 1$ vertices $w_1, \ldots, w_{|B|-1}$ we can pick the vertices greedily.

We let v_0 be yet another vertex adjacent to b_1, and $u_{|B|}$ adjacent to $b_{|B|}$, and choose any further vertices $w_0, v_0, w_{|B|}, u_{|B|}$ such that

$$P = (x, y, u_0, w_0, v_0, b_1, u_1, w_1, v_1, b_2, \ldots, v_{|B|-1}, b_{|B|}, u_{|B|}, w_{|B|}, v_{|B|})$$

is a path on $4|B| + 5$ vertices.

Now we let P' be a path extending P in H of maximum length. We claim that P' is in fact spanning. Suppose not: let u be an end-vertex of P' and v a vertex not on P'. Since P' is maximal every neighbour of u is on P', so $v(P') > h/2 + 9|B| + 10$. If there existed an edge $u'v'$ of $P' - P$ with $u'u$ and $v'v$ edges of H, with v' closer to u on P' than u', then we would have a longer path extending P in H. Counting the edges leaving u and v yields a contradiction.

Finally we let u and v be the end-vertices of the spanning path P'. If uv is an edge of H, or if $u'v'$ is an edge of $P' - P$ (with u' nearer to u on P' than v') such that uv' and $u'v$ are edges of H, then we obtain a cycle T spanning H and containing P as a subpath. Again edge counting reveals that such an edge must exist.

To obtain a cycle T_ℓ with $h - |B| - 2 \leq \ell < h$ we take u to be an end-vertex of the path $T - P$ and v its successor on $T - P$. If we can find two further vertices u' and v' on $T - P$ (in that order from u along $T - P$) with $h - \ell$ vertices between them and with uu' and vv' edges of H then we would obtain a cycle T_ℓ of length ℓ. Again simple edge counting reveals that such a pair of vertices exists. To obtain a cycle T_ℓ with $3 \leq \ell < h - |B| - 2$ we note that $H - B$ has minimum degree $h/2 + 8|B| + 10 > (h - |B|)/2 + 1$ and thus contains a cycle of every possible length using the edge xy.

The cycle T_ℓ satisfies the condition that no four consecutive vertices contain more than one vertex of B, since either it preserves P as a subpath or it contains no vertices of B at all; similarly the path from x to y within T_ℓ satisfies the required conditions. □

Before embarking upon the proof of Lemma 9 we give an outline of the method.

We recall that the Szemerédi partition supplied to the Lemma is essentially the extremal structure: our task is to show that the underlying graph either has the same structure or possesses features which lead to longer squared paths and cycles than required for the conclusion of the Lemma. This is complicated by the fact that the Szemerédi partition is insensitive both to mis-assignment of a sublinear number of vertices and to editing of a subquadratic number of edges: we must assume, for example, that although the vertex set I in the reduced graph R is independent, the vertex set $\bigcup I$ in G may contain some vertices with very high degree into $\bigcup I$, may fail to contain some vertices of G with no neighbours in $\bigcup I$, and may contain a sublinear number of edges meeting every vertex. Fortunately, it is possible to reassign vertices in this case by separating those vertices with 'many' neighbours in $\bigcup I$ and those with 'few'. We are able to show (as Fact 8 below) that after making this reassignment, if edges remain (except possibly at one vertex)—that is, if 'few' turns out to be non-zero—then it is possible to construct very long squared paths and cycles by making use of both the sets $\bigcup B_1$ and $\bigcup B_2$ (which are 'supposed' not to be in the same triangle component).

Once we have demonstrated the existence of a large and (almost) independent set W, the minimum degree condition guarantees that almost every edge from W to the remainder of G is present. We would like to then say that in $V(G) - W$ we can find a long path, which together with vertices from W forms a squared path (and similarly for squared cycles). Unfortunately since $G[W, V(G) - W]$ is not necessarily a complete bipartite graph, this statement is not obviously true: although by definition no vertex outside W has very few neighbours in W, it is certainly possible that two vertices outside W could fail to have a common neighbour in W. But the statement is true for a path possessing sufficiently nice properties—specifically, satisfying the conditions of Lemma 12—and the purpose of Lemma 13 is to provide paths and cycles with those nice properties. The remainder of our proof, then, consists of setting up conditions for application of Lemma 13.

Proof of Lemma 9. If $\delta \geq \frac{2n-1}{3}$ then we appeal to Theorem 1 to find a spanning squared path in G; if $\delta \geq \frac{2n}{3}$ we appeal to Theorem 2 to find C_ℓ^2 for each $\ell \in \{3, 4\} \cup [6, n]$.

Let $\delta' = \delta(R) \geq (\delta/n - d - \varepsilon)m$. Observe that

$$|I| \leq m - \delta' \leq (1 - \delta/n + d + \varepsilon)m \ , \tag{5}$$

because clusters in I have δ' neighbours outside I in R. For $i \in [k]$, fix a cluster $C \in B_i$. Since $\delta' \leq \deg(C) = \deg(C, B_i \cup I) \leq \deg(C, B_i) + |I| \leq$

$\deg(C, B_i) + m - \delta'$, we conclude that

$$|B_i| \geq 2m(2\delta - n)/(3n) \; . \tag{6}$$

Fact 8. *If we have two disjoint edges of G, u_1v_1 and u_2v_2, such the edge u_iv_i has at least $\delta - (2\delta - n)/16$ common neighbours outside $\bigcup I$ for $i = 1, 2$, then G contains both $P^2_{\mathrm{sp}(n,\delta)}$ and C^2_ℓ for each $\ell \in \{3, 4\} \cup [6, \mathrm{sc}(n, \delta)]$.*

Let D be the set of clusters $C \in B_1$ such that either u_1v_1 or u_2v_2 has at most $2dn/m$ common neighbours in C. Using (5) and the hypothesis on $u_1v_1$and u_2v_2 in Fact 8 we get $|D| \leq (2\delta - n)m/(7n)$. Therefore, we conclude from (6) that $B_1 \setminus D \neq \emptyset$. Take $X \in B_1 \setminus D$ arbitrarily. We have $\deg(X, B_1) \geq \deg(X) - |I| \geq \delta' - |I| > |D|$, using (5). Thus there exists a cluster $Y \in \Gamma(X) \cap (B_1 \setminus D)$. Similarly, we can find clusters $X', Y' \in B_2$, such that $X'Y' \in E(R)$, and each of the clusters X', Y' contains at least $2dn/m$ common neighbours of both u_1v_1 and u_2v_2.

Since $\delta_R(B_1), \delta_R(B_2) \geq \delta' - |I|$, we can find greedily a matching M in $R[B_1 \cup B_2]$ with $\delta' - |I|$ edges. Since every cluster in I has at most $m - |I| - \delta'$ non-neighbours outside I, every cluster in I forms a triangle with at least $|M| - (m - |I| - \delta')$ edges of M. Since $\delta' - |I| < |I|$, we may choose greedily clusters in I to obtain a set T of at least $2\delta' - m$ vertex-disjoint triangles formed from edges of M and clusters of I. Let T_1 be the triangles of T contained in $B_1 \cup I$, and T_2 those contained in $B_2 \cup I$. Observe that all the triangles in T_1 are in the same triangle component as the edge XY, and all the triangles in T_2 are in the same triangle component as the edge $X'Y'$.

We can apply Lemma 7 with $X_1 = X_2 = X$, $Y_1 = Y_2 = Y$ to find a squared path starting with u_1v_1 and finishing with u_2v_2 using the triangles T_1. Similarly, using Lemma 7 with $X_1 = X_2 = X'$, $Y_1 = Y_2 = Y'$ we find a squared path (intersecting the first only at u_1, v_1, u_2 and v_2) starting with u_2v_2 and finishing with u_1v_1 using the triangles T_2. Concatenating the two squared paths we have a squared cycle C^2_ℓ in G, where we may choose the lengths of the squared paths such that $3(m_{\mathrm{BL}}+2)^2 \leq \ell \leq 3(1-d)(2\delta'-m)n/m$. Applying Lemma 7 to the copy of K_4 in B_1 directly we obtain C^2_ℓ for each $\ell \in \{3, 4\} \cup [6, 3n/m]$. It follows that G contains both $P^2_{\mathrm{sp}(n,\delta)}$ and C^2_ℓ for each $\ell \in \{3, 4\} \cup [6, \mathrm{sc}(n, \delta)]$ as required. This concludes the proof of Fact 8.

To simplify notation, we let $\xi = \sqrt[4]{\varepsilon + d + \mu}$.

Let W be the vertices of G which do not have more than ξn neighbours in $\bigcup I$. We infer from the fact that I is independent and from the definition of the reduced graph that $|\bigcup I - W| \leq \varepsilon n$.

Recall that $|I| \geq (n - \delta - \mu n)m/n$. Every edge in W has at least $2(\delta - \xi n) - (n - |\bigcup I|) > \delta - (2\delta - n)/16$ common neighbours outside $\bigcup I$. If there are two disjoint edges in W then we are done by Fact 8. Thus assume that no such two edges exist. It follows that there are two vertices in W which meet every edge in W; since neither has more than ξn neighbours in $\bigcup I$ there is a vertex in W adjacent to no vertex of W. Therefore, $n - \delta - \mu n - 2\varepsilon n \leq |\bigcup I| - \varepsilon n \leq |W| \leq n - \delta$.

For each $i \in [k]$ we let A_i be the set of vertices in $\bigcup B_i$ which are adjacent to at least $|W| - \xi^2 n$ vertices of W and to at least $|\bigcup B_i|/2 + 32\xi^2 n$ vertices of $\bigcup B_i$. Because at least $|W|\delta - 2|W|$ edges leave W, the average number of edges from a vertex $v \in G - W$ going to W is at least

$$\frac{|W|\delta - 2|W|}{n - |W|} \geq |W| \frac{\delta - 1}{\delta + \mu n + 2\varepsilon n} \geq |W| - \mu n - 3\varepsilon n \ .$$

In particular, at most $\xi^2 n$ vertices outside W have less than $|W| - \xi^2 n$ neighbours in W. Furthermore, if there existed εn vertices in $\bigcup B_i$ which all have at least $2dn$ neighbours in $G - W - \bigcup B_i$ then by averaging and regularity there would exist a pair of clusters, one in B_i and the other in B_j for some $j \neq i$, adjacent in R. It follows that all but at most εn vertices of $\bigcup B_i$ have at least $\delta - |W| - 2dn > |\bigcup B_i|/2 + 32\xi^2 n$ neighbours in B_i.

The vertices which are neither in W nor any of the sets A_i must be in the original bin set V_0, removed from $\bigcup I$, or removed from one of the sets $\bigcup A_i$. There are at most $\varepsilon n + \varepsilon n + \xi^2 n + k\varepsilon n < 2\xi^2 n$ such vertices.

For each i, $\delta(G[A_i]) \geq |A_i|/2 + 30\xi^2 n$, and $|A_i| \geq 2\delta - n - 3\xi^2 n$ by the above calculation.

First we note that by Lemma 13 we find in A_1 a copy of $C_{2\ell}$ for each $2\ell \in [4, |A_1|]$. By definition of A_1 we can certainly apply Lemma 12 to square this cycle (with $B = \emptyset$). This gives us squared cycles of lengths divisible by three, but not of other lengths.

If we seek a squared cycle $C^2_{3\ell+1}$ then we need to perform a 'parity correction'. Let abc be a triangle in A_1, and then apply Lemma 13 to $A_1 - \{b\}$ to find a path $P = (a, p_2, p_3, \dots, p_{2\ell-1}, c)$ on 2ℓ vertices whose end-vertices are a and c. Finally we apply Lemma 12 to P, taking as the first quadruple (b, a, p_2, p_3) and thereafter every other set of four consecutive vertices on P, finishing with $p_{2\ell-2}, p_{2\ell-1}, c, b)$. This yields a squared cycle on $3\ell + 1$ vertices as required.

If we seek a squared cycle $C^2_{3\ell+2}$ then we perform a similar process, except that we identify not one triangle in A_1 but two triangles with an edge between them.

We note that in the following arguments we can perform either parity correction process to obtain squared cycles of lengths not divisible by 3.

Now we attempt to find the longer squared cycles and the squared path we need.

For each i, we let X_i be A_i together with all vertices in $G-W$ which are adjacent to at least $30\xi^2 n$ vertices of A_i. Because every vertex in $G-W$ has at least $\delta - |W|$ neighbours outside W, every vertex in G_W is in X_i for some i.

Suppose that $|X_i \cap X_j| \geq 2$ for some $i \neq j$. In this case let v_1 and v_2 be distinct vertices of $X_i \cap X_j$. Let u_1 and u_2 be distinct neighbours in A_i of v_1 and v_2 respectively, and similarly y_1 and y_2 in A_j. Applying Lemma 13 to A_i we can find a path from u_1 to u_2 containing any number from 4 to $|A_i|-2$ of vertices we desire; we can find a similar path in A_j from y_1 to y_2. Concatenating these paths with v_1 and v_2 we can find a 2ℓ-vertex cycle $T_{2\ell}$ in $X_1 \cup X_2$ for any desired $10 \leq 2\ell \leq |A_i|+|A_j|-2$. There are no quadruples on $T_{2\ell}$ using both v_1 and v_2; the four quadruples that use one or the other each have at least $(\xi^{4/3} - 3\xi^2)n > 100k$ common neighbours on W, while all the remaining quadruples have at least $|W| - 4\xi^2 n$ common neighbours on W, so applying Lemma 12 we obtain a squared cycle on 3ℓ vertices and (choosing $2\ell > |A_i|+|A_j|-10$) a squared path on at least $\mathrm{sp}(n,\delta)$ vertices. Again it is trivially possible to perform parity corrections (prior to applying Lemma 13) so that in this case we have $C_\ell^2 \subseteq G$ for every $\ell \in \{3,4\} \cup [6, |A_i|+|A_j|-10]$.

Now suppose that for some i every vertex of A_i is adjacent to at least one vertex outside $X_i \cup W$. Since $|A_i| > 31k\xi^2 n$ we can certainly find $31\xi^2 n$ vertices all adjacent to vertices of $X_j - X_i$ for some $i \neq j$. Since no vertex of $X_j - X_i$ is adjacent to $30\xi^2 n$ vertices of A_i by definition of X_i, we find two disjoint edges $u_1 v_1$ and $u_2 v_2$ from A_i to X_j. Choosing distinct neighbours y_1 of v_1 and y_2 of v_2 in A_j and applying the identical logic to the previous case we are done.

So we may assume some vertex in A_i is adjacent only to vertices in $W \cup X_i$; thus $|X_i| \geq \delta - |W| + 1$ for each i.

We now show that we can certainly find $P^2_{\mathrm{sp}(n,\delta)}$ in G.

If for some $i \neq j$ we have $X_i \cap X_j \neq \emptyset$ then as above we obtain a squared path of the desired length.

If the sets X_i are all disjoint, then $k \leq \frac{n-|W|}{\delta-|W|+1}$. Since $|W| \leq n-\delta$ we have $k \leq r_p(n,\delta)$, and the largest of these sets, X_1, has at least $\frac{n-|W|}{r_p(n,\delta)} \geq \frac{\delta}{r_p(n,\delta)}$ vertices. Note that $\delta(G[X_1]) \geq 30\xi^2 n$ while all but at most $2\xi^2 n + k$ vertices of X_1 have at least $|A_1|/2 + 30\xi^2 n > |X_1|/2 + 25\xi^2 n$ neighbours in X_1; so we may mark the vertices of $X_1 - A_1$ as 'bad' and apply Lemma 13 to X_1 to obtain a path T covering X_1 on which every quadruple contains at most one 'bad' vertex. Finally we apply Lemma 12 to obtain a squared path on at least $\mathrm{sp}(n,\delta)$ vertices.

At last, we show that we can find in G the desired long squared cycles.

If there is a cycle of sets (relabelling the indices if necessary) $(X_1, X_2, \ldots, X_s$

for some $3 \leq s \leq k$ such that $X_i \cap X_{i+1 \mod s} = \{v_i\} \neq \emptyset$ for each i, and the v_i are all distinct, then for each i we may choose neighbours $u_i \in A_i$ and y_i in $A_{i+1 \mod s}$ of v_i, and we may insist that all $3s$ vertices are distinct. Applying Lemma 13 to each A_i in turn and concatenating the paths we can find a cycle $T_{2\ell}$ for every $4s \leq 2\ell \leq |A_1| + |A_2|$ on which there are no quadruples using more than one vertex outside $\bigcup_i A_i$. Again we may apply Lemma 12 to $T_{2\ell}$ to obtain a squared cycle on 3ℓ vertices. Finally by performing parity corrections we obtain C_ℓ^2 for every $\ell \in \{3, 4\} \cup [6, |A_1| + |A_2|]$.

If there exists no such cycle of sets, then $\sum_{i=1}^k |X_i| \leq n - |W| + k - 1$. Since we have also $|X_i| \geq \delta - |W| + 1$ for each i and $|W| \leq n - \delta$, we see that $k \leq \mathrm{sc}(n, \delta)$ and by averaging there exists a set X_1 containing at least $2\,\mathrm{sc}(n, \delta)/3$ vertices. As before, we can apply Lemma 13 to X_1 to discover a cycle $T_{2\ell}$ for each $4 \leq 2\ell \leq |X_1|$ on which the 'bad' vertices are separated, and apply Lemma 12 to it to obtain a squared cycle $C_{3\ell}^2$ for each $6 \leq 3\ell \leq \mathrm{sc}(n, \delta)$ as required. Again the parity correction procedure is applicable, so we get C_ℓ^2 for every $\ell \in \{3, 4\} \cup [6, \mathrm{sc}(n, \delta)]$. □

5 Concluding remarks

The proof of Theorem 4. Our results were most difficult to prove for $\delta \approx 4n/7$. This is somewhat surprising given the experience from the partial and perfect packing results of Komlós [9] and Kühn and Osthus [14]; in that setting it becomes steadily more difficult to prove packing results as the minimum degree of the graph (and hence the required size of a packing) increases, with perfect packings as the most difficult case. Yet in our setting it is relatively easy to prove our results when the minimum degree condition is large. This difference occurs because we have to embed triangle-connected graphs; as the minimum degree increases the possibilities for bad behaviour when forming triangle-connections are reduced. This is related to the behaviour of K_4-free graphs: if $\delta(G) > 2v(G)/3$ then G is not K_4-free; if $\delta(G) > 5v(G)/8$ then by the Andrasfai-Erdős-Sós theorem [2] G is forced to be tripartite, while for smaller values of $\delta(G)$ there exist more possibilities.

Extremal graphs. It is straightforward to check that up to some trivial modifications the graphs $G_p(n, \delta)$ and $G_c(n, \delta)$ are the only extremal graphs. However it is not the case that the only extremal graph excluding some C_ℓ^2 of chromatic number four is $K_{n-\delta,n-\delta,2\delta-n}$; when $\delta \leq 3n/5$ there are several quite different extremal graphs. We believe that the graph $G_p(n, \delta)$ remains extremal for squared paths even when δ is not bounded away from $n/2$; although as noted in Section 1 the same is not true for $G_c(n, \delta)$ and squared cycles.

Long squared cycles. In [1] a structural description of graphs avoiding non-trivial (unsquared) cycles of odd length was given. The corresponding result in our setting should be that G contains no non-trivial squared cycles of chromatic number four if it is possible to remove the vertices of an independent set from G to obtain a graph with no non-trivial odd cycles.

In addition, Theorem 5 (ii) states that if any of various odd cycles are excluded from G we are guaranteed even cycles of every length up to $2\delta(G)$, whereas the equivalent statement in our Theorem 4 contains an error term. We believe this error term can be removed, but at the cost of significantly more technical work with both the stability lemma and new extremal results.

Higher powers of paths and cycles. We note that Theorem 2 has a natural generalisation to higher powers of cycles, the so called Pósa-Seymour Conjecture; this was also proved for all sufficiently large n by Komlós, Sárközy and Szemerédi [12]. We conjecture a natural generalisation of Theorem 4 for higher powers of paths and cycles.

Given k, n and δ, we construct an n-vertex graph $G_p^{(k)}(n,\delta)$ by partitioning the vertices into an 'interior' set of $\ell=(k-1)(n-\delta)$ vertices upon which we place a complete balanced $k-1$-partite graph, and an 'exterior' set of $n-\ell$ vertices upon which we place a disjoint union of $\lfloor (n-\ell)/(\delta-\ell+1)\rfloor$ almost-equal cliques. We then join every 'interior' vertex to every 'exterior' vertex. We construct $G_c^{(k)}(n,\delta)$ similarly, permitting the cliques in the 'exterior' vertices to overlap in cut-vertices of the 'exterior' set if this reduces the size of the largest clique while preserving the minimum degree δ.

Conjecture 14. *Given $\nu>0$ and k there exists n_0 such that whenever $n\geq n_0$ and G is an n-vertex graph with $\delta(G)=\delta>\frac{k-1}{k}n+\nu n$, the following hold.*

(i) *If $P_\ell^k\subseteq G_p^{(k)}(n,\delta)$ then $P_\ell^k\subseteq G$.*

(ii) *If $C_{(k+1)\ell}^k\subseteq G_c^{(k)}(n,\delta)$ for some integer ℓ, then $C_{(k+1)\ell}^k\subseteq G$.*

(iii) *If $C_\ell^k\subseteq G_c^{(k)}(n,\delta)$ with $\chi(C_\ell^k)=k+2$ and $C_\ell^k\not\subseteq G$ for some integer ℓ, then $C_{(k+1)\ell}^k\subseteq G$ for each integer $\ell<k\delta-(k-1)n-\nu n$.*

It seems likely that again the νn error term in the last statement is not required, but again (at least for powers of cycles) it is required in the minimum degree condition.

Acknowledgement

This project was started at DocCourse 2008, organised by the research training group Methods for Discrete Structures, Berlin. In particular, we would

like to thank Mihyun Kang and Mathias Schacht for organising this nice event.

References

[1] P. Allen, *Minimum degree conditions for cycles*, (2009), submitted.

[2] B. Andrásfai, P. Erdős, and V. T. Sós, *On the connection between chromatic number, maximal clique and minimal degree of a graph*, Discrete Math. **8** (1974), 205–218.

[3] K. Corrádi and A. Hajnal, *On the maximal number of independent circuits in a graph*, Acta Math. Acad. Sci. Hungar. **14** (1963), 423–439.

[4] G. A. Dirac, *Some theorems on abstract graphs*, Proc. London Math. Soc. (3) **2** (1952), 69–81.

[5] P. Erdős and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091.

[6] P. Erdős, *Problem 9*, Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963) (M. Fiedler, ed.), Publ. House Czechoslovak Acad. Sci., Prague, 1964, p. 159.

[7] G. Fan and H. A. Kierstead, *The square of paths and cycles*, J. Combin. Theory Ser. B **63** (1995), no. 1, 55–64.

[8] ———, *Hamiltonian square-paths*, J. Combin. Theory Ser. B **67** (1996), no. 2, 167–182.

[9] J. Komlós, *Tiling Turán theorems*, Combinatorica **20** (2000), no. 2, 203–218.

[10] J. Komlós, G. N. Sárközy, and E. Szemerédi, *On the square of a Hamiltonian cycle in dense graphs*, Random Structures Algorithms **9** (1996), no. 1-2, 193–211.

[11] ———, *Blow-up lemma*, Combinatorica **17** (1997), no. 1, 109–123.

[12] ———, *Proof of the Seymour conjecture for large graphs*, Annals of Combinatorics **2** (1998), 43–60.

[13] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud., vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 295–352.

[14] D. Kühn and D. Osthus, *The minimum degree threshold for perfect graph packings*, Combinatorica **29** (2009), no. 1, 65–107.

[15] V. Nikiforov and R. H. Schelp, *Cycle lengths in graphs with large minimum degree*, Journal of Graph Theory **52** (2006), no. 2, 157–170.

[16] I. Reiman, *Über ein Problem von K. Zarankiewicz*, Acta Math. Acad. Sci. Hungar. **9** (1958), 269–273.

[17] M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, 1968, pp. 279–319.

[18] E. Szemerédi, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Orsay, 1976), Colloques Internationaux CNRS, vol. 260, CNRS, 1978, pp. 399–401.

[19] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452.

A Proof of Lemma 7

For the proof of Lemma 7 we apply the following version of the Blow-up Lemma of Komlós, Sárközy and Szemerédi [11].

Lemma 15 (Blow-up Lemma [11]). *Given fixed $c, d > 0$, for any sufficiently small $\varepsilon > 0$ the following holds. Let H be any graph with $V(H) = V_1 \dot\cup V_2 \dot\cup V_3$ and $|V_i| \geq \frac{1}{6}|V(H)|$, in which each bipartite graph $H[V_i, V_j]$ is $(2\varepsilon, d)$-regular and furthermore $\delta_{V_i}(V_j) \geq \frac{1}{2}d|V_i|$ for each $1 \leq i, j \leq 3$.*

Let F be any subgraph of the complete tripartite graph with parts V_1, V_2 and V_3 such that the maximum degree of F is at most four. Assume further, that at most four vertices x_i of F are endowed with sets $C_{x_i} \subseteq V_j$ such that $x_i \in V_j$ and $|C_x| \geq c|V_j|$

Then there is an embedding $\psi : V(F) \to V(H)$ of F into H with $\psi(x_i) \in C_{x_i}$ for $i \in [4]$.

We also say that the vertices x_i in Lemma 15 are *image restricted* to C_{X_i}.

Proof of Lemma 7. Let G be an n-vertex graph, and R' an (ε, d)-reduced graph for it on m vertices.

Fix a set $\mathcal{T}' = \left\{T'_1, \dots, T'_{\mathrm{CTF}(R')/3}\right\}$ of vertex-disjoint triangles in a triangle-component of R' covering $\mathrm{CTF}(R')$ vertices. For each triangle $T'_i = X'_{i,1}X'_{i,2}X'_{i,3}$ we may by regularity for each $j \in [3]$ remove at most $\varepsilon|X'_{i,j}|$ vertices from $X'_{i,j}$ to obtain a set $X_{i,j}$ such that each pair $(X_{i,j}, X_{i,k})$ is not only $(2\varepsilon, d)$-regular but also satisfies $\delta_{X_{i,k}}(X_{i,j}) \geq (d - 3\varepsilon)|X_{i,k}|$. We let R be the $(2\varepsilon, d)$-reduced graph corresponding to the new vertex partition given by replacing each $X'_{i,j}$ with $X_{i,j}$; then every edge of R' carries over to R, and we let $\mathcal{T}$ be the corresponding set of $\mathrm{CTF}(R')/3$ vertex disjoint triangles in R.

Fact 9. *Let $X_1, \dots, X_5$ be vertices of R (not necessarily distinct), and Z be any set of at most $\varepsilon|X_3|$ vertices of G. Suppose that X_3X_4 and X_3X_5 are edges of R. Suppose furthermore that we have two vertices $u \in X_1$ and $v \in X_2$ such that uv is an edge of G, u and v have at least $(d-\varepsilon)^2|X_3|$ common neighbours in X_3, and v has at least $(d-\varepsilon)|X_4|$ neighbours in X_4.*

Then there is a vertex $w \in X_3 - Z$ adjacent to u and v such that v and w have at least $(d-\varepsilon)^2|X_4|$ common neighbours in X_4 and w has at least $(d-\varepsilon)|X_5|$ neighbours in X_5.

Proof. Let W be the set of common neighbours of u and v in X_3. Since $X_3X_4 \in E(R)$, at most $\varepsilon|X_3|$ vertices of W have fewer than $(d-\varepsilon)|\Gamma(v) \cap X_4| \geq (d-\varepsilon)^2|X_4|$ common neighbours with v in X_4. Since $X_3X_5 \in E(R)$ at most $\varepsilon|X_3|$ vertices of W have fewer than $(d-\varepsilon)$ neighbours in X_5. Finally since $3\varepsilon|X_3| < (d-\varepsilon)^2|X_3|$ we can find a vertex of $W - Z$ satisfying the desired properties. ☐

Given a triangle walk $W = (E_1, \dots)$ in R and an orientation $\overrightarrow{U_1V_1}$ of the first edge E_1 we wish to find eventually a squared path in G following W, whose first two vertices are in U_1 and V_1, in that order. First we give a sequence of vertices of R which has the property that every vertex in the sequence is adjacent to the two preceding vertices (as is the case for a squared path).

We construct this sequence of vertices of R iteratively as follows. Let $Q_1 = (U_1, V_1)$. Now for each $2 \leq i \leq |W|$ successively, we define Q_i as follows. The last two vertices U_{i-1}, V_{i-1} of Q_{i-1} are an orientation of E_{i-1}. If $E_i = U_{i-1}V_i$ we create Q_i by appending (V_i, U_{i-1}) to Q_{i-1}; if $E_i = V_{i-1}V_i$ we append (V_i) to Q_{i-1} to create Q_i. At each step the final two vertices of Q_i are an orientation of E_i; furthermore every vertex of Q_i is adjacent in R to the two vertices preceding it in Q_i. Finally, we let $Q(W, \overrightarrow{U_1V_1}) = Q_{|W|}$.

For each $1 \leq i \leq \mathrm{CTF}(R)/3 - 1$ let W_i be a fixed triangle walk in R whose first edge is in T_i and whose last is in T_{i+1}. We suppose (repeating edges in the triangle walk W_i if necessary) that each triangle walk contains at least ten edges, and that each walk W_j with more than ten edges is of minimal length. We have $|W_i| \le \binom{m}{2}$ for each i.

We prove first that G contains $C_{3\ell}^2$ for each $3\ell \le (1-d)\,\mathrm{CTF}(R)n/m$. When $\ell \le 3(1-d)n/m$ we have $C_{3\ell}^2 \subseteq K_{(1-d)n/m,(1-d)n/m,(1-d)n/m}$ and thus by Lemma 15 we can find $C_{3\ell}^2$ as a subgraph of G (whose vertices are in T_1, with no restrictions required). Otherwise, let UV be the first edge of the triangle walk W_1. Let W' be the triangle walk obtained by concatenating $W_1, \dots, W_{\mathrm{CTF}(R)/3-1}$ and removing the last edge of W_i if it is identical to the first edge of W_{i+1}.

We choose two adjacent vertices u and v of G in U and V respectively, such that u and v have $(d-\varepsilon)^2 n/m$ common neighbours in both the third vertex of T_1 and the third vertex of $Q(W',\overrightarrow{UV})$, such that v has $(d-\varepsilon)n/m$ neighbours in the third vertex of T_1, and such that v has $(d-\varepsilon)n/m$ neighbours in the third vertex of $Q(W',\overrightarrow{UV})$ (which is possible by regularity of the various pairs). Now we apply Fact 9 with the vertices u and v and the first five vertices of $Q(W',\overrightarrow{UV})$ to obtain a third vertex v'. Now by repeatedly applying Fact 9 we construct a sequence of vertices P' (starting with u, v), where the ith vertex of P' is in $Q(W',\overrightarrow{UV})$, and the vertices are all distinct (noting that $3|W'| < \varepsilon n/m$). Thus P' is a squared path running from T_1 to $T_{\mathrm{CTF}(R)/3-1}$ following all the triangle walks W_i.

We construct similarly (and without re-using vertices) for each $1 \le i \le \mathrm{CTF}(R)/3-1$ a squared path P_i following the triangle walk W_i. However, we use the opposite orientation for the first edge: that is, instead of constructing P_1 from $Q(W_1,\overrightarrow{UV})$ we use $Q(W_1,\overrightarrow{VU})$, and similarly for each P_i we use the opposite orientation of the first edge of W_i to that used in P'. We note that the total number of vertices on all of these squared paths is not more than $6m\binom{m}{2} < \varepsilon n/m$. Finally, we remove from T_1 all vertices of $P' \cup P_1 \cup \dots \cup P_{\mathrm{CTF}(R)/3-1}$; it satisfies the conditions of Lemma 15 and thus we may embed a squared path S_1 into T_1, with the four restrictions that its first vertex is a common neighbour of the first two vertices of P', its second a neighbour of the first vertex of P', its penultimate vertex a neighbour of the first vertex of P_1 and its final vertex a common neighbour of the first two vertices of P_1 (noting that by choice of the first two vertices of P' and of P_1 the sets to which these vertices are restricted are indeed of size cn/m when $c \le d/4$). This squared path may have $3k + f_1$ vertices, where $f_1 \in \{0, 1, 2\}$ is fixed (by the restrictions of the start and end vertices) and we may choose any integer $k \in [10, (1-d)n/m]$. Similarly we may apply Lemma 15 to each T_i, $2 \le i \le \mathrm{CTF}(R)/3$, to obtain squared paths S_i whose length we may (up to the similar restrictions) choose.

Finally $S = P' \cup S_1 \cup P_1 \cup \ldots \cup P_{\mathrm{CTF}(R)/3-1} \cup S_{\mathrm{CTF}(R)/3}$ forms a squared cycle in G. It is not immediately obvious that the number of vertices of S is divisible by three—but note that

$$|Q(W, \overrightarrow{UV})| + |Q(W, \overrightarrow{VU})| \equiv 1 \mod 3$$

for any triangle walk W (with at least two edges) and first edge UV, by construction. It follows that indeed $S = C_{3k}^2$ for some integer k, and we may choose any $3k \in [6m^3, (1-d)\,\mathrm{CTF}(R)n/m]$, as required.

When every triangle-component of R contains K_4 we must also obtain squared cycles whose lengths are not divisible by three. Observe that if $ABCD$ is a copy of K_4 in R, then the vertex sequences ABC, $ABCDABC$ and $ABCDABCDABC$ each start and end with the same pair and (by use of Fact 9) can be used to construct squared paths in G which take any of the three possible lengths modulo three. We construct C_ℓ^2 for $\ell \in \{3,4\} \cup [6,20]$ within a copy of K_4 in R directly (by the above methods). To obtain C_ℓ^2 with $21 \le \ell \le 3(1-d)n/m$ we remove at most $2\varepsilon n/m$ vertices from each of A, B and C to obtain a triangle satisfying the conditions of Lemma 15, construct a short path following the appropriate vertex sequence for $\ell \mod 3$ and apply Lemma 15 to obtain C_ℓ^2. Finally, to obtain longer squared cycles we perform the same construction as above, with the exception that W' is any triangle walk to and from a copy of K_4, and so $Q(W', \overrightarrow{UV})$ may be taken (using one of the three vertex sequences above) to have any desired number of vertices modulo three (and not more than $6m^2$ in total).

Lastly, when we are required to construct a squared path between two specified edges u_1v_1 (with $2dn/m$ common neighbours in both X_1 and Y_1) and u_2v_2 (with $2dn/m$ common neighbours in both X_2 and Y_2) using triangles T in R, we apply the identical strategy, noting that the conditions on u_1v_1 and u_2v_2 are already suitable for application of Fact 9. □