FILLING THE GAP BETWEEN TURÁN'S THEOREM AND PÓSA'S CONJECTURE

PETER ALLEN*, JULIA BÖTTCHER*, AND JAN HLADKÝ†

ABSTRACT. Much of extremal graph theory has concentrated either on finding very small subgraphs of a large graph (Turán-type results) or on finding spanning subgraphs (Dirac-type results). In this paper we are interested in finding intermediate-sized subgraphs. 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 the containment of a spanning squared path and a spanning squared cycle, respectively. Our results show that such minimum degree conditions constitute not merely an interpolation between the corresponding Turán-type and Dirac-type results, but exhibit other interesting phenomena.

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 appears as a subgraph of G (which we denote by $H \subseteq G$). Turán's theorem [21] 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 (with chromatic number at least 3), 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 matching 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 supergraph 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, where the square of a graph, F^2, is obtained from F by adding edges between all pairs of vertices with distance 2 in F. 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.

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 a graph on n vertices 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, Sárközy & 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_5^2 is the 5-chromatic K_5.

In this paper we address the question of what happens between these two extrema of target graphs with constant order and order n. 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 12(a)). This gives a linear dependence between the forced size of a matching in the host graph and its minimum degree. A more general form of the result of Corrádi and Hajnal [3] mentioned earlier is a variant of this linear dependence 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 could 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 disjoint cliques C and C' of orders $2\delta - n$ and $n - \delta$, and an independent set I of order $n - \delta$ such that all vertices of C and C' are connected to all vertices of I but not to 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 piecewise linear but jumps at certain points. (These jumps can be viewed as phase transitions for the appearance of squared paths or cycles.) 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

$$
\begin{aligned}
\text{sp}(n, \delta) &:= \min \left\{ \left\lceil \tfrac{3}{2} \lceil \delta/r_p(n, \delta) \rceil + \tfrac{1}{2} \right\rceil, n \right\}, \quad \text{and} \\
\text{sc}(n, \delta) &:= \min \left\{ \left\lfloor \tfrac{3}{2} \lceil \delta/r_c(n, \delta) \rceil \right\rfloor, n \right\}.
\end{aligned}
\tag{1}
$$

Observe that $\text{sc}(n, \delta) \le \text{sp}(n, \delta)$ and that for almost every $\alpha \in (0, 1)$ we have $\lim_{n \to \infty} \text{sc}(n, \alpha n)/n = \lim_{n \to \infty} \text{sp}(n, \alpha n)/n$. The dependence between $\text{sp}(n, \delta)$ and δ is illustrated in Figure 1.

Our main theorem now states states that $\text{sp}(n, \delta)$ and $\text{sc}(n, \delta)$ are the 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 also contains any shorter squared cycle with length divisible by 3 (see (i) of Theorem 4). We shall show below that these results are tight by explicitly constructing extremal 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 $3 \le \ell \le \text{sc}(n, \delta)$ with $\ell \ne 5$. If any one of these 'extra' squared cycles with chromatic number 4 is not present in the host graph G, then (ii) of Theorem 4 guarantees even much longer squared cycles C_ℓ^2 in G, where ℓ is a multiple of 3.
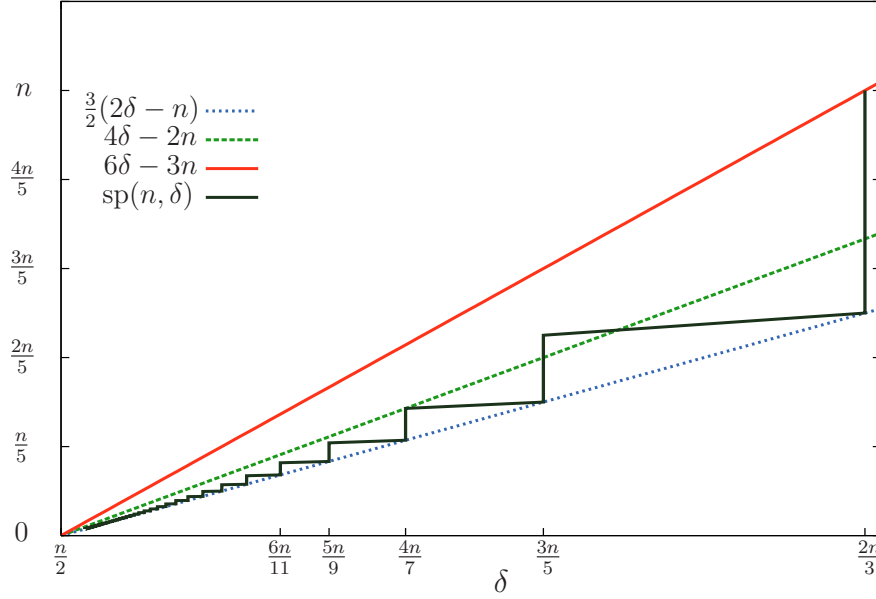
FIGURE 1. The behaviour of $\mathrm{sp}(n,\delta)$.

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 $C^2_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $3 \leq \ell \leq \mathrm{sc}(n,\delta)$ such that 3 divides ℓ.*
- (ii) *Either $C^2_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $3 \leq \ell \leq \mathrm{sc}(n,\delta)$ and $\ell \neq 5$, or $C^2_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $3 \leq \ell \leq 6\delta - 3n - \nu n$ such that 3 divides ℓ.*

The proof of this result relies on Szemerédi's Regularity Lemma[1] 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 [18]). Then, letting 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 $\delta(G) > \frac{1}{2}n + \frac{1}{5}\sqrt{n}$ but G does not contain a copy of C^2_6.

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^2_ℓ for any $\ell \geq 3$ not divisible by 3 or $\ell \geq 3(2\delta - n)$. For the *first part* of (i), let $G_p(n,\delta)$

---

[1]We refer to [14] for a survey on applications of the Regularity Lemma on graph embedding problems.

be the n-vertex graph obtained from the disjoint union of an independent set Y on $n - \delta$ vertices and $r := r_p(n, \delta)$ cliques $X_1, \ldots, X_r$ with $|X_1| \leq \cdots \leq |X_r| \leq |X_1| + 1$ 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), we 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 $m \leq \mathrm{sc}(n, \delta)$.

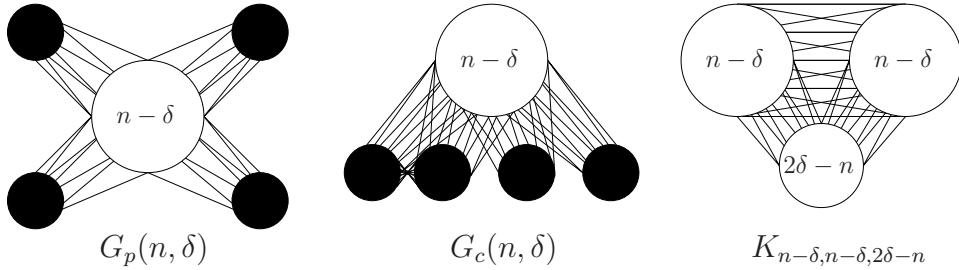$$G_p(n, \delta) \qquad\qquad G_c(n, \delta) \qquad\qquad K_{n-\delta, n-\delta, 2\delta-n}$$

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

Before closing this introduction let us remark that similar phenomena to those 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 the extremal graph is 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 [17] 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 [19], the Regularity Method which pivots around the joint application of Szemerédi's celebrated Regularity Lemma [20], and the so-called Blow-up Lemma by Komlós, Sárközy and Szemerédi [11]. The combination of these three 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, which we call the 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 this section.

Notation. For a graph G we write $V(G)$ and $E(G)$ to denote its vertex set and edge set, respectively, and set $v(G) = |V(G)|$, $e(G) = |E(G)|$ and $e(X, Y) = |\{xy \in E(G) : x \in X, y \in Y\}|$ 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 say that a statement $\mathsf{S}(\epsilon, \epsilon')$ holds for positive real numbers $\varepsilon \gg \varepsilon'$, then we mean that, given an arbitrary $\varepsilon > 0$, we can find an $\epsilon'' > 0$ such that $\mathsf{S}(\epsilon, \epsilon')$ holds for all $\epsilon' \in (0, \epsilon'']$.

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 $e_1, \ldots, e_p$ 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 *vertices of a triangle component* C_i are all vertices v such that some edge uv of G is contained in C_i. We define the *size $|C|$ of a triangle component C* to be the number of vertices 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. We define the *size* of T to be 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

$$(2) \qquad \mathrm{CTF}\left(G_p(n, \delta)\right) = 3 \left\lfloor \frac{\mathrm{sp}(n, \delta)}{3} \right\rfloor .$$

Suppose that a graph G contains a square-path of length ℓ. Then obviously, $\mathrm{CTF}(G) \geq 3\lfloor \ell/3 \rfloor$. Thus, (2) together with Theorem 4(*i*) says that $G_p(n, \delta)$ minimises CTF among all graphs of order n and minimum degree δ.

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, 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 $d(U, W) = e(U, W)/|U||W|$. A pair (U, W) is ε-*regular* if $|d(U', W') - d(U, W)| \leq \varepsilon$ for all $U' \subseteq U$ and $W' \subseteq W$ with $|U'| \geq \varepsilon|U|$ and $|W'| \geq \varepsilon|W|$. An ε-*regular partition* of G is a partition $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ of V with $|V_0| \leq \varepsilon|V|$, $|V_i| = |V_j|$ for all $i, j \in [k]$, and such that for all but at most εk^2 pairs $(i, j) \in [k]^2$, the pair (V_i, V_j) is ε-regular.

Given some $0 < d < 1$ and a pair of disjoint vertex sets (V_i, V_j) in a graph G, we say that (V_i, V_j) is (ε, d)-*regular* if it is ε-regular and has density at least d. We say that an ε-regular partition $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ of a graph G is an (ε, d)-*regular partition* if the following is true. For every $1 \leq i \leq k$, and every vertex $v \in V_i$, there are at most $(\varepsilon + d)n$ edges incident to v which are not contained in (ε, d)-regular pairs of the partition.

Given an (ε, d)-regular partition $V_0 \dot\cup V_1 \dot\cup \ldots \dot\cup V_k$ of a graph G, we define a graph R, called the *reduced graph* of the partition of G, where

$R = (V(R), E(R))$ has $V(R) = \{V_1, \ldots, V_k\}$ and $V_i V_j \in E(R)$ whenever (V_i, V_j) is an (ε, d)-regular pair. We will usually omit the partition, and simply say that G has (ε, d)-*reduced graph* R. We call the partition classes V_i with $i \in [k]$ *clusters* of G. Observe that our definition of the reduced graph R implies that for $T \subseteq V(R)$ we can for example refer to the set $\bigcup T$, which is a subset of $V(G)$.

The celebrated Szemerédi Regularity Lemma [20] states that every large graph has an ε-regular partition with a bounded number of parts. Here we state the so-called degree form of this lemma (see, e.g., [13, Theorem 1.10]).

Lemma 6 (Regularity Lemma, degree form)**.** *For every $\varepsilon > 0$ and every integer m_0, there is m_1 such that for every $d \in [0, 1]$ every graph $G = (V, E)$ on $n \geq m_1$ vertices has an (ε, d)-reduced graph R on m vertices with $m_0 \leq m \leq m_1$.*

For our purpose it is convenient to work with even a different version of the regularity lemma, which takes into account that we are dealing with graphs of high minimum degree. This lemma is an easy corollary of Lemma 6. A proof can be found, e.g., in [16, Proposition 9].

Lemma 7 (Regularity Lemma, minimum degree form)**.** *For all ε, d, γ with $0 < \varepsilon < d < \gamma < 1$ and for every m_0, there is m_1 such that every graph G on $n > m_1$ vertices with $\delta(G) \geq \gamma n$ has an (ε, d)-reduced graph R on m vertices with $m_0 \leq m \leq m_1$ and $\delta(R) \geq (\gamma - d - \varepsilon)m$.*

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 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 9).

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 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 the fact 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 8(*iii*)).

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

Lemma 8 (Embedding Lemma). *For all $d > 0$ there exists $\varepsilon_{\mathrm{EL}} > 0$ with the following property. Given $0 < \varepsilon < \varepsilon_{\mathrm{EL}}$, for every $m_{\mathrm{EL}} \in \mathbb{N}$ there exists $n_{\mathrm{EL}} \in \mathbb{N}$ such that the following hold for any graph G on $n \geq n_{\mathrm{EL}}$ vertices with (ε, d)-reduced graph R' on $m \leq m_{\mathrm{EL}}$ vertices.*

(i) $C_{3\ell}^2 \subseteq G$ *for every $\ell \in \mathbb{N}$ with $3\ell \leq (1 - d) \operatorname{CTF}(R') \frac{n}{m}$.*

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

Furthermore, let T be a connected triangle factor in a triangle component C of R with $K_4 \subseteq C$, let $u_1 v_1, u_2 v_2 \in E(G)$ be disjoint edges, and suppose that there are (not necessarily disjoint) edges $X_1 Y_1, X_2 Y_2 \in C$ such that the edge $u_i v_i$ has at least $2d\frac{n}{m}$ common neighbours in each cluster X_i and Y_i for $i = 1, 2$. Then

(iii) $P_\ell^2 \subseteq G$ *for every $\ell \in \mathbb{N}$ with $6(m + 2)^3 < \ell < (1 - d)|T|\frac{n}{m}$, such that P_ℓ^2 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_ℓ^2 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 8). 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 9 (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 9 (S3)).

This approach is called the Stability Method and the following lemma states that it is feasible for our purposes. This lemma 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 9 (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, each 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.*

By the discussion above, 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. In this proof we shall again make use of the embedding lemma, Lemma 8. Accordingly Lemma 10 inherits the upper bound m_{EL} on the number of clusters from Lemma 8.

Lemma 10 (Extremal Lemma). *For every $\nu > 0$, given $0 < \eta, d < 10^{-8}\nu^4$ there exists $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ and every m_{EL}, there exists N such that the following holds. Suppose that*

(i) *G is an n-vertex graph with $n > N$ and $\delta(G) = \delta > \frac{n}{2} + \nu n$,*

(ii) *R is an (ε, d)-reduced graph of G of order $m \leq m_{\mathrm{EL}}$,*

(iii) *each triangle component of R contains a copy of K_4.*

(iv) *$V(R) = I \dot\cup B_1 \dot\cup B_2 \dot\cup \cdots \dot\cup B_k$ with $k \geq 2$,*

(v) *I is an independent set with $|I| \geq (n - \delta - 11\eta n)m/n$,*

(vi) *for each $i \in [k]$ we have $0 < |B_i| \leq 19m(2\delta - n)/(10n)$, and for every $j \in [k] \setminus \{i\}$ there are no edges between B_i and B_j in R.*

Then G contains $P_{\mathrm{sp}(n,\delta)}^2$ and C_ℓ^2 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 the ingredients for the proof of our main theorem, which uses the Regularity Lemma (in form of Lemma 7) to construct a regular partition with reduced graph R of the host graph G, the Stability Lemma (Lemma 9) to conclude that R either contains a big connected triangle factor or is near-extremal, the Embedding Lemma (Lemma 8) to find long squared paths and cycles

in G in the first case, and the Extremal Lemma (Lemma 10) 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, as follows. First, we choose $\mu := \nu/2$. We then choose $\eta > 0$ to be small enough for both Lemma 9 and Lemma 10. Now we set $d > 0$ to be small enough for Lemma 10 and such that $d \leq \nu/10$ and $d \leq \eta/10$. For this d Lemma 8 then produces a constant $\varepsilon_{\mathrm{EL}}$. We choose $\varepsilon > 0$ to be smaller than $\min\{\varepsilon_{\mathrm{EL}}, \nu/10\}$ and sufficiently small for Lemma 10. We choose m_0 to be sufficiently large to apply Lemma 9 to any graph with at least m_0 vertices. We then choose m_{EL} such that Lemma 7 guarantees the existence of an (ε, d)-regular partition with at least m_0 and at most m_{EL} parts. Finally we choose $n_0 > n_{\mathrm{EL}}$ to be sufficiently large for both Lemma 8 and Lemma 10.

Let $n > n_0$ and $\delta \in (n/2 + \nu n, n-1]$. Let G be any n-vertex graph with $\delta(G) \geq \delta$. Observe that it suffices to show that $P^2_{\mathrm{sp}(n,\delta)} \subseteq G$ and that (ii) of Theorem 4 holds. We first apply Lemma 7 to G to obtain an (ε, d)-reduced graph R on $m_0 \leq m \leq m_{\mathrm{EL}}$ vertices. Let $\delta' = \delta(R) \geq (\delta/n - d - \varepsilon)m > m/2 + \mu m$. Then we apply Lemma 9 to R. There are three possibilities.

First, we could find that $\mathrm{CTF}(R) \geq 3(2\delta' - m)$. In this case by Lemma 8 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 $(1 - d) \cdot 3(2\delta' - m)n/m > 6\delta - 3n - \nu n$. Noting that $P^2_\ell \subseteq C^2_\ell$ we conclude that $P^2_{\mathrm{sp}(n,\delta)} \subseteq G$ and $C^2_\ell \subseteq G$ for each integer $\ell \leq 6\delta - 3n - \nu n$ such that 3 divides ℓ, i.e., the second case of Theorem 4(ii) holds.

Second, we could find that $\mathrm{CTF}(R) \geq \min(\mathrm{sp}(m, \delta' + \eta m), \frac{11m}{20})$ and that every triangle component of R contains a copy of K_4. By Lemma 8 we are guaranteed that for every $\ell \in [6, (1 - d)\,\mathrm{CTF}(R)n/m] \setminus \{5\}$ we have $C^2_\ell \subseteq G$. By choice of η and d we have $(1 - d)\,\mathrm{CTF}(R)n/m > \mathrm{sp}(n,\delta) \geq \mathrm{sc}(n,\delta)$, 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 $C^2_\ell \subseteq G$, i.e., the first case of Theorem 4(ii) holds.

Third, we could find that R is near-extremal. Then R contains an independent set on at least $m - \delta' - 11\eta m$ 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 10. 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, i.e., the first case of Theorem 4(ii) holds. $\qquad\square$

3. Triangle Components and the proof of Lemma 9

In this section we provide a proof of our stability result for connected triangle factors, Lemma 9. Distinguishing different cases, we analyse the sizes and the structure of the triangle components in the graph G under study. 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 11).

Let G be a graph with triangle components $C_1, \ldots, C_r$. 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, written $\mathrm{int}(C_i)$, is 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, by definition the graph $G_p(n, \delta)$ has $r_p(n, \delta)$ triangle components; its interior is the independent set Y (using the notation of the construction of $G_p(n, \delta)$ on page 6 in Section 1), with the component exteriors being the cliques $X_1, \ldots, X_r$.

The following lemma collects some observations about triangle components.

Lemma 11. *Let G be an n-vertex graph with $\delta(G) = \delta > n/2$. 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 $U := \{v \colon uv \in C\}$, 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 $|M| - 1$ vertices of M, by maximality 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.

Since $\delta > n/2$, we have that $\Gamma(u, u') \neq \emptyset$ for any two vertices u and u'. Now, if $u \in \partial(C)$, $u' \in \partial(C')$, $x \in \Gamma(u, u')$, and uu' was an edge, then $uu'x$ would form a triangle. Then u and u' would be together in some triangle component C'', contradicting the fact that they are in the exterior. Therefore, the assertion (b) follows.

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).   □

Now let us sketch the proof of Lemma 9. Lemma 11(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 11(c) 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 9 we shall usually (i.e., for many values of δ) use this fact in order to find a big matching M in such spots (Proposition 12(a) below asserts that this is possible). Clearly all edges in such a matching are triangle connected and hence it will remain to extend M 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 M. We will usually find that $\Gamma(u, v)$ is so big that a simple greedy strategy allows us to construct the triangles. For estimating $\Gamma(u, v)$ 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 11($b$) 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 $\mathrm{sp}(n, \delta) > 4\delta - 2n$ (see Figure 1). We solve this problem by using a different strategy in this range of δ. We will still start with a big connected matching M as before, but use a Hall-type argument to extend M 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 permits us to 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: namely δ close to $\frac{3}{5}n$ and $\frac{4}{7}n$. Observe that in both ranges the number of triangle components of $G_p(n, \delta)$ changes (from 2 to 3 for $\frac{3}{5}n$, and from 3 to 4 for $\frac{4}{7}n$) and thus the value $\mathrm{sp}(n, \delta)$ as a function in δ jumps (see Figure 1). Roughly speaking, the reason that these two ranges need to be treated separately is that again $\mathrm{sp}(n, \delta)$ 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 proof of Fact 17).

As explained above, we will apply the following simple observations about matchings in graphs of given minimum degree.

Proposition 12.

(a) Let $G = (X, E)$ be a graph with minimum degree δ. Then G has a matching covering $2\min(\delta, \lfloor |X|/2 \rfloor)$ vertices.

(b) Let $G = (A \dot\cup B, E)$ be a bipartite graph with parts A and B, such that every vertex in A has degree at least a and every vertex in B has degree at least b. Then G has a matching covering $2\min(a + b, |A|, |B|)$ vertices.

Proof. We first prove (a). Let M be a maximum matching in G, and assume that M contains less than $\min(\delta, \lfloor |X|/2 \rfloor)$ edges. In particular, there are two vertices $x, y \in X$ not covered by M. Clearly, all neighbours of x and y are covered by M.

We claim that there is an edge uv in M with $xu, yv \in E$. Indeed, suppose that this is not the case. Then $|e \cap \Gamma(x)| + |e \cap \Gamma(y)| \le 2$ for each $e \in M$. We therefore have

$$\delta + \delta \le |\Gamma(x)| + |\Gamma(y)| = \sum_{e \in M}(|e \cap \Gamma(x)| + |e \cap \Gamma(y)|) \le 2|M|,$$

contradicting the fact that $\delta > |M|$.

Now, let $uv \in M$ be an edge as in the claim above. Since $xu, yu \in E$ we get that x, u, v, y is an M-augmenting path, a contradiction.

Next we prove (b). Let M be a maximum matching in G. We are done unless there are vertices $u \in A$ and $v \in B$ not contained in M. There cannot be an edge $xy \in M$ such that uy and xv are edges of G by maximality of M, since then u, y, x, v was an M-augmenting path. But u has at least a neighbours in $V(M) \cap B$, and v at least b neighbours in $V(M) \cap A$, so there must be at least $a + b$ edges in M. $\square$

Before turning to the proof of Lemma 9 let us quickly collect some analytical data about $\mathrm{sp}(n, \delta)$ and $r_p(n, \delta) =: r$. It is not difficult to check that

(3)
$$\frac{(r+1)n - r}{2(r+1) - 1} \le \delta < \frac{rn - r + 1}{2r - 1} \quad \text{and}$$
$$\frac{n - \delta}{2\delta - n + 1} \le r < \frac{\delta + 1}{2\delta - n + 1}.$$

For the proof of Lemma 9 it will be useful to note in addition that given $\mu > 0$, for every $0 < \eta < \eta_0 = \eta_0(\mu)$, there is $n_1 = n_1(\eta)$ such that the following holds for all $n \ge n_1$. For all $\delta, \delta' > \frac{n}{2} + \mu n$, where δ is such that $\mathrm{sp}(n, \delta + \eta n) \le \frac{11}{20}n$, and where δ' is such that we have $r_p(n, \delta') \ge 3$ and either $r_p(n, \delta') \ge 5$ or $r_p(n, \delta') = r_p(n, \delta' + \eta n)$, we have

(4) $\mathrm{sp}(n, \delta + \eta n) \le \dfrac{3}{2} \min\left(\dfrac{\delta}{r_p(n, \delta + \eta n) - 1} - 2, \dfrac{\delta + 3\eta n}{r_p(n, \delta + \eta n)} - 2\right),$

(5)
$\mathrm{sp}(n, \delta + \eta n) \le \frac{19}{20} \cdot 3(2\delta - n) - 2 \le 6\delta - 3n - 100\eta n, \quad \text{and}$
$\mathrm{sp}(n, \delta' + \eta n) \le 4\delta' - 2n,$

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

Proof of Lemma 9. Given μ and any $0 < \eta < \min(\frac{1}{1000}, \eta_0(\mu), 2\mu^2/3)$, where $\eta_0(\mu)$ is as above (4), let $n_0 := \max(n_1(\eta), 2/\eta)$ with $n_1(\eta)$ as above (4). Let $n \geq n_0$. This in particular means that we may assume the inequalities (4) and (5) in what follows. Define $\gamma := \delta/n$, and $r := r_p(n, \delta)$ and $r' := r_p(n, \delta + \eta n)$.

If G has only one triangle component then Theorem 3 guarantees that $\mathrm{CTF}(G) \geq 6\delta - 3n$ and so we are in Case (S1). Thus we may assume in the following that G has at least two triangle components. Then Lemma 11(*a*) implies that $\mathrm{int}(C) \neq \emptyset$ 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 be a vertex of C, and $U := \{v \colon uv \in C\}$. By Lemma 11(*c*) we have $\delta(G[U]) \geq 2\delta - n$. Because C contains no copy of K_4, U contains no triangle. By Turán's theorem we have $|U| \geq 2(2\delta - n)$, and so by Proposition 12(*a*) the set U contains a matching M with $2\delta - n$ edges. Finally we choose greedily for each $e \in M$ a distinct 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 $|\Gamma(e)| \geq 2\delta - n$ 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 $\mathrm{CTF}(G) \geq 6\delta - 3n$ and we are in case (S1). Henceforth we assume that every triangle component of G contains a copy of K_4.

We continue by considering the case $\frac{3n-2}{5} \leq \delta < \frac{2n-1}{3}$. The following observation readily implies the lemma in this range, as we will see in Fact 14.

Fact 13. *If $\delta(G) \geq (\frac{3}{5} - 2\eta)n$, G has exactly 2 triangle components, $\mathrm{int}(G)$ is independent, and either $|\mathrm{int}(G)| < n - \delta - 11\eta n$ or the exterior X of the triangle component with most vertices satisfies $|X| \geq \frac{19}{10}(2\delta - n)$, then $\mathrm{CTF}(G) \geq \min(\mathrm{sp}(n, \delta + \eta n), \frac{11}{20}n)$.*

Proof of Fact 13. First, by Lemma 11(*b*) a vertex $x \in X$ cannot have neighbours in the exterior of the other triangle component, so $\Gamma(x) \subseteq X \cup \mathrm{int}(G)$. Thus $\delta(G[X]) \geq \delta - |\mathrm{int}(G)|$, which by Proposition 12(*a*) means that there is a matching M in $G[X]$ with

$$(6) \qquad |M| = \min(\delta - |\mathrm{int}(G)|, \lfloor |X|/2 \rfloor)$$

edges.

We aim to pair off edges of M with vertices of $\mathrm{int}(G)$ to form a sufficiently large number of vertex-disjoint triangles. To see that a triangle factor resulting from this process will be connected, observe that all edges of M are in X, and since X is a triangle component exterior, the edges of M are triangle connected. To form triangles from edges of M and vertices of $\mathrm{int}(G)$, we introduce an auxiliary bipartite

graph H with vertex set $M \dot\cup \text{int}(G)$, where $uv \in M$ is adjacent in H to $w \in \text{int}(G)$ iff uvw is a triangle of G. Every vertex of X has at least $\delta - |X|$ neighbours in $\text{int}(G)$, and so every edge of M has at least $a := 2(\delta - |X|) - |\text{int}(G)|$ common neighbours in $\text{int}(G)$. At the same time, since $\text{int}(G)$ is independent, every vertex of $\text{int}(G)$ has at least $\delta - (n - |\text{int}(G)| - |X|)$ neighbours in X, of which all but $|X| - 2|M|$ must be in M. So every vertex of $\text{int}(G)$ must have at least

$$b := \delta - (n - |\text{int}(G)| - |X|) - (|X| - 2|M|) - |M| = \delta - n + |\text{int}(G)| + |M|$$

edges of M in its neighbourhood. By Proposition $12(b)$ there is a matching in H on at least $\min(a+b, |M|, |\text{int}(G)|)$ edges, and hence a connected triangle factor in G with so many triangles. Observe that

$$\text{(7)} \qquad \begin{aligned} a + b &= 2\delta - 2|X| - |\text{int}(G)| + \delta - n + |\text{int}(G)| + |M| \\ &= 3\delta - n - 2|X| + |M| \,. \end{aligned}$$

Since there are two triangle components in G, there is a vertex u in a triangle component exterior which is not X. Therefore u has no neighbour in X, so $|X| < n - \delta$. Since $\delta \geq (\frac{3}{5} - 2\eta)n$, by (7) we have

$$\text{(8)} \qquad\qquad a + b > |M| - 10\eta n \,.$$

Furthermore,

$$\text{(9)} \qquad\qquad \text{if} \quad |X| \leq (\tfrac{2}{5} - 3\eta)n \,, \quad \text{then} \quad a + b \geq |M| \,.$$

By Lemma $11(a)$ we have $|\text{int}(G)| \geq 2\delta - n \geq \frac{n}{5} - 4\eta n$. Since $\eta \leq \frac{1}{1000}$ we have

$$3|\text{int}(G)| \geq \frac{3n}{5} - 12\eta n > \frac{11n}{20} \,.$$

Thus we have $\text{CTF}(G) \geq \frac{11n}{20}$ if we find a matching in H covering $\text{int}(G)$. It remains, then, to check that we have

$$\text{(10)} \qquad 3\min(a+b, |M|) \geq \min(\text{sp}(n, \delta + \eta n), \frac{11}{20}n).$$

We distinguish two cases.

Case 1: $a+b < |M|$. By (9) this forces $|X| > (\frac{2}{5} - 3\eta)n$. Since we have $|M| = \min(\delta - |\text{int}(G)|, \lfloor |X|/2 \rfloor)$ by (6), there are two possibilities. If $|M| = \lfloor |X|/2 \rfloor$ then we have

$$a + b \overset{(8)}{\geq} \left\lfloor \frac{|X|}{2} \right\rfloor - 10\eta n > \frac{n}{5} - 12\eta n > \frac{11n}{60} \,,$$

which proves (10) in this subcase. If, on the other hand, $|M| = \delta - |\text{int}(G)|$, then we use that $\text{int}(G)$ is independent, which implies $\text{int}(G) \leq n - \delta$ and thus

$$a + b \overset{(8)}{\geq} |M| - 10\eta n = \delta - |\text{int}(G)| - 10\eta n \geq 2\delta - n - 10\eta n$$

$$\overset{(5)}{\geq} \tfrac{1}{3}\,\text{sp}(n, \delta + \eta n) \,,$$

which proves (10) in this subcase.

Case 2: $a+b \geq |M|$. In this case, H contains a matching of size $|M|$, so we have $\mathrm{CTF}(G) \geq 3|M| = 3\min(\delta - |\mathrm{int}(G)|, \lfloor |X|/2 \rfloor)$. Again there are two possibilities, depending on $|M|$. If $|M| = \delta - |\mathrm{int}(G)|$, we are done by (5) exactly as before. If, on the other hand, $|M| = \lfloor |X|/2 \rfloor$, then (10) holds (and hence we are done) unless

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

We now assume (11) in order to derive a contradiction, and make a final subcase distinction.

First assume that $\mathrm{sp}(n, \delta + \eta n) < \tfrac{11}{20}n$. Then $r' \geq 2$ and hence (11) and (1) imply

$$|X| < \tfrac{1}{2}(\delta + \eta n) + 3 < \tfrac{51}{100}\delta < \tfrac{19}{10}(2\delta - n),$$

because $\delta \geq (\tfrac{3}{5} - 2\eta)n$ and $\eta \leq \tfrac{1}{1000}$. Furthermore, since G has two triangle components whose exterior is of size at most X by assumption we have $|\mathrm{int}(G)| > n - 2|X| = n - \delta - \eta n - 6$, a contradiction to the the conditions of Fact 13.

Now assume that $\mathrm{sp}(n, \delta + \eta n) \geq \tfrac{11}{20}n$. Then we have $\delta > (\tfrac{2}{3} - 2\eta)n$. By Lemma 11$(a)$ we have $|X| \leq n - \delta < (\tfrac{1}{3} + 2\eta)n$ and so $|X| < \tfrac{19}{10}(2\delta - n)$. Further $|\mathrm{int}(G)| \geq n - 2|X| \geq 2\delta - n > \tfrac{n}{3} - 4\eta n > n - \delta - 11\eta n$, which again contradicts the conditions of Fact 13.   □

Fact 14. *Lemma 9 is true for $\tfrac{3n-2}{5} \leq \delta < \tfrac{2n-1}{3}$.*

Proof of Fact 14. 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. Since $uv \in \mathrm{int}(G)$ 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 $|\mathrm{int}(G)| \leq n - \delta$. Hence, by Lemma 11$(a)$, G cannot have more than two triangle components. In particular, all vertices in $\mathrm{int}(G)$ lie in both triangle components of G. So if $|\mathrm{int}(G)| \geq n - \delta - 11\eta n$ then $\mathrm{int}(G)$ is the desired large independent set for Case (S3). If moreover all triangle component exteriors are of size $\tfrac{19}{10}(2\delta - n)$ at most we are in Case (S3). Otherwise (if $\mathrm{int}(G)$ is small or a triangle component exterior is large) Fact 13 gives $\mathrm{CTF}(G) \geq \min(\mathrm{sp}(n, \delta + \eta n), \tfrac{11}{20}n)$ which is Case (S2).   □

For the remainder of the proof, we suppose $\delta < \tfrac{3n-2}{5}$ and accordingly $r \geq 3$ and $r' \geq 2$. For dealing with this case we first establish two 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 15. *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 $\delta(G[U_1]) \geq \delta_1$ then $\mathrm{CTF}(G) \geq \min(3\lfloor |U_1|/2 \rfloor, 3\delta_1, 2\delta - n + |U_2|)$.*

Proof of Fact 15. By Proposition $12(a)$ we can find a matching M' in U_1 covering

$$\min(2\lfloor |U_1|/2 \rfloor, 2\delta_1)$$

vertices. Let M be a subset of M' covering $\min(2\lfloor |U_1|/2 \rfloor, 2\delta_1, (4\delta - 2n + 2|U_2|)/3)$ vertices. 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 $|\Gamma(x,y)| \geq 2\delta - (n - |U_2|)$. We claim that this implies that T can be constructed, covering all of M. Indeed, in each step of the greedy procedure we have strictly more than $2\delta - (n - |U_2|) - 3|M| \geq 0$ common neighbours of $e \in M$ available. 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. □

Below, our goal will be to show that $\mathrm{int}(G)$ is an independent set. The following fact prepares us for this step.

Fact 16. *Let uv be an edge in $\mathrm{int}(G)$. Unless $r' = 2$ at least one vertex, u or v, is contained in at most $r' - 1$ triangle components.*

Proof of Fact 16. Let C_1 be the triangle component containing $uv \in \mathrm{int}(G)$ along with the (non-empty) common neighbourhood $\Gamma(u, v)$ (and perhaps some other neighbours of u or v separately). Suppose that $C \neq C_1$, and u is a vertex of C. Then by Lemma $11(c)$, there are at least $2\delta - n + 1$ neighbours x of u such that the edge ux is in C. Now suppose that u lies in at least $r' - 1$ triangle components other than C_1. It follows that there is a set $U_u \subseteq \Gamma(u)$ of vertices x such that ux is not in C_1, with $|U_u| \geq (r' - 1)(2\delta - n + 1)$, since no edge lies in two distinct triangle components. Suppose furthermore that v too lies in at least $r' - 1$ triangle components other than C_1. Then there exists an analogously defined set U_v. Since all vertices of $\Gamma(u, v)$ form triangles of C_1 with u and v, the three sets $\Gamma(u, v)$, U_u and U_v are pairwise disjoint, and thus $|U_u \cup U_v| \geq (2r' - 2)(2\delta - n + 1)$. Now given any $x \in \Gamma(u, v)$, since ux and vx are both in C_1, x cannot be adjacent to any vertex of $U_u \cup U_v$. But then $\delta \leq d(x) < n - (2r' - 2)(2\delta - n + 1)$ which is equivalent to $2r' - 2 < (n - \delta)/(2\delta - n + 1)$. By (3) the right-hand side is at most r and thus we get $2r' - 2 < r$. Since $r \leq r' + 1$ however this is a contradiction unless $r' \leq 2$. □

We assume from now on, that

$$(12) \qquad\qquad \mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n),$$

that is, we are not in Cases (S1) or (S2). Our aim is to conclude that then $(*)$ $\mathrm{int}(G)$ 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 $r = r' + 1 = 3$ and $r = r' + 1 = 4$ (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, which are treated by Fact 17, require a somewhat technical case analysis, which we prefer to defer to the end of the section.

Fact 17. *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 18. *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.*

Proof of Fact 18. Recall that we have $r \geq 3$ at this point of the proof. Moreover, the cases $r = r' + 1 = 3$ and $r = r' + 1 = 4$ are handled by Fact 17. So we assume we are not in these cases; in particular, $r' \geq 3$. We will show that then each vertex of $\mathrm{int}(G)$ is contained in at least r' triangle components. Once we establish this, Fact 16 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 to (12). Indeed, let $w \in \mathrm{int}(G)$ and suppose that there are $k > 1$ triangle components $C_1, \ldots, C_k$ containing w. For $i \in [k]$ let U_i be the set of neighbours u of w such that $uw \in C_i$. By Lemma 11(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. No vertex in U_1 has a neighbour in U_2, since the components are distinct. In addition, all edges in $G[U_1]$ are triangle connected, because $U_1 \subseteq \Gamma(w)$. Therefore Fact 15 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 $\min(3\lfloor \delta/(2r' - 2) \rfloor, 4\delta - 2n)$ vertices. Now since (4) holds, we have $\frac{3}{2}\delta/(r' - 1) - 2 \geq \mathrm{sp}(n, \delta + \eta n)$. Since $r \geq r' \geq 3$ and we have excluded the case $r = r' + 1 = 4$, by (5) we have $4\delta - 2n \geq \mathrm{sp}(n, \delta + \eta n)$. It follows that T covers at least $\mathrm{sp}(n, \delta + \eta n)$ vertices, in contradiction to (12). □

Fact 19. *We are in Case* (S3).

Proof of Fact 19. Fact 18 tells us that $\mathrm{int}(G)$ is an independent set. By Lemma 11(a) and the fact that $\delta > n - \delta$ we have that every triangle

component in G has an exterior, and by Lemma 11(b) that there are no edges between any triangle component exteriors. Hence, to show that we are in Case (S3), it is enough to prove that

$$(13) \quad |\operatorname{int}(G)| := \alpha \geq n - \delta - 11\eta n \qquad \text{and} \qquad |X_1| \leq \frac{19}{10}(2\delta - n)$$

for the biggest triangle component exterior X_1 in G. Suppose for a contradiction that this is not the case. We first claim that this forces G to have exactly r' triangle components.

Indeed, assume G has $k \geq r' + 1$ triangle components. Each of these components C has vertices in its exterior $\partial(C)$, and so by Lemma 11(b) the minimum degree of G implies $|\partial(C)| \geq \delta - \alpha + 1 \geq 2\delta - n + 1$. We let these triangle component exteriors be $X_1, \ldots, X_k$, with X_1 being the biggest. Since $n = |X_1 \dot\cup \ldots \dot\cup X_k \dot\cup \operatorname{int}(G)|$, we have $(r' + 1)(\delta - \alpha) + \alpha < n$. We distinguish two cases.

Case 1: (13) fails because $\alpha < n - \delta - 11\eta n$. Then we obtain

$$(r' + 1)\delta < n + r'\alpha < n + r'(n - \delta - 11\eta n)$$
$$= (r' + 1)n - (9r' - 1)\eta n - r'\delta - (2r' + 1)\eta n.$$

Straightforward manipulation gives

$$\delta + \eta n < \frac{(r' + 1)n - (9r' - 1)\eta n}{2(r' + 1) - 1}.$$

Since $(9r' - 1)\eta n \geq 9r' - 1 \geq r'$ this contradicts (3) applied to $r' = r_p(n, \delta + \eta n)$.

Case 2: (13) fails because $|X_1| > \frac{19}{10}(2\delta - n)$. Let x be any vertex in X_2. Since x has at least δ neighbours, none of which are in $X_1 \dot\cup X_3 \dot\cup \ldots \dot\cup X_k$, we have

$$1 + \delta + \frac{19}{10}(2\delta - n) + (k - 2)(2\delta - n + 1) \leq n, \text{ hence}$$

$$\frac{19}{10}(2\delta - n) + (r' - 1)(2\delta - n) < n - \delta.$$

By (3) we have $r' \geq (n - \delta - \eta n)/(2\delta + 2\eta n - n + 1)$. Combined with the last inequality, this gives

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

Now provided that $\eta < 2\mu^2/3$, and since $2\delta - n \geq 2\mu n$, we have

$$(2\delta - n + 2\eta n + 1)(1 - \mu) < 2\delta - n + 3\eta n - \mu(2\delta - n)$$
$$\leq 2\delta - n + 3\eta n - 2\mu^2 n < 2\delta - n,$$

and we obtain $\frac{9}{5}\mu n + (1 - \mu)(n - \delta - \eta n) < n - \delta$ which is a contradiction since $n - \delta < n/2$ and $\eta < \mu$.

Hence, if (13) fails, then G has indeed exactly r' triangle components.

Now we use this fact in order to derive a contradiction to (12). Observe that, if $r' = 2$, and accordingly $\delta \geq (\frac{3}{5} - 2\eta)n$, then Fact 13

implies that (13) holds, because according to (12) we have $\mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n)$. In the remainder we assume $r' \geq 3$.

Since every vertex in X_1 has neighbours only in X_1 and $\mathrm{int}(G)$, and $|\mathrm{int}(G)| \leq n - \delta$, we have $\delta(G[X_1]) \geq 2\delta - n$. Furthermore, since no vertex in X_1 has neighbours in either X_2 or X_3, and $|X_2 \dot\cup X_3| \geq 2(2\delta - n + 1)$, we can apply Fact 15 to obtain

$$\mathrm{CTF}(G) \geq \min\big(3\lfloor |X_1|/2 \rfloor, 3(2\delta - n), 2\delta - n + 2(2\delta - n + 1)\big)$$
$$= \min\big(3\lfloor |X_1|/2 \rfloor, 3(2\delta - n)\big).$$

Now by (5), $\mathrm{CTF}(G) \geq 3(2\delta - n)$ is a contradiction to (12), so to complete our proof it remains to show that if (13) fails, then $\mathrm{CTF}(G) \geq 3\lfloor |X_1|/2 \rfloor$ is also a contradiction to (12). Again, we distinguish two cases.

Case 1: (13) fails because $\alpha < n - \delta - 11\eta n$. Since X_1 is the largest exterior, we have $|X_1| \geq (\delta + 11\eta n)/r'$. But we have by (4) that

$$\mathrm{sp}(n, \delta + \eta n) \leq \frac{3}{2}\frac{\delta + 3\eta n}{r'} - 2 < 3\Big\lfloor \frac{\delta + 11\eta n}{2r'} \Big\rfloor,$$

so that $\mathrm{CTF}(G) \geq 3\lfloor |X_1|/2 \rfloor$ is indeed a contradiction to (12).

Case 2: (13) fails because $|X_1| > \frac{19}{10}(2\delta - n)$. Then $\mathrm{CTF}(G) \geq 3\lfloor |X_1|/2 \rfloor \geq \frac{57}{20}(2\delta - n) - 2$, which by (5) is a contradiction to (12), as desired. □

This completes, modulo the proof of Fact 17, the proof of Lemma 9. □

It remains to show Fact 17. Note that we can use all facts from the proof of Lemma 9 that precede Fact 17. We will further assume that all constants and variables are set up as in this proof.

Proof of Fact 17. Recall that we assumed (12), i.e., $\mathrm{CTF}(G) < \mathrm{sp}(n, \delta + \eta n)$, in this part of the proof of Lemma 9. We distinguish two cases.

Case 1: $r = 3$ and $r' = 2$. In this case $\delta(G) \in [(\frac{3}{5} - 2\eta)n, (\frac{3}{5} + \eta)n]$. Trivially each vertex of $\mathrm{int}(G)$ is contained in at least $r' = 2$ triangle components. Suppose for a contradiction that there is an edge uv in $\mathrm{int}(G)$. 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 $V_1 := \{y : vy \in C\}$ and let $U_2 := \Gamma(u) \setminus U_1$ and $V_2 := \Gamma(v) \setminus V_1$. Observe that $U_2 \cap V_2 = \emptyset$.

By definition x is not in, and has no neighbour in, $U_2 \dot\cup V_2$. It follows that $|U_2 \dot\cup V_2| < n - \delta \leq (\frac{2}{5} + 2\eta)n$. On the other hand, by Lemma 11(c), we have $|U_2|, |V_2| > 2\delta - n \geq \frac{1}{5}n - 4\eta n$, and thus

$$|U_2|, |V_2| \in \big[(\tfrac{1}{5} - 4\eta)n, (\tfrac{1}{5} + 6\eta)n\big].$$

Since $d(u) \geq \delta \geq (\frac{3}{5} - 2\eta)n$, we have $|U_1| \geq \delta - |U_2| \geq (\frac{2}{5} - 8\eta)n$. But no vertex in U_2 is adjacent to any vertex in U_1. This implies that every vertex in U_2 is adjacent to all but at most $n - \delta - |U_1| \leq 10\eta n$ vertices

outside U_1. Since $\eta < \frac{1}{1000}$ we have $|U_2| > 20\eta n$, so $\delta(G[U_2]) > |U_2|/2$, and by Proposition 12(a), U_2 contains a matching M_u with $\lfloor |U_2|/2 \rfloor$ edges. Since each vertex of U_2 has at most $10\eta n$ non-neighbours outside U_1, each pair of vertices has common neighbourhood covering all but at most $20\eta n$ vertices of $V(G) \setminus U_1$. In particular, the common neighbourhood of each edge of M_u covers all but at most $20\eta n$ vertices of $V(G) \setminus U_1$. Similarly, V_2 contains a matching M_v with $\lfloor |V_2|/2 \rfloor$ edges, and the common neighbourhood of each edge covers all but at most $20\eta n$ vertices of $V(G) \setminus V_1$.

Since $20\eta n < |U_2|/4$ and $U_2 \cap V_1 = \emptyset$, the common neighbourhood of each edge of M_v contains more than half of the edges of M_u. By symmetry, the reverse is also true. Thus all edges in $M_u \dot\cup M_v$ are in the same triangle component of G. Finally, each edge of $M_u \dot\cup M_v$ has at least $\delta - 10\eta n - |U_2 \dot\cup V_2| \geq (\frac{1}{5} - 24\eta)n$ common neighbours outside $U_2 \dot\cup V_2$. Choosing greedily for each edge of $M_u \dot\cup M_v$ in succession distinct common neighbours outside $U_2 \dot\cup V_2$, we obtain a connected triangle factor with $\min(\lfloor |U_2|/2 \rfloor + \lfloor |V_2|/2 \rfloor, (\frac{1}{5} - 24\eta)n) = (\frac{1}{5} - 24\eta)n$ triangles. But then $\mathrm{CTF}(G) \geq (\frac{3}{5} - 72\eta)n > n/2 > \mathrm{sp}(n, \delta + \eta n)$, a contradiction to (12). This proves Fact 17 for the case $r = 3$ and $r' = 2$.

Case 2: $r = 4$ and $r' = 3$. This implies that $(\frac{4}{7} - 2\eta)n \leq \delta(G) \leq (\frac{4}{7} + \eta)n$, and consequently $\mathrm{sp}(n, \delta + \eta n) < (\frac{2}{7} + 2\eta)n$. We first prove two statements about the structure of G which are forced by (12).

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

Proof of (Ψ). For the lower bound on $|U'|$, observe that by (c) of Lemma 11 we have $\delta(G[U']) \geq 2\delta - n \geq (\frac{1}{7} - 4\eta)n$. To obtain the upper bound, again by Lemma 11(c) we have $\delta(G[U]) \geq 2\delta - n$, and since the sets U and U' are neighbours of u in different triangle components C and C', there are no edges from U to U'. Furthermore, since any edge in $G[U]$ forms a triangle with u using an edge from u to U, all edges in $G[U]$ are in C. Now by Fact 15 we have

$$\mathrm{CTF}(G) \geq \min(3\lfloor |U|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U'|).$$

Since $|U| \geq \delta/2$ we have $3\lfloor |U|/2 \rfloor \geq (\frac{3}{7} - 3\eta)n - 2 > \mathrm{sp}(n, \delta + \eta n)$. By (5) we have $3(2\delta - n) > \mathrm{sp}(n, \delta + \eta n)$. Because (12) holds, we have $2\delta - n + |U'| < \mathrm{sp}(n, \delta + \eta n) < (\frac{2}{7} + 2\eta)n$, and therefore $|U'| < (\frac{1}{7} + 6\eta)n$. Now the claimed lower and upper bounds on $|U|$ follow from $U = \Gamma(u) \setminus U'$, and from the fact that no vertex in U' has a neighbour in U, respectively.   □

(Ξ) If a vertex 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 $i \in [3]$.

Proof of (Ξ). Assume that U_1 is the largest of the three sets. By (c) of Lemma 11 we have $\delta(G[U_i]) \geq 2\delta - n \geq (\frac{1}{7} - 4\eta)n$ for each i, so $|U_i| > (\frac{1}{7} - 4\eta)n$ for each i. As in the previous case, there can be no edge from U_1 to $U_2 \dot\cup U_3$, and all edges in U_1 are triangle-connected. Thus by Fact 15 we have

$$\text{CTF}(G) \geq \min \left(3\lfloor |U_1|/2 \rfloor, 3(2\delta - n), 2\delta - n + |U_2 \dot\cup U_3|\right).$$

Now since $\text{sp}(n, \delta + \eta n) < (\frac{3}{7} - 10\eta)n$ and (12) holds, we have

$$3\lfloor |U_1|/2 \rfloor < \text{sp}(n, \delta + \eta n) \leq (\frac{2}{7} + 2\eta)n$$

which implies $|U_1| < (\frac{4}{21} + 2\eta)n$. Since $|U_2|, |U_3| \leq |U_1|$ this completes the desired upper bounds. The lower bounds follow from $|U_1| + |U_2| + |U_3| \geq \delta \geq (\frac{4}{7} - 2\eta)n$. $\qquad\qquad\square$

Next we show that
(Θ) $\text{int}(G)$ is an independent set.

Proof of (Θ). Assume for a contradiction that there is an edge $uv \in \text{int}(G)$. By Fact 16 one of the vertices of this edge, say u, is in only 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 $V_1, \ldots, V_k$ according to the triangle component containing the edge to v. Assume further that $\Gamma(u, v) \subseteq U_1 \cap V_1$, so that $U_2, V_2, \ldots, V_k$ are pairwise disjoint. Let $x \in \Gamma(u, v)$. Since x has neighbours in neither U_2 nor V_2, and since by Lemma 11(c) we have $|V_2| > (\frac{1}{7} - 4\eta)n$, we conclude that $\delta \leq d(x) \leq n - 1 - |U_2| - |V_2|$. In particular, $|U_2| < (\frac{3}{7} - 8\eta)n$ because $\delta \geq (\frac{4}{7} - 2\eta)n$, and therefore by ($\Psi$) we have

$$(\frac{1}{7} - 4\eta)n < |U_2| < (\frac{1}{7} + 6\eta)n .$$

Next we want to derive analogous bounds for $|V_2|$. For this purpose we first show that $k = 2$.

Indeed, if we had $k = 3$, then by (Ξ)

$$d(x) \leq n - 1 - |U_2| - |V_2| - |V_3|$$
$$\leq n - 1 - (\frac{1}{7} - 4\eta)n - 2(\frac{4}{21} - 6\eta)n < (\frac{10}{21} + 16\eta)n < \delta ,$$

and this contradicts $\delta(G) \geq \delta$. Similarly, if $k \geq 4$, then by Lemma 11(c) we have $|V_i| \geq (\frac{1}{7} - 4\eta)n$ for each i, and hence

$$d(x) \leq n - 1 - |U_2| - |V_2| - |V_3| - |V_4| < (\frac{3}{7} + 16\eta)n < \delta ,$$

which too is a contradiction. It follows that $k = 2$ as claimed.

Hence, we can argue analogously as before (for U_2) that $|V_2| > (\frac{3}{7} - 8\eta)$ would contradict $d(x) \geq \delta$. Consequently, by (Ψ) we have

$$(\frac{1}{7} - 4\eta)n < |V_2| < (\frac{1}{7} + 6\eta)n .$$

We now argue that this yields a contradiction to (12) in much the same way as we argued in the $r = r' + 1 = 3$ case. Every vertex of U_2 is adjacent to all but at most $n - |U_1| - \delta \leq 10\eta n$ vertices of $V(G) \backslash U_1$. Since $|U_2| > 20\eta n$, by Proposition 12(a) there is a matching M_u in U_2 covering all but at most one vertex of U_2. Each edge of M_u has at least $\delta - 10\eta n \geq (\frac{4}{7} - 12\eta)n$ common neighbours outside U_1. Similarly, in V_2 there is a matching M_v covering all but at most one vertex of V_2, each edge of which has at least $(\frac{4}{7} - 12\eta)n$ common neighbours outside V_1. Since $\Gamma(u, v) = U_1 \cap V_1$, we have $U_1 \cap V_2 = \emptyset$. It follows that every edge of M_v has more than half of the edges of M_u in its common neighbourhood, and thus the edges $M_u \dot\cup M_v$ are triangle connected. Choosing greedily for each edge in $M_u \dot\cup M_v$ in succession a distinct common neighbour outside $M_u \dot\cup M_v$, we obtain a connected triangle factor with as many triangles as there are edges in $M_u \dot\cup M_v$. Since $|U_2|, |V_2| > (\frac{1}{7} - 4\eta)n$, we have $\mathrm{CTF}(G) > (\frac{3}{7} - 12\eta)n - 3 > \mathrm{sp}(n, \delta + \eta n)$, contradicting (12). This completes the proof that $\mathrm{int}(G)$ is an independent set. $\qquad\qquad\square$

It remains to show that each vertex $u \in \mathrm{int}(G)$ is contained in at least $r' = 3$ triangle components. Assume for a contradiction that this is not the case and that some vertex u is only contained in 2 triangle components, C and C'. Let U and U', respectively, be the neighbours of u on edges in C and C'. Without loss of generality $|U| \geq |U'|$. Because $\mathrm{int}(G)$ is an independent set, U and U' are contained in the exteriors of C and C'. By Lemma 11(b) there are thus no edges between U and $\partial(C')$. By Lemma 11(c) we have $\delta(G[U]) \geq 2\delta - n$, and since $U \subseteq \partial(C)$ every edge of $G[U]$ is in C. It follows that we may apply Fact 15 to obtain

$$\mathrm{CTF}(G) \geq \min\left(3\lfloor |U|/2 \rfloor, 3(2\delta - n), 2\delta - n + |\partial(C')|\right).$$

Since $|U| \geq \delta/2$ we have $3\lfloor |U|/2 \rfloor \geq (\frac{3}{7} - 3\eta)n - 2 > \mathrm{sp}(n, \delta + \eta n)$. By (5) we have $3(2\delta - n) > \mathrm{sp}(n, \delta + \eta n)$. Since (12) holds, we conclude that $2\delta - n + |\partial(C')| < \mathrm{sp}(n, \delta + \eta n) < (\frac{2}{7} + 2\eta)n$, and therefore $|\partial(C')| < (\frac{1}{7} + 6\eta)n$.

Now any vertex in $\partial(C')$ has neighbours only in $\partial(C') \dot\cup \mathrm{int}(G)$, and therefore $|\mathrm{int}(G)| \geq \delta - |\partial(C')| \geq (\frac{3}{7} - 8\eta)n$. The vertex u has neighbours only in $U' \subseteq \partial(C')$ and U, and therefore

$$|U| \geq \delta - |U'| \geq \delta - |\partial(C')| \geq (\tfrac{3}{7} - 8\eta)n\,.$$

By Lemma 11(c) we have $\delta(G[U]) \geq 2\delta - n \geq (\frac{1}{7} - 4\eta)n$, and since $|U| > (\frac{2}{7} - 8\eta)n$ we obtain by Proposition 12(a) a matching M in U with at least $(\frac{1}{7} - 4\eta)n$ edges. Now each vertex in $\mathrm{int}(G)$ is adjacent to all but at most $n - \delta - |\mathrm{int}(G)| \leq 10\eta n$ vertices outside $\mathrm{int}(G)$. In particular, each vertex in $\mathrm{int}(G)$ is adjacent to all but at most $10\eta n$ vertices of M, and is therefore a common neighbour of all but at most $10\eta n$ edges

of M. We now match greedily vertices of $\text{int}(G)$ with distinct edges of M forming triangles. Since $|\text{int}(G)| > |M|$, we will be forced to halt only when we come to a vertex $x \in \text{int}(G)$ which is not a common neighbour of any remaining edge of M, i.e., when we have used all but at most $10\eta n$ edges of M. It follows that we obtain a triangle factor T with at least $(\frac{1}{7} - 14\eta)n$ triangles. Since each triangle uses an edge of $M \subseteq G[U] \subseteq G[\partial(C)]$, T is a connected triangle factor, and we have $\text{CTF}(G) \geq (\frac{3}{7} - 42\eta)n > \text{sp}(n, \delta + \eta n)$ in contradiction to (12). $\square$

4. Near-extremal graphs

In this section we provide the proof of Lemma 10. To prepare this proof we start with two useful lemmas. The first will be used to construct squared paths and squared cycles from simple paths and cycles.

Lemma 20. *Given a graph G, let $T = (t_1, t_2, \ldots, t_{2l})$ 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 the vertices in $Q_{\sigma(i)}$ have at least i common neighbours in W, then there is a squared path*

$$(q_1, t_1, t_2, q_2, t_3, t_4, q_3, \ldots, t_{2\ell}, q_{\ell+1})$$

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_{3l}^2.

Proof. We need only ensure that for each i one can choose q_i such that 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 $Q_{\sigma(i)}$. $\square$

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

Lemma 21. *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 xy 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. $\qquad\square$

Before embarking upon the proof of Lemma 10 we give an outline of the method. We recall that the Szemerédi partition supplied to the Lemma is essentially the extremal structure. We shall show that the underlying graph either also has an extremal structure, or possesses features which actually 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 $\bigcup I$ may fail to contain some vertices of G with no neighbours in $\bigcup I$, and may contain a small number of edges meeting every vertex. However, observe that by the definition of an (ε, d)-regular partition, there are no vertices of $\bigcup I$ with more than $(\varepsilon + d)n$ neighbours in $\bigcup I$. Fortunately, it is possible to apply the following strategy in this case.

We start by separating those vertices with 'few' neighbours in $\bigcup I$, which we shall collect in a set W, and those with 'many'. We are then able to show (as Fact 23 below) that, if there are two vertex disjoint edges in W, then the sets $\bigcup B_1$ and $\bigcup B_2$ are in the same triangle component of G ('unexpectedly', since B_1 and B_2 are in different triangle components in R). We shall show that in this case it is possible to construct very long squared paths and cycles by making use of Lemma 8.

Hence we can assume that there are not two disjoint edges in W, which in turn implies that W is almost independent and will give us rather precise control about the size of W. In addition, the minimum degree condition will guarantee that almost every edge from W to the remainder of G is present. We would like to then say that in $V(G) \setminus W$ we can find a long path, which together with vertices from W forms a squared path (and similarly for squared cycles). Unfortunately since $G[W, V(G) \setminus W]$ 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 20—and the purpose of Lemma 21 is to provide paths and cycles with those nice properties. The remainder of our proof, then, consists of setting up conditions for the application of Lemma 21.

Proof of Lemma 10. Given $\nu > 0$, suppose the parameters $\eta > 0$ and $d > 0$ satisfy the following inequalities.

$$(14) \qquad \eta \leq \frac{\nu^4}{10^8} \qquad \text{and} \qquad d \leq \frac{\nu^4}{10^8}$$

Given $d > 0$, Lemma 8 returns a constant $\varepsilon_{\mathrm{EL}} > 0$. We set

$$(15) \qquad \varepsilon_0 = \min\left(\frac{\nu^4}{10^8}, \varepsilon_{\mathrm{EL}}\right).$$

Given m_{EL} and $0 < \varepsilon < \varepsilon_0$, Lemma 8 returns a constant n_{EL}. We set

$$(16) \qquad N = \max\left(1000 m_{\mathrm{EL}}^4, 100 \eta^{-1} \nu^{-1}, n_{\mathrm{EL}}\right).$$

Now let G, R, and the partition $V(R) = I \dot\cup B_1 \dot\cup \ldots \dot\cup B_k$ satisfy conditions (i)–(vi) of the lemma.

If $\delta(G) = \delta \geq \frac{2n-1}{3}$ then we can appeal to Theorem 1 to find a spanning squared path in G; if $\delta \geq \frac{2n}{3}$ then we can appeal to Theorem 2 to find C_ℓ^2 for each $\ell \in [3, n] \setminus \{5\}$. Therefore, the definition of $\mathrm{sp}(n, \delta)$ and $\mathrm{sc}(n, \delta)$ imply that we may assume $\delta < 2n/3$ in the following, and that we only need to find

(17)   squared paths and squared cycles of length at most $11n/20$.

We now start by investigating the sizes of I and of the B_i. Define $\delta' = (\delta/n - d - \varepsilon)m$. Since R is an (ε, d)-reduced graph we have

$$(18) \qquad\qquad \delta(R) \geq \delta' = (\delta/n - d - \varepsilon)m \,.$$

Observe that moreover

$$(19) \qquad\qquad |I| \leq m - \delta' \leq \left(1 - \frac{\delta}{n} + d + \varepsilon\right)m \,,$$

by (v) because clusters in I have δ' neighbours outside I in R. For $i \in [k]$, fix a cluster $C \in B_i$. By assumption (vi) C has neighbours only in $B_i \cup I$ in R. 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 have

$$|B_i| > \deg(C, B_i) \geq 2\delta' - m \geq \frac{m}{n}\big(2(\delta - dn - \varepsilon n) - n\big)$$
$$= \frac{m}{n}\big(2\delta - n - (d + \varepsilon)n\big) \,.$$

Now since $2\delta - n \geq 2\nu n$ by (i), we conclude from (14) and (15) that

$$(20) \qquad\qquad |B_i| \geq \frac{2m(2\delta - n)}{3n} \geq \frac{4}{3}\nu m \,.$$

We next show that each B_i is part of exactly one triangle component of R.

Fact 22. *For each $1 \leq i \leq k$ the following holds. All edges in $R[B_i]$ are triangle connected in R.*

Proof of Fact 22. By assumption (vi) we have

$$(21) \qquad\qquad |B_i| \leq 19m(2\delta - n)/(10n) \leq 39(2\delta' - m)/20 \,,$$

where the second inequality comes from (14) and (15). Since we have $\delta_R(B_i) \geq 2\delta' - m > |B_i|/2$, the graph $R[B_i]$ is connected. It follows that if there are two edges in $R[B_i]$ which are not triangle-connected, then there are two adjacent edges in $R[B_i]$ with this property. That is, there are vertices P, Q and Q' of B_i such that PQ is in triangle component C and PQ' is in triangle component C' with $C \neq C'$.

We now show that there are at least $2\delta' - m$ edges leaving P in $R[B_i]$ which are in C. There are two possibilities. First, suppose there are no C-edges from P to I. In this case, the common neighbourhood $\Gamma(PQ)$ must lie entirely in B_i. Every vertex of $\Gamma(PQ)$ makes a C-edge with P, and we have $|\Gamma(PQ)| \geq 2\delta' - m$ as required. Second, suppose that

there is a C-edge PP' with $P' \in I$. Since I is an independent set in R, the set $\Gamma(PP')$ lies entirely within B_i, and has size at least $2\delta' - m$. Again, every edge from P to $\Gamma(PP')$ is a C-edge, as desired.

By the identical argument, there are at least $2\delta' - m$ edges leaving P in $R[B_i]$ which are in C'. Since no edge is in both C and C', there are at least $2(2\delta' - m)$ edges leaving P in $R[B_i]$, so $|B_i| \geq 2(2\delta' - m)$. This contradicts (21). It follows that all edges of B_i are triangle connected, as desired.   □

We next define a set W of those vertices in G which have few neighbours in $\bigcup I$. The intuition is that W consists of $\bigcup I$ and only a few more vertices of G. To simplify notation, we set $\xi = \sqrt[4]{\varepsilon + d + 11\eta}$. By (14) and (15), we have

$$(22) \qquad\qquad \xi \leq \nu/100 \,.$$

Let W be the vertices of G which do not have more than ξn neighbours in $\bigcup I$. Since $\xi > d + \varepsilon$, by the independence of I and by the definition of an (ε, d)-regular partition, we have $\bigcup I \subseteq W$. By assumption (v) we have $|I| \geq (n - \delta - 11\eta n)m/n$. Hence every edge in W has at least

$$(23) \qquad 2(\delta - \xi n) - \left(n - |\bigcup I|\right) > \frac{\delta - (2\delta - n)}{16}$$

common neighbours outside $\bigcup I$, where we use assumption (i) that $2\delta - n > 2\nu n$, (14) and (22).

By this observation, the next fact implies that we are done if there are two vertex disjoint edges in W.

Fact 23. *If $u_1 v_1$ and $u_2 v_2$ are vertex disjoint edges of G such that for $i = 1, 2$ the edge $u_i v_i$ has at least $\delta - (2\delta - n)/16$ common neighbours outside $\bigcup I$, then G contains $P^2_{\mathrm{sp}(n,\delta)}$ and C^2_ℓ for each $\ell \in [3, \mathrm{sc}(n, \delta)] \setminus \{5\}$.*

Proof of Fact 23. Let D' be the set of clusters $C \in V(R) \setminus I$ such that $u_1 v_1$ has at most $2dn/m$ common neighbours in C. By the hypothesis, $u_1 v_1$ has at least $\delta - (2\delta - n)/16$ common neighbours outside $\bigcup I$. Of these, at most εn are in the exceptional set V_0 of the regular partition, and at most $2dn|D'|/m$ are in $\bigcup D'$. The remaining common neighbours must all lie in $\bigcup(V(R) \setminus (I \cup D'))$, and hence we have the inequality

$$\delta - \frac{2\delta - n}{16} - \varepsilon n - \frac{2dn|D'|}{m} \leq (m - |I| - |D'|)\frac{n}{m}$$
$$\overset{(v)}{\leq} n - (n - \delta - 11\eta n) - |D'|\frac{n}{m} \,.$$

Simplifying this, we obtain

$$\frac{n - 2dn}{m}|D'| \leq 11\eta n + \varepsilon n + \frac{2\delta - n}{16} \,,$$

and by (14) and (15), we get $|D'| \leq (2\delta - n)m/(14n)$.

Now let D be the set of clusters $C \in V(R) \setminus I$ such that either $u_1 v_1$ or $u_2 v_2$ has at most $2dn/m$ common neighbours in C. The same analysis holds for $u_2 v_2$, so we obtain

$$(24) \qquad\qquad |D| \leq \frac{(2\delta - n)m}{7n} \, .$$

Therefore, we conclude from (20) that $B_1 \setminus D \neq \emptyset$. Take $X \in B_1 \setminus D$ arbitrarily. We have

$$\deg(X, B_1) \overset{(vi)}{\geq} \deg(X) - |I| \geq \delta' - |I| \overset{(19)}{\geq} \delta' - \left(1 - \frac{\delta}{n} + d + \varepsilon\right)m$$

$$\overset{(18)}{\geq} \left(\frac{\delta}{n} - d - \varepsilon\right)m - \left(1 - \frac{\delta}{n} + d + \varepsilon\right)m$$

$$\overset{(14),(15)}{\geq} \frac{1}{2}(2\delta - n)\frac{m}{n} \overset{(24)}{>} |D| \, .$$

Thus there exists a cluster $Y \in \Gamma(X) \cap (B_1 \setminus D)$. Hence we have clusters $X, Y \in B_1 \setminus D$ such that $XY \in E(R)$. Analogously, we can find clusters $X', Y' \in B_2 \setminus D$ such that $X'Y' \in E(R)$.

Since $\delta_R(B_1), \delta_R(B_2) \geq \delta' - |I| \geq 2\delta' - m$, 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 $|M| - (m - |I| - \delta') = 2\delta' - m$ edges of M. In addition, by assumption (v), (14), and since $\delta < 2n/3$ we have $|I| > (\frac{1}{3} - 11\eta)m \geq \frac{1}{4}m$. Therefore we may choose greedily clusters in I to obtain a set T of at least

$$\min\left\{2\delta' - m, |I|\right\} \geq \min\left\{2\delta' - m, \frac{1}{4}m\right\}$$

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 $B_2 \cup I$.

By Fact 22, since each triangle in T_1 contains an edge of B_1, all triangles in T_1 are in the same triangle component as the edge XY. Similarly all the triangles in T_2 are in the same triangle component as the edge $X'Y'$.

Noting that ε satisfies (15) and $n > N$ satisfies (16), we can apply Lemma 8 with $X_1 = X_2 = X$, $Y_1 = Y_2 = Y$ to find a squared path starting with $u_1 v_1$ and finishing with $u_2 v_2$ using the triangles T_1. Similarly, using Lemma 8 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 $u_2 v_2$ and finishing with $u_1 v_1$ using the triangles T_2. Choosing appropriate lengths for these squared paths and concatenating them we get a squared cycle C_ℓ^2 in G, for any $36(m_{\text{EL}} + 2)^3 \leq \ell \leq 3(1 - d)\min\{2\delta' - m, m/4\}n/m$. Applying Lemma 8 to the copy of K_4 in B_1 directly we obtain C_ℓ^2 for each $\ell \in [3, 3n/m] \setminus \{5\}$. By (16) we

have $3n/m > 36(m_{\text{EL}} + 2)^3$, and by (5), (14), (15), and (17) we have $3(1-d)(2\delta' - m)n/m > \text{sp}(n, \delta) \geq \text{sc}(n, \delta)$ and $3(1-d)n/4 \geq 11n/20 > \text{sp}(n, \delta) \geq \text{sc}(n, \delta)$. It follows that G contains both $P^2_{\text{sp}(n,\delta)}$ and C^2_ℓ for each $\ell \in [3, \text{sc}(n, \delta)] \setminus \{5\}$ as required.   □

By (23), if there are two vertex disjoint edges in W, then we are done by Fact 23. Thus we assume in the following that no such two edges exist. This implies that there are two vertices in W which meet every edge in W. Since neither of these two vertices has more than ξn neighbours in $\bigcup I \subseteq W$, while $|I| > (\frac{1}{3} - 11\eta)m$ by (v) and because $\delta < 2n/3$, there is a vertex in W adjacent to no vertex of W. We conclude that

$$(25) \qquad n - \delta - 11\eta n \leq |\bigcup I| \leq |W| \leq n - \delta.$$

Our next goal is to extract from each set $\bigcup B_i$ a large set A_i of vertices which are adjacent to almost all vertices in W and are such that $G[A_i]$ has minimum degree somewhat above $|A_i|/2$. Because at least $|W|\delta - 2|W|$ edges leave W, the total number of non-edges between W and $V(G) \setminus W$ is at most

$$|W||V(G) \setminus W| - |W|(\delta - 2) \leq (n-\delta)(\delta + 11\eta n - \delta + 2) \overset{(25)}{\leq} 11\eta n^2 + 2n.$$

In particular, by the definition of ξ, by (14) and (16),

$$(26) \qquad \left|\{v \in V(G) \setminus W : \deg(v, W) < |W| - \xi^2 n\}\right| \leq \xi^2 n.$$

In addition, by assumption (vi) we have $|B_i| \leq 19m(2\delta - n)/(10n)$, which together with $\delta \leq 2n/3$, (14), (15) and (22) implies

$$(27) \qquad \left|\bigcup B_i\right| \leq \frac{19}{10}(2\delta - n) \leq \frac{19}{20}\delta < \delta - \xi n - (d + \varepsilon)n.$$

However, by assumption (vi) and the definition of an (ε, d)-regular partition, vertices in $\bigcup B_i$ send at most $(d + \varepsilon)n$ edges to $V(G) - \bigcup B_i - \bigcup I$. It follows from the definition of W that

$$\bigcup B_i \cap W = \emptyset \qquad \text{for all } i \in [k].$$

Furthermore, (14), (15) and (22) imply that $v \in \bigcup B_i$ has at least

$$(28) \quad \delta - |W| - (d + \varepsilon)n \overset{(25)}{\geq} 2\delta - n - (d + \varepsilon)n \overset{(27)}{>} |\bigcup B_i|/2 + 32\xi^2 n$$

neighbours in $\bigcup B_i$.

Now, 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. In the rest of this paragraph we determine some important properties of the sets A_i. By (26) we have

$$(29) \qquad \left|\bigcup_{i \in [k]} (\bigcup B_i) \setminus A_i\right| \leq \xi^2 n \qquad \text{for all } i \in [k].$$

But the vertices which are neither in W nor any of the sets A_i must be either in the exceptional set V_0 or in $\bigcup B_i \setminus A_i$ for some i. Hence we have

$$(30) \qquad \left| V_0 \cup \bigcup_{i \in [k]} \left(\bigcup B_i \right) \setminus A_i \right| \leq \varepsilon n + \xi^2 n < 2\xi^2 n \,.$$

Accordingly (28) implies that

$$(31) \qquad \delta(G[A_i]) \geq |A_i|/2 + 30\xi^2 n \,,$$

and since $|B_i| > \delta' - |I| \geq 2\delta' - m$ we have

$$(32) \quad |A_i| \geq \left| \bigcup B_i \right| - 2\xi^2 n \geq (1 - \varepsilon) \frac{n}{m} |B_i| - 2\xi^2 n \geq 2\delta - n - 3\xi^2 n$$

for each $i \in [k]$, where we used the definition of ξ, (14), 15, and (18) in the last inequality.

In the remainder of the proof we utilize the sets A_i in order to find the desired squared path and squared cycles. We start by showing that we obtain squared cycles on ℓ vertices for each $\ell \in [3, \frac{3}{2}|A_1|] \setminus \{5\}$. To see this note first that by Lemma 21 (with $B = \emptyset$) we find in A_1 a copy of $C_{2\ell'}$ for each $2\ell' \in \left[4, \min\{|A_1|, 2\frac{n}{4}\} \right]$. By the definition of A_1 every quadruple of consecutive vertices on such a cycle has at least $|W| - 4\xi^2 n$ common neighbours in W, and by the definition of ξ, (14), (15), and (25) we have $|W| - 4\xi^2 n \geq n/4$. Hence we can apply Lemma 20 to G and W to square this cycle. This gives us squared cycles of lengths ℓ with

$$3 \leq \ell \leq \min \left\{ \frac{3}{2}|A_1|, 3\frac{n}{4} \right\} \stackrel{(17)}{=} \frac{3}{2}|A_1|$$

such that ℓ is divisible by three, but not of lengths not divisible by three.

If we seek a squared cycle $C_{3\ell'+1}^2$ or $C_{3\ell'+2}^2$ (with $3\ell' + 2 \neq 5$) then we need to perform a process which we will call *parity correction* and which we explain in the following two paragraphs. We shall use this parity correction process also in all remaining steps of the proof to obtain squared cycles of lengths not divisible by 3.

For obtaining a squared cycle of length $3\ell' + 1$ we proceed as follows. We pick (using Turán's theorem) a triangle abc in A_1 and *clone* the vertex b, i.e., we insert a dummy vertex b' into G with the same adjacencies as b. Then we apply Lemma 21 to $A_1 - \{b\}$ to find a path $P = (a, p_2, p_3, \ldots, p_{2\ell'-1}, c)$ on $2\ell'$ vertices whose end-vertices are a and c. Finally we apply Lemma 20 to the path bPb', taking $Q_1 = (b, a)$, $Q_2 = (b, a, p_2, p_3)$ as the first quadruple 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 path $(q_1, b, a, \ldots, c, b')$ on $3(\ell'+1)$ vertices, which gives a squared cycle $(b, a, \ldots, c)$ in G (without q_1 and the clone vertex b') on $3\ell' + 1$ vertices as required.

If we seek a squared cycle of length $3\ell' + 2$ with $\ell' > 1$ on the other hand, then we perform a similar process, except that we identify not one triangle in A_1 but two triangles abc, xyz connected with an edge cx (which we obtain by the Erdős-Stone theorem). We apply Lemma 21 to find a path $P = (a, \ldots, z)$ in $A_1 \setminus \{b, c, y, z\}$ on $2\ell'$ vertices. We then apply Lemma 20 once to the path bPy and once to (b, c, x, y). Omitting the first vertex on each of the resulting squared paths and concatenating, we get a squared cycle $C_{3\ell'+2}^2$.

Hence we do indeed obtain squared cycles C_ℓ^2 for all $\ell \in [3, \frac{3}{2}|A_1|] \setminus \{5\}$. It remains to show that we can also find C_ℓ^2 for all ℓ with $\frac{3}{2}|A_1| \leq \ell \leq \mathrm{sc}(n, \delta)$ and that we can find $P_{\mathrm{sp}(n,\delta)}^2$.

For this purpose, we first re-incorporate the vertices that are neither in W nor in any of the sets A_i by examining in which of the A_i they have many neighbours. More precisely, for each $i \in [k]$, we let X_i be A_i together with all vertices in $V(G) \setminus W$ which are adjacent to at least $30\xi^2 n$ vertices of A_i. Because every vertex in $V(G) \setminus W$ has at least $\delta - |W|$ neighbours outside W, by (25) every vertex in $G - W$ is in X_i for at least one i. Moreover, by the definition of an (ε, d)-regular partition, assumption (vi) and since $A_j \subseteq \bigcup B_j$, we have for all $j \in [k]$ with $j \neq i$ that

$$(33) \qquad\qquad A_j \cap X_i = \emptyset\,.$$

Hence it follows from (30) that

$$(34) \qquad |X_i| < |A_i| + 2\xi^2 n \qquad \text{and} \qquad |X_1 - A_1| \leq 2\xi^2 n\,.$$

We finish the proof by distinguishing three cases.

Case 1: $|X_i \cap X_j| \geq 2$ for some $i \neq j$. 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 21 to A_i we can find a path from u_1 to u_2 of length ℓ' for any $4 \leq \ell' \leq |A_i| - 2$. 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 $2\ell'$-vertex cycle $T_{2\ell'}$ in $X_1 \cup X_2$ for any $10 \leq 2\ell' \leq |A_i| + |A_j| - 2$. There are no quadruples of consecutive vertices on $T_{2\ell'}$ using both v_1 and v_2. The four quadruples that use either v_1 or v_2 each have at least $(\xi - 3\xi^2)n > 100k$ common neighbours in W, where the inequality follows from (16), (22), from

$$(35) \qquad\qquad k \leq \nu^{-1}\,,$$

and from $\xi - 3\xi^2 > 0$. All other quadruples have at least $|W| - 4\xi^2 n$ common neighbours in W. So applying Lemma 20 we obtain a squared cycle on $3\ell'$ vertices. Again it is possible to perform parity corrections (prior to applying Lemma 21) so that in this case we have $C_\ell^2 \subseteq G$ for every $\ell \in [3, \frac{3}{2}(|A_i| + |A_j| - 10)] \setminus \{5\}$. By (32), we have $\mathrm{sc}(n, \delta) \leq \mathrm{sp}(n, \delta) < \frac{3}{2}(|A_i| + |A_j| - 10)$.

Case 2: for some i every vertex of A_i is adjacent to at least one vertex outside $X_i \cup W$. Since

$$|A_i| \overset{(29)}{\geq} \left|\bigcup B_i\right| - \xi^2 n \overset{(20)}{\geq} \frac{4}{3}\nu(1-\varepsilon)n - \xi^2 n \overset{(22)}{\geq} 13\xi n \overset{(22),(35)}{>} 31k\xi^2 n$$

we can certainly find $j \neq i$ such that there are $31\xi^2 n$ vertices in A_i all adjacent to vertices of $X_j \setminus X_i$. Since no vertex of $X_j \setminus X_i$ 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 $u_1, u_2 \in A_i$ to $v_1, v_2 \in X_j$. Choosing distinct neighbours y_1 of v_1 and y_2 of v_2 in A_j and applying the same reasoning as in the previous case we are done.

Case 3: for each $i \neq j$ we have $|X_i \cap X_j| \leq 1$, and for each i 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 first focus on finding a squared path on $\mathrm{sp}(n, \delta)$ vertices in G, and then turn to the squared cycles which will complete the proof. If for some $i \neq j$ we have $|X_i \cap X_j| = 1$ then we obtain a squared path of the desired length as in Case 1. There we required two vertices in $X_i \cap X_j$ to obtain a squared cycle (which must return to its start), but one vertex suffices for a squared path to cross from X_i to X_j.

So, assume that the sets X_i are all disjoint. It follows that $k \leq (n - |W|)/(\delta - |W| + 1)$. Since $|W| \leq n - \delta$ by (25), this implies

$$k \leq \frac{n - (n - \delta)}{\delta - (n - \delta) + 1} = \frac{\delta}{2\delta - n + 1}\,.$$

Now if $k \geq r_p(n, \delta) + 1$ then we would have

$$r_p(n, \delta) + 1 \leq k \leq \frac{\delta}{2\delta - n + 1}\,,$$

and so

$$r_p(n, \delta) \leq \frac{n - \delta - 1}{2\delta - n + 1}\,,$$

but by (3) we have $r_p(n, \delta) \geq \frac{n-\delta}{2\delta - n + 1}$, so

$$k \leq r_p(n, \delta)\,.$$

Thus the largest of the sets X_i, say X_1, has at least

$$(36) \qquad |X_1| \geq \frac{n - |W|}{k} \overset{(25)}{\geq} \frac{\delta}{k} \geq \frac{\delta}{r_p(n, \delta)}$$

vertices.

We now want to apply Lemma 21 with $H = G[X_1]$ and 'bad' vertices $B = X_1 - A_1$. Note that by (34) there are at most $2\xi^2 n$ vertices in $B = X_1 - A_1$, and so we have

$$|B| \overset{(34)}{\leq} 2\xi^2 n \overset{(22)}{\leq} \frac{\nu\delta}{100} \overset{(35)}{\leq} \frac{\delta}{100k} \leq \frac{|H|}{100}\,.$$

Moreover, $\delta(H) = \delta(G[X_1]) \geq 30\xi^2 n$ by definition of X_1, and therefore every vertex in B has at least $30\xi^2 n \geq 9 \cdot 2\xi^2 n \geq 9|B|$ neighbours in H. In addition, vertices v in $H - B \subseteq A_1$ satisfy

$$\deg(v, X_1) \overset{(31)}{\geq} \frac{|A_1|}{2} + 30\xi^2 n \overset{(34)}{>} \frac{|X_1|}{2} + 25\xi^2 n$$
$$= \frac{|H|}{2} + 25\xi^2 n \overset{(16)}{\geq} \frac{|H|}{2} + 9|B| + 10.$$

Hence we can indeed apply Lemma 21, to obtain a path T covering $\min\{X_1, n/2\}$ vertices on which every quadruple of consecutive vertices contains at most one 'bad' vertex. Finally we want to apply Lemma 20 to the graph $G[X_1 \cup W]$ and the cycle T with the following ordering σ of the quadruples of consecutive vertices in T: σ is such that all quadruples containing vertices from B come first, followed (by an arbitrary ordering of) all other quadruples. There are at most $2 \cdot 2\xi^2 n$ quadruples containing vertices from $B = X_1 - A_1$, and by the definition of A_1 and of W, each of them has at least $(\xi - 3\xi^2)n \geq \xi^2 n$ common neighbours in W. All remaining quadruples have, by the definition of A_1, by (25) and since $\delta \leq 2n/3$, at least $|W| - 4\xi^2 n \geq \frac{n}{4} \geq \frac{1}{2}\min\{|X_1|, \frac{n}{2}\}$ common neighbours in W. Hence, we can indeed apply Lemma 20 to obtain a squared path on at least $\frac{3}{2}\min\{|X_1|, n/2\} \geq \mathrm{sp}(n, \delta)$ vertices, where the inequality follows from the definition of $\mathrm{sp}(n, \delta)$, from (17), and from (36).

At last, we show that we can find in G the desired long squared cycles in Case 3. Assume first that 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 $X_i \cap X_{i+1 \bmod s} = \{v_i\}$ 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 $A_{i+1 \bmod s}$ of v_i, and we may insist that all these $3s$ vertices are distinct. Similarly as before we can apply Lemma 21 to each $G[A_i]$ in turn and concatenate the resulting paths, in order to find a cycle $T_{2\ell'}$ for every $8s \leq 2\ell' \leq |A_1| + |A_2|$ on which there are no quadruples using more than one vertex outside $\bigcup_i A_i$. Again (checking the conditions similarly as before) we may apply Lemma 20 to $T_{2\ell'}$ to obtain a squared cycle on 3ℓ vertices. Finally by performing parity corrections we obtain C_ℓ^2 for every $\ell \in [3, \frac{3}{2}(|A_1| + |A_2|)] \setminus \{5\}$.

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 $|W| \leq n - \delta$, it follows from the definition of $r_c(n, \delta)$ (by establishing a relation similar to (3)) that $k \leq r_c(n, \delta)$, and by averaging, that the largest of the sets X_i, say X_1, contains at least $2\,\mathrm{sc}(n, \delta)/3$ vertices. As before, we can apply Lemma 21 to X_1 to obtain a cycle $T_{2\ell'}$ for each $4 \leq 2\ell' \leq |X_1|$ on which the 'bad' vertices from $B = X_1 - A_1$ are separated, and apply Lemma 20 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 $\ell \in [3, \mathrm{sc}(n, \delta)] \setminus \{5\}$. $\qquad\qquad\qquad\qquad\square$

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 [15]. In the setting of these results 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 paralleled by the behaviour of K_4-free graphs: For any minimum degree $\delta(G) > 2v(G)/3$ the graph G is not K_4-free. However, if $\delta(G) > 5v(G)/8$ then by the Andrásfai-Erdős-Sós theorem [2] the K_4-free graph G is forced to be tripartite, while for smaller values of $\delta(G)$ there exist more possibilities.

Extremal graphs. It is straightforward to check (from our proofs) that up to some trivial modifications the graphs $G_p(n, \delta)$ and $G_c(n, \delta)$ are the only 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.

However it is not the case that the only extremal graph excluding some C_ℓ^2 of chromatic number four is $K_{n-\delta, n-\delta, 2\delta-n}$ (cf. (*ii*) of our main theorem, Theorem 4). Let us briefly explain this. Suppose ℓ is not divisible by three. Since C_ℓ^2 has no independent set on more than $\lfloor \ell/3 \rfloor$ vertices, whenever we remove an independent set from C_ℓ^2 we must leave some three consecutive vertices, which form a triangle. Now suppose that we can find a graph H on δ vertices with minimum degree $2\delta - n$ which is both triangle-free and contains no even cycle on more than $2(2\delta - n)$ vertices. Then the graph G obtained by adding an independent set of size $n - \delta$ to H, all of whose vertices are adjacent to all of H, contains no squared cycle of length indivisible by three and no squared cycle with more than $3(2\delta - n)$ vertices.

To mention one possible H, take $\delta = \frac{6n}{11}$ and let H be obtained as follows. We take the disjoint union of three copies of $K_{n/11, n/11}$ and fix a bipartition. Now we add three vertex disjoint edges within one of the two partition classes, one between each copy of $K_{n/11, n/11}$. The resulting triangle-free graph has no even cycle leaving any copy of

$K_{n/11,n/11}$. Hence all even cycles have at most $\frac{2n}{11}$ vertices. However, it has odd cycles of up to $\frac{6n}{11} - 3$ vertices.

Long squared cycles. 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, requiring both a new version of 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 conjecture was 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 24. *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 $C_\ell^k \nsubseteq G$ 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

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, *Embedding large subgraphs into dense graphs*, Surveys in Combinatorics (S. Huczynka, J. Mitchell, and C. Roney-Dougal, eds.), Cambridge University Press, 2009, pp. 137–167.

15. D. Kühn and D. Osthus, *The minimum degree threshold for perfect graph packings*, Combinatorica **29** (2009), 65–107.

16. D. Kühn, D. Osthus, and A. Taraz, *Large planar subgraphs in dense graphs*, J. Comb. Theory, Ser. B **95** (2005), no. 2, 263–282.

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

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

19. 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.

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

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

Appendix A. Proof of Lemma 8

For the proof of Lemma 8 we apply the following version (which is a special case) of the Blow-up Lemma of Komlós, Sárközy and Szemerédi [11].

Lemma 25 (Blow-up Lemma [11]). *Given fixed $c, d > 0$, there exist $\varepsilon_0 > 0$ and n_{BL} such that for any $0 < \varepsilon < \varepsilon_0$ the following holds. Let H be any graph on at least n_{BL} vertices 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 $(3\varepsilon, d)$-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 ($i \in [4]$) of F are endowed with sets $C_{x_i} \subseteq V_j$ such that $x_i \in V_j$ and $|C_{x_i}| \geq c|V_j|$

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

We also say that the x_i in Lemma 25 are *image restricted* to C_{x_i}.

Proof of Lemma 8. Given d, we let $c = d^2/4$. Now Lemma 25 gives values $\varepsilon_0 > 0$ and n_{BL}. We choose $\varepsilon_{\mathrm{EL}} = \min(\varepsilon_0, d^2/24)$. Given $\varepsilon < \varepsilon_{\mathrm{EL}}$ and m_{EL} we choose

$$n_{\mathrm{EL}} = \max\left(2m_{\mathrm{EL}}n_{\mathrm{BL}}, \frac{6m^4}{\varepsilon}\right).$$

Let $n \geq n_{\mathrm{EL}}$, let G be an n-vertex graph, and let R' be an (ε, d)-reduced graph of G on $m \leq m_{\mathrm{EL}}$ vertices.

Fix a set $\mathcal{T}' = \{T_1', \ldots, T_{\mathrm{CTF}(R')/3}'\}$ 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 set of $\mathrm{CTF}(R')/3$ vertex disjoint triangles in R corresponding to $\mathcal{T}'$. We set $r = \mathrm{CTF}(R')/3$.

Our strategy now is as follows. We shall first fix a collection of suitable triangle walks $W_1, \ldots, W_{r-1}$ and W' in R. Next, for each of these triangle walks $W = (E_1, E_2, \ldots)$ we do the following. Let $\overrightarrow{U_1 V_1}$ be (a suitable) orientation of the first edge E_1 of W. We shall construct a sequence $Q(W, \overrightarrow{U_1 V_1})$ of vertices of R whose first two vertices are U_1 and V_1, in that order, and 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). Then we use this sequence $Q(W, \overrightarrow{U_1 V_1})$ to obtain a squared path in G following W, whose first two vertices are in U_1 and V_1. Finally, connecting suitable paths appropriately will lead to a proof of (i), (ii), and (iii).

We first construct the triangle walks $W_1, \ldots, W_{r-1}$ and W'. For each $1 \leq i \leq r - 1$ 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 W_i contains at least ten edges, that the first edge of W_{i+1} is not the same as the last edge of W_i, and such that each walk with more than ten edges is of minimal length. We have $|W_i| \leq \binom{m}{2}$ for each i. Let W' be the triangle walk obtained by concatenating $W_1, \ldots, W_{r-1}$.

Next, we describe how to construct the sequence $Q(W, \overrightarrow{A_1 B_1})$ for any triangle walk $W = (E_1, E_2, \ldots)$ in R and any orientation $\overrightarrow{A_1 B_1}$ of its first edge E_1. We construct $Q(W, \overrightarrow{A_1 B_1})$ iteratively as follows. Let $Q_1 = (A_1, B_1)$. Now for each $2 \leq i \leq |W|$ successively, we define Q_i as follows. The last two vertices A_{i-1}, B_{i-1} of Q_{i-1} are an orientation of E_{i-1}. If $E_i = A_{i-1} B_i$ we create Q_i by appending (B_i, A_{i-1}) to Q_{i-1}; if $E_i = B_{i-1} B_i$ we append $(B_i, A_{i-1}, B_{i-1}, B_i)$ 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 $Q(W, \overrightarrow{A_1 B_1}) = Q_{|W|}$.

We shall need the following observations concerning the lengths of sequences constructed in this way. It is easy to check by induction that for any triangle walk W with at least two edges whose first edge is $U_1 V_1$, we have

$$(37) \qquad |Q(W, \overrightarrow{U_1 V_1})| + |Q(W, \overrightarrow{V_1 U_1})| \equiv 1 \mod 3 \,.$$

Now consider the concatenation W' of the walks W_i. Let $U_1 V_1$ be the first edge of W_1. If we construct $Q(W', \overrightarrow{U_1 V_1})$ then the first edge $U_i V_i$ and the last edge $U'_i V'_i$ of each W_i obtains an orientation, say $\overrightarrow{U_i V_i}$ and $\overrightarrow{U'_i V'_i}$. Clearly, there are sequences $\tilde{Q}_i$ of vertices in T_i for $1 < i < r$, such that $Q(W', \overrightarrow{V_1 U_1})$ is the concatenation of

$$Q(W_1, \overrightarrow{V_1 U_1}), \tilde{Q}_2, Q(W_2, \overrightarrow{V_2 U_2}), \ldots, \tilde{Q}_{r-1}, Q(W_{r-1}, \overrightarrow{V_{r-1} U_{r-1}}) \,.$$

Further we let $\tilde{Q}_1 = T_1 - U_1 V_1$ and $\tilde{Q}_r = T_r - U'_{r-1} V'_{r-1}$. We define $f_i = |\tilde{Q}_i| \mod 3$ for $i \in [r]$. Together with (37) we obtain

$$|Q(W', \overrightarrow{U_1 V_1})| + |Q(W_1, \overrightarrow{V_1 U_1})| + \sum_{1 < i < r} \left(|Q(W_i, \overrightarrow{V_i U_i})| + f_i \right) \equiv 1 \mod 3$$

and hence

$$(38) \quad |Q(W', \overrightarrow{U_1 V_1})| + \sum_{i \in [r-1]} \left(|Q(W_i, \overrightarrow{V_i U_i})| + f_i \right) + f_r \equiv 0 \mod 3 \,.$$

This will enable us to construct cycles of lengths divisible by three later.

In order to construct squared paths in G from short vertex sequences in R we use the following fact.

Fact 26. *Let X_1, X_2, X_3 be vertices of R (not necessarily distinct), and Z be any set of at most $2\varepsilon |X_1|$ vertices of G. Suppose that $X_1 X_2$*

and X_1X_3 are edges of R. Suppose furthermore that we have two vertices u and v of G such that u and v have at least $(d-2\varepsilon)^2|X_1|$ common neighbours in X_1, and v has at least $(d-2\varepsilon)|X_2|$ neighbours in X_2.

Then there is a vertex $w \in X_1 - Z$ adjacent to u and v such that v and w have at least $(d-2\varepsilon)^2|X_2|$ common neighbours in X_2 and w has at least $(d-2\varepsilon)|X_3|$ neighbours in X_3.

Proof of Fact 26. Let W be the set of common neighbours of u and v in X_1. Since $X_1X_2 \in E(R)$, at most $2\varepsilon|X_1|$ vertices of W have fewer than $(d-2\varepsilon)|\Gamma(v) \cap X_2| \geq (d-2\varepsilon)^2|X_2|$ common neighbours with v in X_2. Since $X_1X_3 \in E(R)$ at most $2\varepsilon|X_1|$ vertices of W have fewer than $(d-2\varepsilon)$ neighbours in X_3. Finally since $6\varepsilon|X_1| < (d-2\varepsilon)^2|X_1|$ we can find a vertex of $W \setminus Z$ satisfying the desired properties. □

With these building bricks at hand we can now turn to the proofs of (i), (ii), and (iii).

Proof of (i), i.e., G contains $C_{3\ell}^2$ for each $3\ell \leq (1-d)\mathrm{CTF}(R)n/m$: When $\ell \leq (1-d)n/m$ we have $C_{3\ell}^2 \subseteq K_{(1-d)n/m,(1-d)n/m,(1-d)n/m}$ and thus by Lemma 25 we can find $C_{3\ell}^2$ as a subgraph of G (whose vertices are in T_1, with no restrictions required). Otherwise we use the following strategy. Let UV be the first edge of the triangle walk W_1.

Our first goal will be to construct a squared path P' in G which 'connects' T_1 to T_2, T_2 to T_3, and so on. For this purpose we choose two adjacent vertices u and v of G in U and V respectively, such that u and v have $(d-2\varepsilon)^2n/m$ common neighbours in both the third vertex of T_1 and the third vertex of $Q(W', \overrightarrow{UV})$, such that v has $(d-2\varepsilon)n/m$ neighbours in the fourth vertex of $Q(W', \overrightarrow{UV})$, and such that u has $(d-2\varepsilon)n/m$ neighbours in V. This is possible by the regularity of the various pairs. (Observe that the required sizes for the neighbourhoods and joint neighbourhoods are chosen large enough for an application of Lemma 25 in the triangle T_1.) Now we apply Fact 26 with the vertices u and v and the third, fourth and fifth vertices of $Q(W', \overrightarrow{UV})$ to obtain a third vertex v' in the third vertex of $Q(W', \overrightarrow{UV})$ such that u and v are adjacent to v'. By repeatedly applying Fact 26 we construct a sequence of vertices P' (starting with u, v, v'), where the ith vertex of P' is in the ith set of $Q(W', \overrightarrow{UV})$ and is adjacent to its two predecessors, and where the vertices are all distinct (noting that $3|W'| < \varepsilon n/m$). Thus P' is a squared path running from T_1 to T_{r-1} following all the triangle walks W_i.

In addition we construct similarly (and without re-using vertices) for each $1 \leq i \leq r - 1$ a squared path P_i following the triangle walk W_i. However, this time 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'. Again, for each P_i we insist that the

first two vertices have suitable neighbourhoods in T_i, and the last two in T_{i+1}, for an application of Lemma 25 in these triangles. Again, this is possible by regularity.

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 $P = P' \cup P_1 \cup \cdots \cup P_{r-1}$. Since at most $\varepsilon n/m$ vertices are removed, and each cluster of T_1 has size at least $(1-3\varepsilon)n/m$, even after removal all three pairs remain $(3\varepsilon, d)$-regular and each cluster still has size at least $(1-4\varepsilon)n/m$.

Thus the conditions of Lemma 25 are satisfied, and hence 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 because $c = d^2/4$). In this way we can construct a squared path on $3\ell_1 + f_1$ vertices for any integer $\ell_1 \in [10, (1-d)n/m]$ (since $3 \cdot 4\varepsilon < d$), where $f_1 \in \{0, 1, 2\}$ is as defined above (38). Similarly we may apply Lemma 25 to each T_i ($2 \le i \le r$), after removing P from T_i, to obtain squared paths S_i of length $3\ell_i + f_i$ for any integer $\ell_i \in [10, (1-d)n/m]$.

Finally $S = P' \cup S_1 \cup P_1 \cup \ldots \cup P_{r-1} \cup S_r$ forms a squared cycle in G. It follows from (38) that the length of S is divisible by three. We conclude that indeed $S = C_{3k}^2$, where we may choose any integer k with $3k \in [6m^3, (1-d)\,\mathrm{CTF}(R)n/m]$, as required.

Proof of (ii): When every triangle component of R contains K_4 we also want to 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. Hence, with the help of Fact 26, these sequences can be used to construct squared paths in G of length 3 (which is 0 mod 3), length 7 (1 mod 3), and length 11 (2 mod 3).

We construct C_ℓ^2 for $\ell \in [3, 20] \setminus \{5\}$ 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 25, construct a short path in A, B, C, D following the appropriate vertex sequence for $\ell \bmod 3$ and apply Lemma 25 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 $2m^2$ in total).

Proof of (iii): 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 an application of Fact 26. $\qquad\square$