Minimising the number of triangular edges

Vytautas Gruslys* Shoham Letzter*

March 19, 2017

Abstract

We consider the problem of minimising the number of edges that are contained in triangles, among n-vertex graphs with a given number of edges. For sufficiently large n, we prove an exact formula for this minimum, which partially resolves a conjecture of Füredi and Maleki.

1 Introduction

Mantel [9] proved that a triangle-free graph on n vertices has at most $\lfloor n^2/4 \rfloor$ edges. In other words, a graph on n vertices with at least $\lfloor n^2/4 \rfloor + 1$ edges contains a triangle. A natural question arises from this classical result: how many triangles must such a graph have? And, indeed, Rademacher [11] extended Mantel's result by showing that any graph on n vertices with $\lfloor n^2/4 \rfloor + 1$ edges contains at least $\lfloor n/2 \rfloor$ triangles, a bound that can readily be seen to be best possible (see Figure 1).

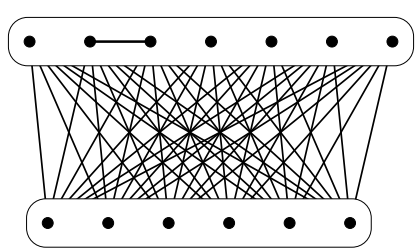

Figure 1: An n-vertex graph with $\lfloor n^2/4 \rfloor + 1$ edges and $\lfloor n/2 \rfloor$ triangles.

*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, CB3 0WB Cambridge, UK; e-mail: {v.gruslys,s.letzter}@dpmms.cam.ac.uk.

2010 *Mathematics Subject Classification*: 05C35.

Erdős [1] conjectured that a further generalisation holds: any graph on n vertices with at least $\lfloor n^2/4 \rfloor + l$ edges contains at least $l\lfloor n/2 \rfloor$ triangles, for every $1 \leq l < \lfloor n/2 \rfloor$. Erdős [1, 2] proved his conjecture for $l \leq cn$ for some constant $c > 0$. It is not hard to see that the bound on the number of triangles is best possible. Indeed, this bound can be achieved by adding l edges that do not span a triangle to the larger part of the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ (see Figure 2). The bound on l can also be easily seen to be best possible.

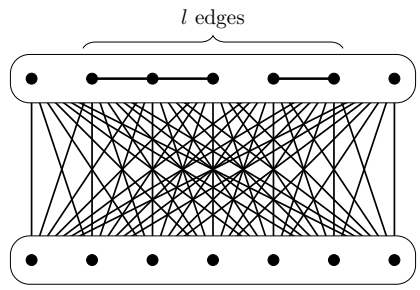

Figure 2: An n-vertex graph with $\lfloor n^2/4 \rfloor + l$ edges and $l\lfloor n/2 \rfloor$ triangles.

Erdős's conjecture was resolved by Lovász and Simonovits [7], who also characterised [8] the n-vertex graphs with $\lfloor n^2/4 \rfloor + l$ edges that minimise the number of triangles, for every $l \leq cn^2$ and some fixed $c > 0$. Razborov [12] asymptotically determined the minimal possible number of triangles in an n-vertex graph with $\lfloor n^2/4 \rfloor + l$ edges where $l = \Omega(n^2)$.

In this paper we consider a similar problem, concerning the number of edges that are contained in a triangle (we shall call such edges *triangular edges*), rather than the number of triangles. The first result in this direction was obtained by Erdős, Faudree and Rousseau [3] who proved that any n-vertex graph with $\lfloor n^2/4 \rfloor + 1$ edges has at least $2\lfloor n/2 \rfloor + 1$ triangular edges. This bound is best possible (see Figure 1).

It is very natural, similarly to the question about the number of triangles, to ask how many triangular edges an n-vertex graph with e edges must have, where e is an integer satisfying $\lfloor n^2/4 \rfloor < e \leq \binom{n}{2}$. After some thought, a natural example comes to mind. Given integers a, b, c, we denote by $G(a, b, c)$ the graph on $n = a + b + c$ vertices, which consists of a clique A of size a and two independent sets B and C of sizes b and c respectively, such that all edges between B and $A \cup C$ are present, and there are no edges between A and C (see Figure 3).

Note that the graph $G(a, b, c)$ has $\binom{a}{2} + b(a + c)$ edges and, as long as $a \geq 2$ and $b \geq 1$, precisely bc of them are non-triangular. We remark that the extremal example (depicted in Figure 1) for the aforementioned result by Erdős, Faudree and Rousseau [3] is isomorphic to $G(2, \lfloor n/2 \rfloor, \lceil n/2 \rceil - 2)$.

Füredi and Maleki [4] conjectured that the minimisers of the number of triangular edges are graphs of the form $G(a, b, c)$, or subgraphs of such graphs.

Conjecture 1. *Let n and $e > \lfloor n^2/4 \rfloor$ be integers and let G be an n-vertex graph with e edges that*

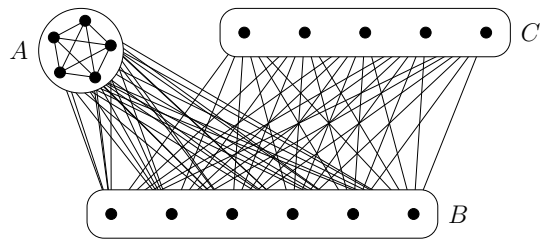

Figure 3: The graph $G(a, b, c)$ (here $a = 5, b = 6, c = 5$).

minimises the number of triangular edges. Then G is isomorphic to a subgraph of a graph $G(a, b, c)$ for some a, b, c.

The condition that G is isomorphic to a subgraph of a graph $G(a, b, c)$ (rather than to $G(a, b, c)$ itself) is due to the fact that we specify the exact number of edges, so the minimiser may be isomorphic to $G(a, b, c)$ with a few edges removed from $A \cup B$.

The conjecture implies, in particular, that every n-vertex graph with e edges has at least $g(n, e)$ triangular edges, where $g(n, e)$ is defined by

$$g(n, e) = \min \left\{ e - bc : a + b + c = n, \binom{a}{2} + b(a + c) \geq e \right\}.$$

Füredi and Maleki [4] proved an approximate version of the latter statement, which reads as follows.

Theorem 2 (Füredi and Maleki [4])**.** *Every n-vertex graph with e edges has at least $g(n, e) - 3n/2$ triangular edges.*

It is worth noting that, if $e \geq (1/4 + \Omega(1))n^2$ and $e \leq (1/2 - \Omega(1))n^2$, then $g(n, e) = \Omega(n^2)$ and $\binom{n}{2} - g(n, e) = \Omega(n^2)$. Therefore, Theorem 2 is asymptotically sharp in the range where the edge density is bounded away from 1/2 and 1 by small positive constants.

Our main result is an exact version of Theorem 2: we shall prove that an n-vertex graph with e edges has at least $g(n, e)$ triangular edges, provided that n is large enough. However, the bound on n does not depend on e, that is, as long as $n \geq n_0$ for some n_0, our theorem holds for any e such that $\lfloor n^2/4 \rfloor \leq e \leq \binom{n}{2}$.

Before we precisely state our result, we make a few remarks. Firstly, it turns out to be more convenient to consider the clearly equivalent problem of maximising the number of non-triangular edges among n-vertex graphs with e edges. Thus, given a graph G, we denote by $t(G)$ the number of non-triangular edges in G. Secondly, given n and e, instead of restricting our attention to n-vertex graphs with exactly e edges, we consider n-vertex graphs with at least e edges. Since the removal of a triangular edge cannot decrease the number of non-triangular edges, this slight reformulation does not change the problem, and yet it allows us to concentrate on graphs $G(a, b, c)$ without having to consider their subgraphs.

We are now ready to state our main result.

Theorem 3. *There exists n_0 such that, for any graph G on at least n_0 vertices, there exists a graph $H = G(a, b, c)$ (for some integers a, b, c) such that $|H| = |G|$, $e(H) \geq e(G)$ and $t(H) \geq t(G)$.*

We note that Theorem 3 comes close to proving Conjecture 1 (for sufficiently large n) as it shows that the minimum number of triangular edges is attained by a graph $G(a, b, c)$ or a subgraph of $G(a, b, c)$. However, we do not prove that such graphs are the only minimisers.

1.1 Structure of the paper

The proof of our main result, Theorem 3, is divided into three parts, according to the number of edges in the graph G. We treat separately graphs that are close to being bipartite, that is, whose number of edges is close to $n^2/4$; graphs that are close to being complete, that is, whose number of edges is close to $\binom{n}{2}$; and the middle range, where the number of edges is bounded away from both $n^2/4$ and $\binom{n}{2}$ by a constant factor of n^2.

We state Theorems 4 to 6, which are the theorems corresponding to the aforementioned three ranges, in Section 2, and give an overview of their proofs. In Section 3 we introduce some notation and describe the tools that we shall use to prove these theorems. We prove the theorems in Sections 4 to 6. Theorem 4, which deals with graphs with about $n^2/4$ edges, is proved in Section 4; the proof of Theorem 5, for the middle range, which is the most difficult of the three and is the heart of this paper, is given in Section 5; and Theorem 6 is proved in Section 6. We conclude the paper with Section 7 where we make a few remarks and mention some open problems.

2 Overview

We split the proof of Theorem 3 into three parts, according to the number of edges in the graph. We state the theorems corresponding to these three parts here.

The following theorem deals with e that is close to $n^2/4$, that is, $e \leq (1/4 + \delta)n^2$, where δ is a sufficiently small constant.

Theorem 4. *There exist n_0 and $\delta > 0$ such that the following holds. Let G be a graph with $n \geq n_0$ vertices and e edges, where $n^2/4 \leq e \leq (1/4 + \delta)n^2$. Then there exists a graph $H = G(a, b, c)$ such that $|H| = n$, $e(H) \geq e$ and $t(H) \geq t(G)$.*

The next theorem considers the case where e is bounded away from $n^2/4$ and $\binom{n}{2}$, namely, where $(1/4 + \delta)n^2 \leq e \leq (1/2 - \delta)n^2$ for a constant $\delta > 0$.

Theorem 5. *For every $\delta > 0$ there exists n_0 such that the following holds. Let G be a graph with $n \geq n_0$ vertices and e edges, where $(1/4 + \delta)n^2 \leq e \leq (1/2 - \delta)n^2$. Then there exists a graph $H = G(a, b, c)$ such that $|H| = n$, $e(H) \geq e$ and $t(H) \geq t(G)$.*

Finally, we consider the remaining case, where e is close to $\binom{n}{2}$, that is, $e \geq (1/2 - \delta)n^2$ for a sufficiently small constant $\delta > 0$.

Theorem 6. *There exist n_0 and $\delta > 0$ such that the following holds. Let G be a graph with $n \geq n_0$ vertices and e edges, where $e \geq (1/2 - \delta)n^2$. Then there exists a graph $H = G(a, b, c)$ such that $|H| = n$, $e(H) \geq e$ and $t(H) \geq t(G)$.*

It is clear that Theorems 4 to 6 imply Theorem 3: we first take a small $\delta > 0$ that works for Theorems 4 and 6 and then we choose a sufficiently large n_0 that works for all three theorems.

We now give some insight into our proofs. The rough plan for the proof of each of the theorems is the same. Assuming that G is an n-vertex graph with at least e edges that maximises the number of non-triangular edges, we first obtain rough information about the structure of the graph. In each of the cases, we partition the vertices of G into parts A, B, C, which relate to the three parts in a graph $G(a, b, c)$, in a way that will be explained in the proofs. In the next stage we use lower bounds on the number of non-triangular edges (coming from examples $G(a, b, c)$) to estimate the sizes of the sets A, B, C. The final stage uses the estimates on the sizes and some case specific arguments to conclude that G has the required structure, namely, that it is isomorphic to the graph $G(|A|, |B|, |C|)$.

The proofs of the two extremal cases, where e is close to either $n^2/4$ or $\binom{n}{2}$, are considerably easier than that of the middle range. The main reason for this is that in the extremal cases it is fairly easy to show that the graph G should be close to a graph $G(a, b, c)$, whereas in the middle range getting any handle on the structure of the graph is hard, and the initial structural properties that we find are less restrictive than in the two extremal cases.

We introduce two tools, which will be helpful in the proof of the middle range. The first one is a process of 'compression' that allows us to 'simplify' a graph without decreasing the number of edges or non-triangular edges. The second is the 'exchange lemma', which allows us to 'exchange' edges to non-triangular edges and vice versa. In other words, it allows us to replace a graph by another graph with (somewhat) fewer edges, but more non-triangular edges and vice versa. Both of these tools will be presented and explained in greater detail in Section 3.

3 Tools

In this section we introduce the tools that we will use throughout the paper. We start by describing some notation and simple definitions in Subsection 3.1. We introduce the notion of weighted graphs

in Subsection 3.2 and list some results by Füredi and Maleki [4] that involve weighted graphs. An important tool in the proof of the middle range is the so-called Exchange Lemma, Lemma 13. We prove Lemma 13 and explain its importance in Subsection 3.3. Our last tool is the notion of *compressed graphs* which is a class of graphs with somewhat restrictive structure. In Subsection 3.4, we give our definition of a compressed graph and prove Lemma 16, which shows that, in order to prove Theorem 3, it suffices to prove it for compressed graphs.

3.1 Notation

The following notation is standard. Write $|G|$ for the order of a graph G and $e(G)$ for the number of edges in G. We denote the degree of a vertex u of G by $\deg_G(u)$, or $\deg(u)$ if G is clear from the context. Given a set U of vertices of G, we denote by $G[U]$ the graph induced by G on U.

We now turn to notation that is more specific to our context. An edge $e \in E(G)$ is called *triangular* if it is an edge of at least one triangle in G. Similarly, we say that e is *non-triangular* if it is not an edge of any triangle. We denote by $t(G)$ the number of non-triangular edges of G.

Given a vertex u, a vertex v is a *triangular neighbour* of u, if uv is a triangular edge. Similarly, the *triangular neighbourhood* of u is the set of triangular neighbours of u, and the *triangular degree* of u is the number of triangular edges adjacent to u. The notions of a *non-triangular neighbour*, *non-triangular neighbourhood* and *non-triangular degree* are defined similarly. We denote the non-triangular degree of u in G by $\deg_{\text{Non-}\Delta}(u)$. A vertex u is called *triangular* if $\deg_{\text{Non-}\Delta}(u) = 0$, that is, if all edges adjacent to u are triangular.

We say that a set of vertices $U \subseteq V(G)$ is a *set of clones* if any two vertices in U have the same neighbourhood in G. In particular, a set of clones is an independent set. For example, in $G(a, b, c)$ the sets B and C are sets of clones. We remark that the notion of clones will play an important role in the definition of a compressed graph (which is given in Subsection 3.4).

We now introduce the natural notion of an optimal graph.

Definition 7. A graph G on n vertices is called *optimal* if there does not exist a graph H on n vertices such that either $t(H) > t(G)$ and $e(H) \geq e(G)$ or $e(H) > e(G)$ and $t(H) \geq t(G)$.

In other words, G is optimal if it maximises $t(G)$ among graphs with n vertices and at least $e(G)$ edges and, in addition, it maximises $e(G)$ among graphs with n vertices and at least $t(G)$ non-triangular edges.

It clearly suffices to prove the main result, Theorem 3, for optimal graphs. The following observation is a simple property of optimal graphs.

Observation 8. *Let G be an optimal graph and let u, v be vertices of G. Then at least one of $\deg(u) \geq \deg(v) - 1$ and $\deg_{Non\text{-}\Delta}(u) \geq \deg_{Non\text{-}\Delta}(v) - 1$ holds.*

Proof. Suppose that $\deg(u) \leq \deg(v)-2$ and $\deg_{\text{Non-}\Delta}(u) \leq \deg_{\text{Non-}\Delta}(v)-2$. Consider the graph G' obtained by removing the edges incident with u and adding the edges between u and the neighbours of v (do not add the loop uu if u, v are adjacent in G). Then $e(G') \geq e(G)-\deg(u)+\deg(v)-1 > e(G)$ and, similarly, $t(G') > t(G)$, contradicting the assumption that G is optimal. $\square$

We shall use big-O notation extensively throughout this paper, so, for the sake of clarity, we briefly explain how we interpret the symbols O, o and Ω. First of all, we always assume that n is large, so, whenever we write down a statement or an inequality, we only suppose it to hold for sufficiently large n. We write $f(n) = O(g(n))$ if there exists an absolute constant $C > 0$ such that $|f(n)| \leq Cg(n)$. In particular, the expression $f(n) = g(n) + O(h(n))$ consists of the following inequalities: $g(n)-Ch(n) \leq f(n) \leq g(n)+Ch(n)$. Similarly, $f(n) = o(g(n))$ means that $\lim_{n\to\infty} |f(n)|/g(n) = 0$. Finally, we write $f(n) = \Omega(g(n))$ if $f(n) \geq Cg(n)$ for an absolute constant $C > 0$. To ensure that this notation makes sense, we will only write $O(g(n)), o(g(n)), \Omega(g(n))$ for functions $g(n)$ which are positive for sufficiently large n. We remark that $\Omega(g(n))$ always denotes a positive quantity, while $O(g(n))$ and $o(g(n))$ may denote positive and negative quantities.

Throughout this paper, we omit integer parts whenever they do not affect the argument.

3.2 Weighted graphs

Our most basic tool is the concept of a *weighted graph*, which is a graph whose vertices have been assigned non-negative real weights. The *total weight* of a weighted graph G is the sum of the weights of its vertices and is denoted by $|G|$. For technical reasons, throughout this paper we require that the number of vertices of a weighted graph does not exceed its total weight. Equivalently, we require that the average weight of a vertex in a weighted graph is at least 1.

For containment purposes, we identify weighted graphs with their underlying graphs. For instance, given weighted graphs G and H, we say that H is a *weighted subgraph* of G if, as graphs, H is an induced subgraph of G. Note that this definition does not impose any conditions on the weight function of H. In particular, if H is a weighted subgraph of G then the weight in H of a vertex in H may be larger, or smaller, than its weight in G. Similarly, an edge of a weighted graph is *triangular* (*non-triangular*) if it is a triangular (non-triangular) edge of the underlying graph. We remark that, unless explicitly stated otherwise, vertices of zero weight are taken into account when switching to the underlying graph.

Given a weighted graph G with weight function $w : V(G) \to \mathbb{R}^{\geq 0}$ we define $e(G)$ to be the sum of $w(u)w(v)$ over all edges uv of G. Similarly, we define $t(G)$ to be the same sum over the non-triangular edges of G. Note that any graph G can be seen either as a graph or as a weighted graph whose every vertex has weight 1, and the definitions of $|G|, e(G)$ and $t(G)$ are independent of the point of view.

The notions of *degree* and *non-triangular* degree of a vertex may be similarly generalised to weighted graphs. For instance, the degree of a vertex u of a weighted graph is the sum of weights of the neighbours of u. Note that the degree and non-triangular degree of a vertex do not depend on the weight of that vertex itself. We use the notation $\deg(u)$ and $\deg_{\text{Non-}\Delta}(u)$ for the degree and the non-triangular degree of a vertex u in a weighted graph.

We now define *good weighted graphs* (see Figure 4), which are weighted equivalents of the graphs $G(a, b, c)$ (see Figure 3).

Definition 9. We call a weighted graph G *good* if its vertex set can be partitioned into a set K, which induces a clique, and a pair (u, v) of adjacent vertices such that uv is the only non-triangular edge in G.

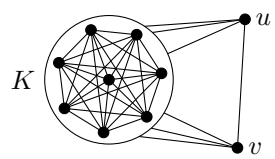

Figure 4: A good weighted graph.

Moreover, if there are no edges between v and K, and if u is adjacent to all vertices in K, then we say that G is a *very good weighted graph*. We remark that, according to this definition, if K consists of a single vertex, then G cannot be very good. On the other hand, if $|K| = 1$, then G is good if and only if uv is the only edge in G, in which case we have $e(G) \leq |G|^2/4$.

Let G be a good graph with $e(G) > |G|^2/4$ and let $\{K, \{u, v\}\}$ be the partition of $V(G)$ as described above. We know that in this case $|K| \geq 2$. Moreover, we may assume that the weight of v does not exceed the weight of u. Since uv is a non-triangular edge, u and v do not have common neighbours in K. Therefore, by removing the edges between v and K and adding all possible edges between u and K, we obtain a very good graph G' such that $|G'| = |G|$, $e(G') \geq e(G)$ and $t(G') = t(G)$.

We observe that, provided that $a \geq 2$, a graph $G(a, b, c)$ can be represented by a very good weighted graph, by replacing the independent parts of sizes b and c by vertices of weight b and c respectively. This is an example of the correspondence between an independent set of clones I and a vertex of weight $|I|$ with the same neighbourhood, which we shall use on multiple occasions. We remark that, in general, good and very good weighted graphs may have non-integer weights.

Motzkin and Straus [10] used weighted graphs to give an alternative proof of Turán's theorem [13]. They pointed out that Turán's theorem for weighted graphs is very easy: given a weighted graph G, there exists a weighted graph H that satisfies $|H| = |G|$ and $e(H) \geq e(G)$, and, as a graph, is a complete subgraph of G. Therefore, among K_{r+1}-free weighted graphs with total weight $\alpha \geq r$, $e(G)$ is maximised when G is a complete graph with r vertices whose every vertex has weight α/r.

If α/r is an integer, then this corresponds to a complete r-partite graph, implying Turán's theorem. However, if α/r is not an integer, then this argument gives only an approximate form of Turán's theorem, and Motzkin and Straus needed an additional argument to recover the full theorem.

Füredi and Maleki [4] modified the aforementioned observation of Motzkin and Straus to also give $t(H) \geq t(G)$ at the cost of making the structure of H more complicated.

Lemma 10 (Füredi and Maleki [4])**.** *Let G be a weighted graph with $t(G) > 0$. Then G contains a weighted subgraph H which is a good weighted graph and satisfies $|H| = |G|$, $e(H) \geq e(G)$ and $t(H) \geq t(G)$.*

We will use both this result and the key observation that leads to its proof. We state and prove this observation next, but we do not present the careful analysis that Motzkin and Straus perform to complete the proof of Lemma 10.

Lemma 11 (Füredi and Maleki [4])**.** *Let G be a weighted graph and suppose that I is an independent set of three vertices. Then there exists a weighted graph H, which can be obtained from G by removing one of the vertices in I and, possibly, changing the weights of the other two vertices in I, such that $|H| = |G|$, $e(H) \geq e(G)$ and $t(H) \geq t(G)$.*

Proof. Denote $I = \{u_1, u_2, u_3\}$, $d_i = \deg(u_i)$ and $t_i = \deg_{\text{Non-}\Delta}(u_i)$. It is not hard to see that there exist reals s_1, s_2, s_3, not all 0, such that $s_1d_1 + s_2d_2 + s_3d_3 \geq 0$, $s_1t_1 + s_2t_2 + s_3t_3 \geq 0$ and $s_1 + s_2 + s_3 = 0$. For real λ we denote by G_λ the weighted graph obtained by adding λs_i to the weight $w(u_i)$ of u_i for each $i \in [3]$; this definition is valid for the values of λ for which $w(u_i)+\lambda s_i \geq 0$ for all $i \in [3]$. Pick $\lambda > 0$ such that $w(u_i) + \lambda s_i \geq 0$ for $i \in [3]$ with equality for at least one value, say 1. Then $|G_\lambda| = |G|$, $e(G_\lambda) \geq e(G)$ and $t(G_\lambda) \geq e(G)$, so the weighted graph $H = G_\lambda \setminus \{u_1\}$ satisfies the requirements of the lemma. $\square$

Füredi and Maleki deduce their main result from Lemma 10. We present their theorem with minor modifications, which make it more suitable for our application.

Corollary 12 (Füredi and Maleki [4])**.** *Let G be a weighted graph G with $|G| = n$. Then there exists a graph $H = G(a, b, c)$ satisfying $|H| = n$, $e(H) \geq e(G)$ and $t(H) \geq t(G) - 5n$.*

Proof. We begin by recalling that, according to our definition of a weighted graph, G has at most n vertices. The parameter $e(G)$ is maximised when the underlying graph of G is complete, in which case $2e(G) = \sum_{u\in V(G)} w(u)(n - w(u)) = n^2 - \sum_{u\in V(G)} w^2(u)$. By the arithmetic-quadratic mean inequality, the right hand side is maximised when $w(u) = 1$ for all u. Therefore, $e(G) \leq \binom{n}{2}$. As a result, we may assume that $t(G) > 5n$ because otherwise the complete graph K_n satisfies the requirements. We may also assume that $e(G) > n^2/4$ because otherwise $G(2, \lfloor n/2\rfloor, \lceil n/2\rceil - 2)$ works.

Let H be a good weighted graph that satisfies $|H| = n$, $e(H) \geq e(G)$ and $t(H) \geq t(G)$, whose existence is ensured by Lemma 10. There exists a partition $\{K, \{u, v\}\}$ of $V(H)$ such that K induces a clique and uv is the only non-triangular edge in H. Denote the sum of weights (in H) of the vertices in K by α and the weights of u and v by β and γ; we may assume that $\beta \geq \gamma$. Trivially, we have $t(H) = \beta\gamma$. Moreover, since no vertex in K is adjacent to both u and v, we also have $e(H) \leq \alpha^2/2 + \alpha\beta + \beta\gamma$.

We now show that for some integers $a, b, c \geq 0$ the graph $G' = G(a, b, c)$ has the desired properties. It is enough to choose a, b, c so that

$$a + b + c = n, \tag{1}$$

$$\binom{a}{2} + (n-b)b \geq \frac{\alpha^2}{2} + (n-\beta)\beta, \tag{2}$$

$$bc \geq \beta\gamma - 5n. \tag{3}$$

Of course, the plan is to set $a \approx \alpha, b \approx \beta, c \approx \gamma$, but there are some tedious details to check. We set $a = \lceil \alpha \rceil + 2$ and, depending on whether $\beta \geq n/2$ or $\beta < n/2$, either $b = \lfloor \beta \rfloor$ or $b = \lceil \beta \rceil$. Finally, we set $c = n - a - b$. Note that from the assumption that $t(G) > 5n$ it follows that $\beta, \gamma > 5$. In particular, since $c \geq \gamma - 4$, c is positive. Now, (1) is immediate from the definition; (3) is immediate from the fact that $b \geq \beta - 1 > 0$ and $c \geq \gamma - 4 > 0$; and the only case when (2) is not immediate is when $(n-1)/2 \leq \beta \leq (n+1)/2$. However, in this case the difference between $(n-\beta)\beta$ and $(n-b)b$ is at most 1, and it is compensated by the difference between $\binom{a}{2}$ and $\alpha^2/2$. □

3.3 Exchange lemma

The following lemma, Lemma 13, will be very useful in the proof of our main result in the middle range. Roughly speaking, it says that there exists a positive number ζ, which we informally call the 'exchange rate', with the following property. For any graph G, not too dense and not too sparse, and any number x, not too big and not too small, we can exchange x edges of G for at least ζx non-triangular edges. That is, there exists a graph H such that $|H| = |G|$, $e(H) \geq e(G) - x$ and $t(H) \geq t(G) + \zeta x$. Similarly, we can exchange x non-triangular edges for at least ζx edges.

This tool is very useful to us, because now we can arrive at a contradiction by finding a graph G whose either parameter $e(G)$ or $t(G)$ is too large, even if the other parameter is slightly smaller than what would normally be needed for a contradiction.

For any positive integer n and real $e \leq \binom{n}{2}$, we denote by $t(n, e)$ the maximum number of non-triangular edges among n-vertex graphs with at least e edges. Note that if $e \leq \lfloor n^2/4 \rfloor$, then $t(n, e) = \lfloor n^2/4 \rfloor$. Moreover, for any n, the function $t(n, e)$ is a non-increasing function of e.

Lemma 13. *For any $\delta > 0$ there exist $\zeta, \varepsilon, C > 0$ and n_0 such that the following holds for any weighted graph G on $n \geq n_0$ vertices and for any real x satisfying $Cn \leq x \leq \varepsilon n^2$.*

1. *If $e(G) \geq e + x$ for some real e satisfying $n^2/4 \leq e \leq (1/2 - \delta)n^2$, then $t(G) \leq t(n, e) - \zeta x$.*

2. *If $t(G) \geq t(n, e) + x$ for some real $e \geq (1/4 + \delta)n^2$, then $e(G) \leq e - \zeta x$.*

Here is a brief overview of the proof of Lemma 13. To prove the first statement, we note that by Lemma 10, we may assume that G is good. We shift the weights of the vertices in G so as to increase $t(G)$ while decreasing $e(G)$ only slightly. An upper bound on $t(G)$ then follows from Corollary 12. The second statement is proved in a similar way.

Proof of Lemma 13. Let $\delta \in (0, 1/10)$. To prove the first statement, suppose that n, e, x satisfy $e \leq (1/2 - \delta)n^2$ and $Cn \leq x \leq \varepsilon n^2$ for constants C and ε that will be determined later. Let G be a weighted graph such that $|G| = n$ and $e(G) \geq e + x$. We note that $t(n, e) \geq \frac{\delta^{3/2}}{2}n^2$. Indeed, the graph $G(a, b, c)$ where $c = \frac{\delta}{2}n$, $b = \sqrt{\delta}\,n$ and $a = n - b - c$ has at least $(1/2 - \delta)n^2$ edges and $\frac{\delta^{3/2}}{2}n^2$ non-triangular edges. By taking ε, ζ to satisfy $\varepsilon\zeta \leq \frac{\delta^{3/2}}{4}$, we may assume that

$$t(G) \geq \frac{\delta^{3/2}}{4}n^2, \tag{4}$$

because otherwise we get $t(G) \leq t(n, e) - \zeta x$ for free. By Lemma 10, we may assume that G is a good weighted graph, so $V(G)$ can be partitioned into a clique K and two adjacent vertices u and v such that uv is the only non-triangular edge. Denote by α the sum of weights of vertices in K and let β and γ be the weights of u and v respectively. By Inequality (4), we have $\beta, \gamma \geq \frac{\delta^{3/2}}{4}n$. Moreover, the removal of the edges spanned by K would make G bipartite, so we have $e(G) \leq n^2/4 + \alpha^2/2 \leq n^2/4 + \alpha n/2$. Recall that $e(G) \geq e + x \geq n^2/4 + x$, and hence $\alpha \geq 2x/n$.

Let G' be a weighted graph obtained by increasing the weight of u by x/n and decreasing the weights of the vertices in K so that the new sum of their weights is $\alpha - x/n$. Trivially, $e(G') \geq e(G) - x \geq e$ and $t(G') = (\beta + x/n)\gamma \geq t(G) + \frac{\delta^{3/2}}{4}x$. Furthermore, it follows from Corollary 12 that $t(G') \leq t(n, e) + 5n$, and hence $t(G) \leq t(n, e) + 5n - \frac{\delta^{3/2}}{4}x \leq t(n, e) - \left(\frac{\delta^{3/2}}{4} - \frac{5}{C}\right)x$. By taking C large and ζ small with respect to δ, we can ensure that $t(G) \leq t(n, e) - \zeta x$.

To prove the second statement, suppose that n, e, x satisfy $e \geq (1/4 + \delta)n^2$ and $10n < x \leq \varepsilon n^2$ for a sufficiently small constant $\varepsilon > 0$. Let G be a weighted graph such that $|G| = n$ and $t(G) \geq t(n, e) + x$. Note that by taking ε, ζ to satisfy $\varepsilon\zeta \leq \delta/2$, we may assume that

$$e(G) \geq \left(\frac{1}{4} + \frac{\delta}{2}\right)n^2, \tag{5}$$

because otherwise we can conclude immediately that $e(G) \leq e - \zeta x$. Furthermore, by Lemma 10, we may assume that G is a good weighted graph. In fact, since $e(G) > n^2/4$, we may assume that G is very good. Let $\{K, \{u, v\}\}$ be a partition of $V(G)$ into a clique K and two vertices u, v such that uv is the only non-triangular edge in G, u is adjacent to all vertices of K and there are no edges between v and K. Moreover, let α be the total weight of vertices in K and let β and γ be the

weights of u and v. As before, it follows from Inequality (5) that $\alpha \geq \sqrt{\delta}\, n$. Moreover, K contains at least two vertices, so in particular a vertex $w \in K$ whose weight does not exceed $\alpha/2$. Let G' be the weighted graph obtained by reducing the weight of v by $x/2n$ (note that $\beta\gamma = t(G) \geq x$, so $\gamma \geq x/n$) and increasing the weight of w by the same amount. Then, since $x > 10n$,

$$t(G') = \beta(\gamma - x/2n) \geq t(G) - x/2 \geq t(n,e) + x/2 > t(n,e) + 5n. \tag{6}$$

Furthermore, since $\alpha \geq \sqrt{\delta}\, n$,

$$e(G') \geq e(G) + \frac{x}{2n} \cdot \frac{\alpha}{2} \geq e(G) + \frac{\sqrt{\delta}}{4} x.$$

By Corollary 12 and Inequality (6), $e(G') < e$, because otherwise there exists a graph H with n vertices, at least e edges and more than $t(n,e)$ non-triangular edges, which contradicts the definition of $t(n,e)$. Thus, $e(G) \leq e - \zeta x$ for any $\zeta \leq \frac{\sqrt{\delta}}{4}$. □

3.4 Compressed graphs

We now present the notion of *compressed* graphs. Many proofs of Turán's theorem, including the one given by Motzkin and Straus [10], first show that, among K_{r+1}-free graphs on a given number of vertices, the greatest number of edges is achieved by a complete r-partitite graph. As a result, it is enough to solve the problem for complete r-partite graphs. In this paper the class of compressed graphs will play a role similar to that of complete r-partite graphs in the proof of Motzkin and Straus. Compressed graphs have fairly simple structure (though not quite as simple as complete r-partite graphs) and we shall see from Lemma 16 that it suffices to prove Theorem 3 for compressed graphs.

In the following definition, as well as the rest of the paper, the logarithm is taken in base 2.

Definition 14. A graph G on n vertices is called *compressed* if the following assertions hold.

1. Any independent set $I \subset V(G)$ can be partitioned into at most $3 \log n$ sets of clones. Moreover, this can be done in such a way that at most four of the sets of clones into which I is partitioned have size larger than $3n^{1/3}$.

2. The set of triangular vertices in G, which we denote by U, induces a clique in G. Furthermore, the vertices of U all have the same neighbourhood outside of U.

To demonstrate how compressed graphs may be of use to us, we mention the following observation.

Observation 15. *Let G be a compressed graph on n vertices and let I be an independent set of size at least $45n^{1/3} \log n$. Then I contains a set of clones of size at least $|I|/5$.*

Indeed, let m be the size of the largest set of clones in I. Then Condition 1 of Definition 14 implies that $|I| \leq 4m + 9n^{1/3} \log n \leq 4m + |I|/5$, so $m \geq |I|/5$.

The following lemma shows that, for the purpose of proving Theorem 3, we may assume without loss of generality that the given graph is compressed.

Lemma 16. *Let G be a graph on n vertices. Then there exists a compressed graph H such that $|H| = n$, $e(H) \geq e(G)$ and $t(H) \geq t(G)$.*

Proof. Given a graph G on n vertices, we let H be a weighted graph with the following properties.

- $|H| = n$, $e(H) \geq e(G)$ and $t(H) \geq t(G)$.
- All vertices of H have integer weights.
- The number of vertices of H is minimal under the first two conditions.
- The number of vertices of weight at least $3n^{1/3}$ is minimal under the first three conditions.

We shall show that the graph, obtained by replacing each vertex of H by a set of clones of size equal to the weight of the vertex, is compressed. To that end, we show that H has no independent set of size larger than $3 \log n$, and that the vertices with weight larger than $3n^{1/3}$ do not contain an independent set of size at least five.

We first show that every independent set of H contains at most $3 \log n$ vertices. Suppose to the contrary that H contains an independent set I of size $m \geq 3 \log n$. For any set $A \subseteq I$ we denote $S_A = \sum_{x \in A} \deg(x)$ and $T_A = \sum_{x \in A} \deg_{\text{Non-}\Delta}(x)$. Trivially, $S_A \leq n^2$ for every $A \subseteq I$. Since $\binom{m}{m/2} \geq 2^m/\sqrt{2m} \geq n^3/\sqrt{2n} > n^2$, it follows from considering sets of size $m/2$ that there exist distinct sets $A, B \subseteq I$ such that $|A| = |B|$ and $S_A = S_B$. By replacing A and B by $A \setminus B$ and $B \setminus A$, we may assume that $A \cap B = \emptyset$. Also, without loss of generality, $T_A \geq T_B$.

Let w be the minimum weight of a vertex in B. Consider the weighted graph H', obtained by increasing the weight of each vertex in A by w and decreasing the weight of each vertex in B by w (and removing vertices whose weight becomes 0). Then $|H'| = |H|$, $e(H') = e(H) + w(S_A - S_B) = e(H)$, $t(H') = t(H) + w(T_A - T_B) \geq t(H)$ and the number of vertices in H' is smaller than the number of vertices in H, contradicting the choice of H. It follows that every independent set of H contains at most $3 \log n$ vertices.

We now show that, given an independent set of five vertices $\{u_1, \ldots, u_5\}$ in H, at least one of the vertices u_i has weight at most $3n^{1/3}$. Indeed, suppose that the weight of each of the vertices exceeds $3n^{1/3}$. For any quintuple of non-negative integers $k = (k_1, \ldots, k_5)$, we denote $S_k = k_1 \deg(u_1) + \ldots + k_5 \deg(u_5)$ and $T_k = k_1 \deg_{\text{Non-}\Delta}(u_1) + \ldots + k_5 \deg_{\text{Non-}\Delta}(u_5)$. Consider only the quintuples k that satisfy $k_1 + \ldots + k_5 = 3n^{1/3}$: there are $\binom{3n^{1/3}+4}{4} \geq \frac{81}{24} n^{4/3}$ such quintuples and for each of

them we have $S_k \leq 3n^{4/3}$. Thus, there exist distinct quintuples k and l, each of whose coordinates are non-negative integers summing to $3n^{1/3}$, such that $S_k = S_l$. Without loss of generality, we may assume that $T_k \geq T_l$.

Consider the weighted graph H', obtained by repeatedly adding $k_i - l_i$ to the weight of each vertex u_i, as long as all weights remain non-negative (note that this process will end because $k_i < l_i$ for some $i \in [5]$). The resulting weighted graph H' has the same number of vertices as H and satisfies $|H'| = |H|$, $e(H') = e(H)$ and $t(H') \geq t(H)$. Furthermore, since $|k_i - l_i| \leq 3n^{1/3}$, for some $i \in [5]$ the weight of u_i in H' is smaller than $3n^{1/3}$. In particular, H' has fewer vertices with weight at least $3n^{1/3}$ than H. This is, again, a contradiction to the choice of H. It follows that every independent set in H has at most four vertices with weight at least $3n^{1/3}$.

Recall that H has integer weights, so we may view it as a graph where a vertex of weight w represents a set of clones of size w. The graph H satisfies Condition 1 of Definition 14. Denote by U the set of triangular vertices in H. Adding all edges missing from $H[U]$ would not form triangles with edges that were previously non-triangular, so we may assume that U induces a clique in H. Let $u \in U$ be a vertex of maximum degree in H. For every $v \in U \setminus \{u\}$, we remove the edges between v and $V(H) \setminus U$ and add the edges between v and the neighbourhood of u in $V(H) \setminus U$. This process does not decrease the total number of edges and, moreover, all edges that were previously non-triangular remain non-triangular. We denote the resulting graph by H' and note that it satisfies Condition 2. Furthermore, if $I \subset V(H) \setminus U$ and $v \in U$ are such that $I \cup \{v\}$ is an independent set in H', then $I \cup \{u\}$ is an independent set in H. Therefore, H' retains Condition 1, and hence H' is compressed. □

4 Almost bipartite

In this section we prove Theorem 4.

Theorem 4. *There exist n_0 and $\delta > 0$ such that the following holds. Let G be a graph with $n \geq n_0$ vertices and e edges, where $n^2/4 \leq e \leq (1/4 + \delta)n^2$. Then there exists a graph $H = G(a, b, c)$ such that $|H| = n$, $e(H) \geq e$ and $t(H) \geq t(G)$.*

Once again, some of the statements and inequalities that we write down only hold for sufficiently large n. Whenever this happens, we assume that n is, indeed, large enough to satisfy them.

Throughout this section we assume that G is a graph with n vertices and $e = (1/4 + \varepsilon)n^2$ edges, where $0 < \varepsilon \leq \delta$ for a small positive constant δ which we will (implicitly) determine later. Moreover, we assume that G is optimal (this means that increasing the number of edges reduces the number of non-triangular edges and vice versa, see Definition 7) and compressed (see Definition 14). In fact, for this proof we only need Condition 2 of Definition 14, which is much simpler than Condition 1.

However, to be able to make the assumptions described above, we have to deal with a small technicality regarding the condition $e(G) \leq (1/4+\delta)n^2$. Indeed, it is true that given G it is always possible to find an optimal and compressed graph G' satisfying $|G'| = n, e(G') \geq e$ and $t(G') \geq t(G)$, but we cannot guarantee that $e(G') \leq (1/4+\delta)n^2$ holds. To deal with this issue, we use the Exchange Lemma (Lemma 13). Indeed, if $e(G') \geq e + \Omega(n^2)$, then Lemma 13 implies that $t(G) \leq t(G') \leq t(n,e) - \Omega(n^2)$. If this happens, then we take a graph H with n vertices and at least e edges, satisfying $t(H) = t(n,e)$. By Corollary 12 there exists a graph $H' = G(a,b,c)$ such that $|H'| = n, e(H') \geq e$ and $t(H') \geq t(n,e) - 5n \geq t(G)$, so we are done in this case. Therefore, we may assume that $e(G') \leq e + o(n^2)$, and so $e(G') \leq (1/4+\delta)n^2$ holds for a relaxed value of δ.

To get a rough idea about how large $t(G)$ is, we derive the following lower bound. Consider the graph $G(a,b,c)$ where $a = \lceil\sqrt{2\varepsilon}\,n\rceil + 1$, $b = \lceil n/2\rceil$ and $c = n - a - b = \lfloor n/2\rfloor - \lceil\sqrt{2\varepsilon}\,n\rceil - 1$. Then $e(G(a,b,c)) = \binom{a}{2} + (n-b)b \geq (1/4+\varepsilon)n^2$ and $t(G(a,b,c)) = bc \geq \left(1/4 - \sqrt{\varepsilon/2} - O(1/n)\right)n^2$. Since G is optimal, it follows that

$$t(G) \geq \left(\frac{1}{4} - \sqrt{\frac{\varepsilon}{2}} - O\left(\frac{1}{n}\right)\right)n^2. \tag{7}$$

Moreover, we have $e \geq \lfloor n^2/4\rfloor + 1$, so in fact $\varepsilon n^2 \geq 1/2$ and therefore $1/n = O(\sqrt{\varepsilon})$. It follows that

$$t(G) \geq \left(\frac{1}{4} - O\left(\sqrt{\varepsilon}\right)\right)n^2. \tag{8}$$

We would like to make a brief comment regarding the use of big-O notation in Inequalities (7),(8) and other similar inequalities. According to our definition, $O(g(n))$ stands for a positive or negative quantity, so $-O(g(n))$ is exactly the same as $+O(g(n))$. We usually choose the sign before the big-O which looks more natural. However, we do not assume that $-O(g(n))$ is necessarily negative of $+O(g(n))$ is necessarily positive.

We divide the proof of Theorem 4 into four parts, represented by the following four propositions. In the first of these propositions we prove that G has the following structure (see Figure 5), which resembles a graph $G(a,b,c)$.

Proposition 17. *There exists a partition $\{A,B,C,D\}$ of $V(G)$ satisfying the following assertions.*

1. *All possible edges between B and C are present in G and are non-triangular. Moreover, $|B|,|C| \geq (1/2 - O(\sqrt{\varepsilon}))n$. In particular, B and C are independent sets.*
2. *There are no edges between A and C nor between B and D.*
3. *The induced subgraphs $G[A]$ and $G[D]$ do not have isolated vertices.*
4. *Every vertex in $A \cup D$ is incident with at most $O(\sqrt{\varepsilon}\,n)$ non-triangular edges of G. Moreover, the sets A and D do not span non-triangular edges (but there may be non-triangular edges between A and D).*

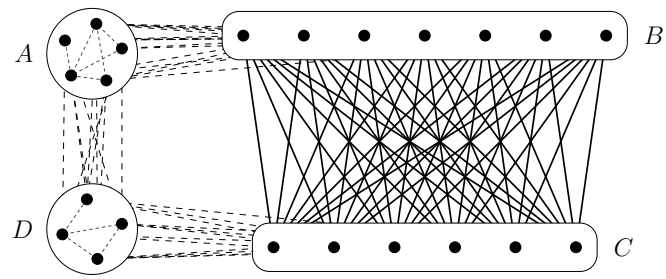

Figure 5: The partition $\{A, B, C, D\}$.

Here the proof of Theorem 4 splits into two cases: $\varepsilon \leq \kappa/n$ and $\varepsilon \geq \kappa/n$, where κ is a small absolute positive constant that will be (implicitly) determined later. If ε is small, then we complete the proof directly.

Proposition 18. *There exists a constant $\kappa > 0$ with the property that if $\varepsilon \leq \kappa/n$, then $G \cong G(a, b, c)$ for some a, b, c.*

If ε is large, then we first obtain sharp estimates for the sizes of the sets A, B, C, D.

Proposition 19. *Let G and A, B, C, D satisfy the conclusions of Proposition 17 and suppose that $|B| \geq |C|$ and that $\varepsilon \geq \kappa/n$ for some constant $\kappa > 0$. Then*

$$\begin{aligned} |A \cup D| &= \left(\sqrt{2\varepsilon} + O_\kappa(\varepsilon)\right) n, \\ |B| &= \left(\frac{1}{2} - O_\kappa(\varepsilon^{3/4})\right) n, \\ |C| &= \left(\frac{1}{2} - \sqrt{2\varepsilon} + O_\kappa(\varepsilon^{3/4})\right) n. \end{aligned}$$

Here $f(n) = O_\kappa(g(n))$ means that there exists a constant $C_\kappa > 0$, which depends on κ, such that $|f(n)| \leq C_\kappa g(n)$.

We can now complete the proof of Theorem 4 in the case where ε is large.

Proposition 20. *Let $\kappa > 0$ be a constant. If $\varepsilon \geq \kappa/n$, then $G \cong G(a, b, c)$ for some a, b, c.*

Proposition 20 is a typical example of a statement that holds for sufficiently large n.

Proof of Theorem 4. Theorem 4 immediately follows from Propositions 18 and 20. □

The rest of this section is devoted to the proofs of Propositions 17 to 20, which are presented in separate subsections.

4.1 Structure of an optimal graph

In this subsection, we prove Proposition 17 (see also Figure 5). Recall that G is a fixed optimal graph with n vertices and $e = (1/4 + \varepsilon)n^2$ edges, where $0 < \varepsilon \leq \delta$ for a fixed small constant $\delta > 0$.

Proposition 17. *There exists a partition $\{A, B, C, D\}$ of $V(G)$ satisfying the following assertions.*

1. *All possible edges between B and C are present in G and are non-triangular. Moreover, $|B|, |C| \geq (1/2 - O(\sqrt{\varepsilon}))n$. In particular, B and C are independent sets.*
2. *There are no edges between A and C nor between B and D.*
3. *The induced subgraphs $G[A]$ and $G[D]$ do not have isolated vertices.*
4. *Every vertex in $A \cup D$ is incident with at most $O(\sqrt{\varepsilon}n)$ non-triangular edges of G. Moreover, the sets A and D do not span non-triangular edges (but there may be non-triangular edges between A and D).*

The first assertion follows fairly easily from the fact that the number of non-triangular edges is almost $n^2/4$. To complete the proof we use basic properties of optimal graphs.

Proof of Proposition 17. Let H be the spanning subgraph of G whose edges are the non-triangular edges of G. We note that H is a triangle-free graph with close to $n^2/4$ edges, which implies that H is close to being a complete bipartite graph. This enables us to find independent (with respect to G) sets U and W of size almost $n/2$ each, such that H contains almost all of the possible edges between them. This idea is rigorously implemented in the following claim.

Claim 21. *There exist disjoint non-empty independent sets $U, W \subseteq V(G)$ such that every vertex in U has at least $(1/2 - O(\varepsilon^{1/4}))n$ non-triangular neighbours in W and vice versa. In particular, $|U|, |W| \geq (1/2 - O(\varepsilon^{1/4}))n$.*

Proof. Inequality (8) states that $e(H) = t(G) \geq (1/4 - c\sqrt{\varepsilon})n^2$ for some absolute constant $c > 0$. From this we deduce that, writing $d = \sqrt{c}$, there are at most $2d\varepsilon^{1/4}n$ vertices in H of degree smaller than $(1/2 - d\varepsilon^{1/4})n$. Indeed, suppose that we can find a set S consisting of exactly $2d\varepsilon^{1/4}n$ vertices of degree smaller than $(1/2 - d\varepsilon^{1/4})n$ in H. Since H is triangle-free, $e(H \setminus S) \leq (n - |S|)^2/4$. Hence,

$$\begin{aligned} e(H) &< \frac{(n-|S|)^2}{4} + |S|\left(\frac{1}{2} - d\varepsilon^{1/4}\right)n \\ &= \frac{n^2}{4} - |S|\left(d\varepsilon^{1/4}n - \frac{|S|}{4}\right) \\ &= \left(\frac{1}{4} - d^2\sqrt{\varepsilon}\right)n^2 \\ &= \left(\frac{1}{4} - c\sqrt{\varepsilon}\right)n^2, \end{aligned}$$

a contradiction.

Let $u \in V(G)$ be any vertex with $\deg_H(u) \geq (1/2 - d\varepsilon^{1/4})n$. Denote by U the set of vertices in $N_H(u)$ that have at least $(1/2 - d\varepsilon^{1/4})n$ neighbours in H. Since the edges of H are non-triangular in G, it follows that U is independent in G. Moreover, $|U| \geq \deg_H(u) - 2d\varepsilon^{1/4}n \geq (1/2 - O(\varepsilon^{1/4}))n$.

Now let $v \in U$ and denote by W the set of vertices in $N_H(v)$ whose degree in H is at least $(1/2 - d\varepsilon^{1/4})n$. As before, W is independent in G and has size at least $(1/2 - O(\varepsilon^{1/4}))n$. Finally, every vertex in U has at least $(1/2 - d\varepsilon^{1/4})n - (n - |U| - |W|) \geq (1/2 - O(\varepsilon^{1/4}))n$ non-triangular neighbours in W, and vice versa. □

Let U and W be the disjoint independent sets given by Claim 21. The following similar claim allows us to enlarge U and W to obtain sets B and C which will be shown to satisfy the requirements of Proposition 17.

Claim 22. *There exist disjoint independent sets $B, C \subseteq V(G)$, satisfying $U \subseteq B$ and $W \subseteq C$ and $|B \cup C| \geq (1 - O(\sqrt{\varepsilon}))n$, such that every vertex in B has at least $2n/5$ non-triangular neighbours in C and vice versa.*

Proof. We first show that there are at most $O(\sqrt{\varepsilon}n)$ vertices of degree at most $21n/50$ in H (where H, the graph of non-triangular edges of G, was defined in the proof of the previous claim). To this end we recall Inequality (8), which states that $e(H) = t(G) \geq (1/4 - c\sqrt{\varepsilon})n$ for some absolute constant c. Importantly, this constant does not depend on δ, so we may choose δ to satisfy $c\sqrt{\delta} \leq 1/200$. Recall that δ is an upper bound for ε, so we have $c\sqrt{\varepsilon} \leq 1/200$.

Suppose that S is a set consisting of exactly $25c\sqrt{\varepsilon}n$ vertices of degree at most $21n/50$ in H. Then, similarly to the previous claim,

$$\begin{aligned} e(H) &\leq \frac{(n - |S|)^2}{4} + |S|\frac{21n}{50} \\ &= \frac{n^2}{4} - |S|\left(\frac{2n}{25} - \frac{|S|}{4}\right) \\ &< \left(\frac{1}{4} - c\sqrt{\varepsilon}\right)n^2, \end{aligned}$$

a contradiction to the choice of c. Therefore, there are at most $O(\sqrt{\varepsilon}n)$ vertices with degree at most $21n/50$ in H.

Recall that every vertex in U has at least $(1/2 - O(\varepsilon^{1/4}))n \geq 2n/5$ non-triangular neighbours in W and vice versa. Here we implicitly assume that δ is small enough to make this inequality true, and we shall do so throughout this proof.

Denote by X the set of vertices in $V(G) \setminus (U \cup W)$ whose degree in H is at least $21n/50$. We note that no vertex in X has neighbours in both U and W. Indeed, suppose that $v \in X$ is adjacent

to $u \in U$ and $w \in W$. Since v is not adjacent to any non-triangular neighbour of either u or w, it has at most $O(\varepsilon^{1/4}n)$ neighbours in U and at most $O(\varepsilon^{1/4}n)$ neighbours in W, implying that $\deg_H(v) \leq O(\varepsilon^{1/4}n)$, a contradiction to the assumption that $\deg_H(v) \geq 21n/50$.

Let Y be the set of vertices in X that are adjacent to vertices in U and, similarly, let Z be the set of vertices in X that have neighbours in W. Then every vertex in Y has at least $21/50n - O(\varepsilon^{1/4}n) \geq 2n/5$ non-triangular neighbours in U and no neighbours in W. In particular, since $|U| \leq n - |W| < 4n/5$, any two vertices in Y share a non-triangular neighbour in U, and hence Y is an independent set in G. Denote $B = Y \cup W$ and $C = Z \cup U$. Then B and C are independent sets such that every vertex in B has at least $2n/5$ non-triangular neighbours in C, and vice versa. Furthermore, $V(G) \setminus (B \cup C)$ is the set of vertices with fewer than $21n/50$ neighbours in H, so $|V(G) \setminus (B \cup C)| = O(\sqrt{\varepsilon}n)$, finishing the proof of Claim 22. $\square$

We can now finish the proof of Proposition 17. Let B and C be as in Claim 22. Since every vertex in $B \cup C$ has at least $2n/5$ non-triangular neighbours, it follows from Observation 8 and the assumption that G is optimal that every vertex in G has degree at least $2n/5 - 1$. We conclude (similarly to the proof of Claim 22) that no vertex in G has neighbours in both B and C. Indeed, suppose that some $v \in V(G)$ is adjacent to some $u \in B$ and $w \in C$. Since u has at least $2n/5$ non-triangular neighbours in C, v is adjacent to at most $|C| - 2n/5$ vertices in C and, similarly, to at most $|B| - 2n/5$ vertices in B. It follows that v has degree at most $n/5$, a contradiction.

Since no vertex in G is adjacent to a vertex in B and a vertex in C, we may add all missing edges between B and C without creating new triangles. However, G is an optimal graph, so in fact all edges between B and C are present in G. Again, since vertices in B and C do not have common neighbours, all edges between B and C are non-triangular.

We may assume that $|B| \geq |C|$. Then $|B| \geq (1/2 - O(\sqrt{\varepsilon}))n$ and hence every vertex in C has non-triangular degree at least $(1/2 - O(\sqrt{\varepsilon}))n$. Again, by Observation 8, every vertex in G has degree at least $(1/2 - O(\sqrt{\varepsilon}))n$. Since B is an independent set, it follows that $n - |B| \geq (1/2 - O(\sqrt{\varepsilon}))n$. Therefore $|C| = |B \cup C| - |B| \geq (1/2 - O(\sqrt{\varepsilon}))n$.

We are now done with the first assertion of Proposition 17, and the remaining ones follow easily. Let A be the set of vertices outside of $B \cup C$ that are adjacent to a vertex in B and, similarly, let D be the set of vertices outside of $B \cup C$ that have a neighbour in C. Then $\{A, B, C, D\}$ forms a partition of G, because a vertex without neighbours in $B \cup C$ would have too small a degree. This establishes the second assertion.

To prove the third assertion, we may assume that every vertex in A has a neighbour in A: if some $u \in A$ has no neighbours in A, then we may add all edges between u and the vertices in B without creating new triangles and then reassign u to C. Similarly, we may assume that every vertex in D has a neighbour in D.

By inspecting the degrees, any two vertices in A have a common neighbour in B. Therefore, there cannot be any non-triangular edges with both ends in A or, similarly, with both ends in D. It remains to check that every vertex in $A \cup D$ is incident with at most $O(\sqrt{\varepsilon}n)$ non-triangular edges. Let $u \in A$ and let $v \in A$ be a neighbour of u. Since u and v have neighbours only in $A \cup D \cup B$ and the degree of v is at least $(1/2 - O(\sqrt{\varepsilon}))n$, it follows that u has at most $|A \cup D \cup B| - (1/2 - O(\sqrt{\varepsilon}))n = O(\sqrt{\varepsilon}n)$ non-triangular neighbours. The same holds for any vertex in D. This establishes the fourth assertion and completes the proof of Proposition 17. $\square$

4.2 Completing the proof if ε is small

We now prove Proposition 18, which completes the proof of Theorem 4 in the case where ε is small.

Proposition 18. *There exists a constant $\kappa > 0$ with the property that if $\varepsilon \leq \kappa/n$, then $G \cong G(a, b, c)$ for some a, b, c.*

Proof. It follows from the assumptions on the sets A, B, C, D that $|A \cup D| = O(\sqrt{\varepsilon}n)$ and that each vertex in $A \cup D$ is incident with at most $O(\sqrt{\varepsilon}n)$ non-triangular edges. Therefore the number of non-triangular edges with an end in $A \cup D$ is $O(\varepsilon n^2) = O(\kappa n)$. We show that, in fact, there are no such edges.

Suppose that uv is a non-triangular edge with $u \in A \cup D$. Without loss of generality, we may assume that $u \in A$ and $v \in B \cup D$. Observe that the neighbours of u are not adjacent to v. Let G' be the graph obtained by adding the edges between v and the neighbours of u in A, removing the edges between u and $A \setminus \{u\}$ and also adding all missing edges between u and B. Then $e(G') \geq e(G)$ and $t(G') \geq t(G) + |B| - O(\kappa n) > t(G)$, where the last inequality holds provided that we choose κ small enough. However, this contradicts the optimality of G, so there cannot be such an edge uv.

It is now easy to finish the proof. By what we have just proved, all the missing edges with both ends in $A \cup D$ may be added without causing a non-triangular edge to become triangular, and hence, since G is optimal, $G[A \cup D]$ is a clique. Similarly, all possible edges between A and B and between D and C are present in G. We may assume that $|B| \geq |C|$. Remove the edges between D and C and add all possible edges between D and B. The resulting graph G' is isomorphic to $G(|A \cup D|, |B|, |C|)$ and satisfies $|G'| = n$, $e(G') \geq e(G)$ and $t(G') \geq t(G)$. However, G is optimal, so we must have $e(G') = e(G)$ and $t(G') = t(G)$. Therefore, it must be the case that $D = \emptyset$ or $|B| = |C|$. If D is empty, then $G = G'$ and we are done. Let us suppose that $|B| = |C|$. Since $e(G') > n^2/4$, G' is not bipartite, and hence $|A \cup D| \geq 2$. Take any vertex $w \in A \cup D$. If we remove all edges between w and B, but add all possible edges between w and C, then we obtain a new graph which has the same number of edges, but more non-triangular edges than G. However, this contradicts the assumption that G is optimal. Therefore, it must be the case that $|B| > |C|$, and so we are done. $\square$

4.3 Sizes of A, B, C, D

In this subsection we prepare for the proof of Theorem 4 in the case where ε is large. In particular, we obtain good bounds for the sizes of the sets $A \cup D$, B and C.

Proposition 19. *Let G and A, B, C, D satisfy the conclusions of Proposition 17 and suppose that $|B| \geq |C|$ and that $\varepsilon \geq \kappa/n$ for some constant $\kappa > 0$. Then*

$$\begin{aligned} |A \cup D| &= \left(\sqrt{2\varepsilon} + O_\kappa(\varepsilon)\right) n, \\ |B| &= \left(\frac{1}{2} - O_\kappa(\varepsilon^{3/4})\right) n, \\ |C| &= \left(\frac{1}{2} - \sqrt{2\varepsilon} + O_\kappa(\varepsilon^{3/4})\right) n. \end{aligned}$$

The proof is just a technical calculation, in which the main tool is the lower bound on $t(G)$ given by Inequality (7).

Proof. Denote $a = |A \cup D|$, $b = |B|$ and $c = |C|$ and write

$$\begin{aligned} a &= \left(\sqrt{2\varepsilon} + \alpha\right) n, \\ b &= \left(\frac{1}{2} - \beta\right) n, \\ c &= \left(\frac{1}{2} - \sqrt{2\varepsilon} + \beta - \alpha\right) n, \end{aligned}$$

where the quantities α and β are defined by these identities. We cannot assume that α and β are positive, but we have $-\sqrt{2\varepsilon} \leq \alpha \leq O(\sqrt{\varepsilon})$, where the second inequality comes from Proposition 17. Since there are at most $O(\varepsilon n^2)$ non-triangular edges with an end in $A \cup D$, we have

$$\begin{aligned} t(G) &\leq bc + O(\varepsilon n^2) \\ &\leq \frac{(n-a)^2}{4} + O(\varepsilon n^2) \\ &\leq \frac{n^2}{4} - \frac{an}{2} + O(\varepsilon n^2). \end{aligned}$$

Combining this with Inequality (7), which states that $t(G) \geq \left(1/4 - \sqrt{\varepsilon/2} - O(1/n)\right) n^2$, we get

$$\frac{1}{4} - \sqrt{\frac{\varepsilon}{2}} - O_\kappa(\varepsilon) \;\leq\; \frac{t(G)}{n^2} \;\leq\; \frac{1}{4} - \frac{(\sqrt{2\varepsilon} + \alpha)}{2} + O(\varepsilon).$$

Therefore, $\alpha \leq O_\kappa(\varepsilon)$. Using the fact that $b \geq c$ and that any vertex in $A \cup D$ sends edges to only

one of B and C, we obtain the following upper bound on the number of edges in G:

$$e(G) \leq b(n-b) + \frac{a^2}{2}.$$

Combining this with the definition $e(G) = (1/4+\varepsilon)n^2$, we get

$$\begin{aligned}\frac{1}{4}+\varepsilon &\leq \left(\frac{1}{2}-\beta\right)\left(\frac{1}{2}+\beta\right) + \frac{\left(\sqrt{2\varepsilon}+\alpha\right)^2}{2} \\ &= \frac{1}{4} - \beta^2 + \varepsilon + \alpha\left(\sqrt{2\varepsilon}+\frac{\alpha}{2}\right).\end{aligned}$$

It follows that $\beta^2 \leq \alpha(\sqrt{2\varepsilon}+\alpha/2)$. In particular, $\alpha \geq 0$ and $\beta = O_\kappa(\varepsilon^{3/4})$, implying the assertions of Proposition 19. □

4.4 Completing the proof if ε is large

We are now able to complete the proof of Theorem 4 under the assumption that $\varepsilon \geq \kappa/n$ for some constant $\kappa > 0$. Here we will use the assumption that G is a compressed graph.

Proposition 20. *Let $\kappa > 0$ be a constant. If $\varepsilon \geq \kappa/n$, then $G \cong G(a,b,c)$ for some a,b,c.*

The proof consists of two stages. In the first stage we use the bounds from Proposition 19 to conclude that D is very small and that very few vertices in A are incident with non-triangular edges. In the second stage we show that if D is non-empty or if there exists a vertex in A with a non-triangular neighbour, then G can be manipulated to obtain a graph with more edges and more non-triangular edges, contradicting the assumption that G is optimal. It follows that G is isomorphic to a graph $G(a,b,c)$.

Proof of Proposition 20. We start by showing that the edges between $B \cup D$ and $A \cup C$ form an almost complete bipartite subgraph. We shall be using the estimates on the size of the sets $A \cup D$, B and C from Proposition 19. To be able to use Proposition 19, we assume, without loss of generality, that $|B| \geq |C|$. Note that κ is an absolute constant (implicitly determined in Proposition 18). Thus, we may remove the dependence on κ in the estimates of these sizes.

Claim 23. *Every vertex in $B \cup D$ is adjacent to all but $O(\varepsilon^{3/4}n)$ vertices in $A \cup C$. Furthermore, $|D| = O(\varepsilon^{3/4}n)$.*

Proof. The non-triangular degree of any vertex in C is at least $|B|$. Hence, by Observation 8, every vertex in G has degree at least $|B|-1$. The vertices in $B \cup D$ are not adjacent to any vertex in B. Since $|B| \geq (1/2 - O(\varepsilon^{3/4}))n$, it follows that every vertex in $B \cup D$ is adjacent to all but $O(\varepsilon^{3/4}n)$ vertices in $V(G) \setminus B = A \cup C \cup D$. Since there are no edges between B and D, $|D| = O(\varepsilon^{3/4}n)$. □

Denote by T the set of triangular vertices in A (recall that a triangular vertex is incident only with triangular edges) and let $S = A \setminus T$. We show that the vertices in S have few neighbours in A.

Claim 24. *Every vertex in S has $O(\varepsilon^{3/4}n)$ neighbours in A.*

Proof. Let $u \in S$ and let v be a non-triangular neighbour of u. Then $v \in B \cup D$, because there are no edges between A and C, and there are no non-triangular edges with both ends in A. Recall that, by Claim 23, v is adjacent to all but $O(\varepsilon^{3/4}n)$ vertices in A. Since uv is non-triangular, u and v have no common neighbours, implying that u has $O(\varepsilon^{3/4}n)$ neighbours in A. □

We conclude that almost all of the vertices in A are triangular.

Claim 25. $|T| \geq (\sqrt{2\varepsilon} - O(\varepsilon^{3/4}))n$.

Proof. By removing the edges with both ends in A or both ends in D from G, we remain with a bipartite graph, so $(1/4+\varepsilon)n^2 = e(G) \leq n^2/4 + e(G[A]) + e(G[D])$. Since $|D| = O(\varepsilon^{3/4}n)$, we have $e(G[D]) = O(\varepsilon^{3/2}n^2)$, and hence $e(G[A]) \geq (\varepsilon - O(\varepsilon^{3/2}))n^2$.

Claim 24 implies that $e(G[A]) - e(G[T]) \leq O(|S|\varepsilon^{3/4}n) \leq O(|A|\varepsilon^{3/4}n) \leq O(\varepsilon^{5/4}n^2)$, where the rightmost inequality is a consequence of Proposition 19. Therefore, $e(G[T]) \geq (\varepsilon - O(\varepsilon^{5/4}))n^2$, and so $|T| \geq (\sqrt{2\varepsilon} - O(\varepsilon^{3/4}))n$, as required. □

Since G is compressed, T induces a clique and any two vertices in T have the same neighbourhood outside of T. In particular, if a vertex $v \in S$ is adjacent to a vertex in T, then v is adjacent to all vertices in T. However, this cannot happen since, by Claim 24, v has at most $O(\varepsilon^{3/4}n)$ neighbours in A, while, by Claim 25, there are at least $\Omega(\sqrt{\varepsilon}n)$ vertices in T. Therefore, there are no edges between T and S.

In the following claim we deduce that, in fact, all vertices in A are triangular. The key observation is that a pair of adjacent vertices in S can be replaced by one vertex in C and one in T, increasing both the number of edges and the number of non-triangular edges.

Claim 26. *The set S is empty.*

Proof. Suppose that S contains a vertex u. By Proposition 17, u has a neighbour $v \in A$. Since there are no edges between T and S, we conclude that $v \in S$. In particular, u and v have no neighbours in T. Now let H be the graph obtained from G by removing the vertices u and v and adding new vertices x and y, where x is joined by edges to all of B and y is joined to all of $B \cup T$. It follows from Claims 23 to 25 that $e(H) \geq e(G) - O(\varepsilon^{3/4}n) + (\sqrt{2\varepsilon} - O(\varepsilon^{3/4}))n > e(G)$. Recall that, by Proposition 17, the non-triangular degree of any vertex in A is at most $O(\sqrt{\varepsilon}n)$, implying that $t(H) \geq t(G) - O(\sqrt{\varepsilon}n) + |B| > t(G)$. Therefore, H has more edges and more non-triangular edges than G, contradicting the optimality of G. Thus, S is empty. □

Similarly, we prove that D is empty. The trick here is to replace two adjacent vertices in D by one vertex in C and one in A.

Claim 27. *The set D is empty.*

Proof. Suppose that D is non-empty, so we may pick adjacent vertices $u, v \in D$. Consider the graph H, obtained by removing the vertices u and v and adding new vertices x and y with x joined to all of B and y joined to all of $A \cup B$. It follows from the bounds given by Proposition 19 and Claim 23 that $e(H) \geq e(G) + (\sqrt{2\varepsilon} - O(\varepsilon^{3/4}))n > e(G)$. Moreover, since $A = T$ is a clique of triangular vertices, the addition of x and y does not destroy any non-triangular edges in $G \setminus \{u, v\}$. Since u and v have at most $O(\sqrt{\varepsilon}n)$ non-triangular neighbours, we have $t(H) \geq t(G) + (1/2 - O(\sqrt{\varepsilon}))n > t(G)$, contradicting the assumption that G is optimal. □

Now the proof of Proposition 20 is complete. Indeed, we know from Claim 26 that $A = T$. This means that A induces a clique and that every vertex in A is adjacent to every vertex in B. Therefore, $G = G(|A|, |B|, |C|)$. □

5 Middle range

In this section we prove Theorem 5, in which we consider the case where the graph is neither close to being complete nor close to being complete bipartite. Out of the three ranges, the middle range turns out to be the hardest to prove. One of the main difficulties that arises here is that, unlike in the other two ranges, we cannot directly conclude that the graph G has structure similar to that of $G(a, b, c)$.

Theorem 5. *For every $\delta > 0$ there exists n_0 such that the following holds. Let G be a graph with $n \geq n_0$ vertices and e edges, where $(1/4 + \delta)n^2 \leq e \leq (1/2 - \delta)n^2$. Then there exists a graph $H = G(a, b, c)$ such that $|H| = n$, $e(H) \geq e$ and $t(H) \geq t(G)$.*

Fix $\delta > 0$. Throughout this section we assume that G is a compressed and optimal graph with n vertices and e edges, where $(1/4 + \delta)n^2 \leq e \leq (1/2 - \delta)n^2$. As in the rest of the paper, the statements that we write down hold for sufficiently large n. Moreover, since δ is fixed, the constants implied by big-O notation may depend on δ.

We split the proof of Theorem 5 into four stages, as described by the four following propositions. In the first stage we show that G has many triangular vertices (that is, vertices that are incident only with triangular edges).

Proposition 28. *G has $\Omega(n)$ triangular vertices.*

In the second stage we conclude that G admits the following structure (see Figure 6). Although Proposition 29 gives much less information than Proposition 17 from Section 4, it still shows that G vaguely resembles a graph $G(a, b, c)$.

Proposition 29. *There exists a partition $\{A, B, C\}$ of $V(G)$ such that all parts have size $\Omega(n)$ and the following properties are satisfied.*

1. *A is the set of triangular vertices in G, it spans a clique and its vertices are adjacent to all of B and none of C.*
2. *B may be partitioned into $O(1)$ sets of clones and a remainder of size $O(\sqrt{n}\log n)$.*
3. *C may be partitioned into $O(1)$ sets of clones, each having $\Omega(n)$ non-triangular neighbours in B, and a remainder of size $O(n^{1/3}\log n)$.*

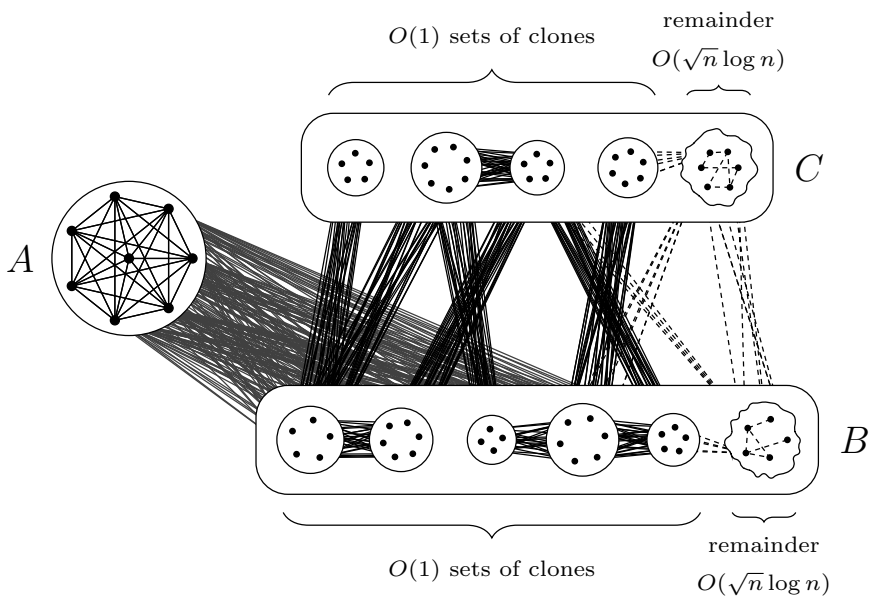

Figure 6: The partition $\{A, B, C\}$

In the third stage we show that the number of edges (and non-triangular edges) in G is close to the number of edges (and non-triangular edges) in $G(|A|, |B|, |C|)$.

Proposition 30. *Let A, B, C be as in Proposition 29 and denote $a = |A|, b = |B|, c = |C|$. Then $e(G) = a^2/2 + ab + bc + O(n^{7/4}\sqrt{\log n})$ and $t(G) = bc + O(n^{7/4}\sqrt{\log n})$.*

In the final fourth stage we complete the proof of Theorem 5.

Proposition 31. *$G \cong G(a, b, c)$ for some a, b, c.*

Proof of Theorem 5. The proof is immediate from Propositions 28 to 31. The only slight technicality is that when we replace a graph with at most $(1/2-\delta)n^2$ edges by an optimal and compressed graph, the number of edges may increase and exceed this bound. However, Lemma 13 implies that this condition is still satisfied for a relaxed value of δ. $\square$

We now turn to the proofs of Propositions 28 to 31. We present them in separate subsections.

5.1 Many triangular vertices

In this subsection we prove Proposition 28.

Proposition 28. *G has $\Omega(n)$ triangular vertices.*

The main ingredients of this proof are a somewhat unexpected application of Lemma 13 and the assumption that G is compressed. First, we conclude from Lemma 10 that G has a large clique. Then, we partition the graph into fairly large independent sets of clones and a very dense part, using the fact that G is compressed. It is then possible to conclude that only few of the vertices of the clique are incident with non-triangular edges.

Proof of Proposition 28. Our first aim is to show that G has a clique of size at least $\Omega(n)$. This can be done fairly easily, as shown in the proof of the following claim.

Claim 32. *G has a clique of size $\Omega(n)$.*

Proof. By Lemma 10, there exists a good weighted subgraph H of G satisfying $|H| = |G| = n$, $e(H) \geq e(G)$, $t(H) \geq t(G)$ (see Definition 9 for the definition of a good weighted graph). Let $\{K, \{u, v\}\}$ be a partition of $V(H)$ into a clique K and an edge uv, which is the only non-triangular edge of H.

Let α be the sum of the weights of vertices in K and let m be the number of vertices in K. Let β and γ be the weights of u and v and suppose that $\beta \geq \gamma$. Note that $\alpha+\beta+\gamma = n$. By the Cauchy-Schwarz inequality, the contribution of the edges in K towards $e(H)$ is maximised if all of these vertices have weight α/m. Therefore this contribution does not exceed $(\alpha/m)^2\binom{m}{2} = (1-1/m)\alpha^2/2$. Moreover, since no vertex is adjacent to both u and v, the contribution of the edges between K and $\{u, v\}$ towards $e(H)$ is maximised when every vertex in K is adjacent to u, but not v. Hence,

$$e(G) \leq e(H) \leq \left(1 - \frac{1}{m}\right)\frac{\alpha^2}{2} + \alpha\beta + \beta\gamma. \tag{9}$$

In particular, since $\beta\gamma \leq n^2/4$, we have $e(G) \leq n^2/4 + \alpha n$. Recall that $e(G) \geq (1/4+\delta)n^2$. It follows that $\alpha \geq \delta n$.

Denote $b = \lceil \beta \rceil$, $c = \lceil \gamma \rceil$ and $a = n - b - c$ and consider the graph $F = G(a, b, c)$. Note that $t(G) \leq t(H) = \beta\gamma \leq bc = t(F)$. Since G is optimal, it follows that $e(G) \geq e(F)$. Therefore,

$$\begin{aligned} e(G) \geq e(F) &= \binom{a}{2} + ab + bc \\ &\geq \frac{(\alpha-2)(\alpha-3)}{2} + (\alpha-2)\beta + \beta\gamma \\ &\geq \frac{\alpha^2}{2} + \alpha\beta + \beta\gamma - 2.5n \\ &= \left(1 - \frac{5n}{\alpha^2}\right)\frac{\alpha^2}{2} + \alpha\beta + \beta\gamma. \end{aligned}$$

Comparing this with (9), we have $m \geq \alpha^2/(5n) \geq \delta^2 n/5$. It follows that G has a clique of size at least $\delta^2 n/5$. $\square$

Recall that G is compressed. Hence, by Observation 15, every independent set of size $5\sqrt{n}$ in G contains a set of clones of size $\sqrt{n}$.

We construct a set $U \subseteq V(G)$ as follows. We start with $U = \emptyset$. At each stage, if the complement $U^{\mathsf{c}} = V(G) \setminus U$ contains an independent set I of size $5\sqrt{n}$, then I contains a set of clones of size at least $\sqrt{n}$. We add this set of clones to U and continue until U^{c} has no independent set of size $5\sqrt{n}$. Observe that the resulting set U is a disjoint union of sets of clones each of size at least $\sqrt{n}$, while the complement U^{c} has no independent set of size $5\sqrt{n}$ (see Figure 7).

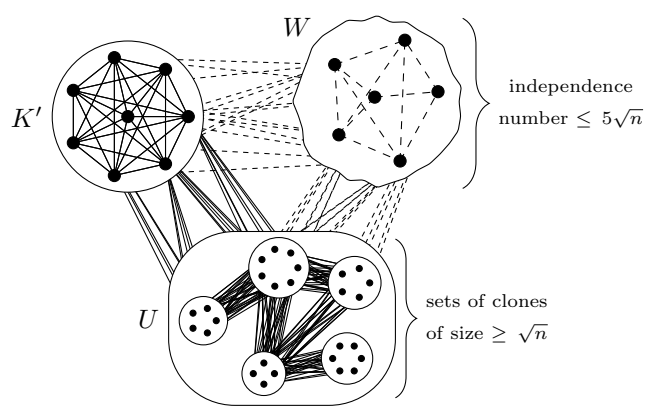

Figure 7: The sets U, W and K'.

In the following claim we deduce from Lemma 13 that $G[U^{\mathsf{c}}]$ is very dense.

Claim 33. *$G[U^{\mathsf{c}}]$ has $O(n^{3/2})$ non-edges.*

Proof. Since $G[U^{\mathsf{c}}]$ has no independent set of size at least $5\sqrt{n}$, every vertex in G has at most $5\sqrt{n}$ non-triangular neighbours in U^{c}. It follows that there are at most $5n^{3/2}$ non-triangular edges with at least one end in U^{c}.

Let l denote the number of non-edges in $G[U^{\mathsf{c}}]$. By adding these edges to G we obtain a graph G' with n vertices and $e(G)+l$ edges such that $t(G') \geq t(G) - 5n^{3/2}$. It follows from Lemma 13 that $l = O(n^{3/2})$. $\square$

Let K be a largest clique in G, so $|K| = \Omega(n)$ by Claim 32. Let $K' = K \setminus U$ and denote $W = U^{\mathsf{c}} \setminus K'$ (see Figure 7). Note that, since U contains no clique of size greater than $\sqrt{n}$, we have $|K'| \geq |K| - \sqrt{n} = \Omega(n)$. In the following claim we use the structure of U and Claim 33 to deduce that almost all vertices in K' are triangular.

Claim 34. *All but $O(\sqrt{n})$ vertices in K' are triangular.*

Proof. Since K' is a clique, any vertex in the complement $V(G) \setminus K'$ sends at most one non-triangular edge to K'. In fact, if $u \in V(G) \setminus K'$ has a non-triangular neighbour in K', then u has no other neighbours in K'.

Denote by m the number of vertices in W that have a non-triangular neighbour in K'. Then the number of missing edges in $G[U^{\mathsf{c}}]$ is at least $m(|K'|-1) = \Omega(mn)$. From Claim 33 we conclude that $m = O(\sqrt{n})$. Therefore, there are $O(\sqrt{n})$ vertices in K' with a non-triangular neighbour in U^{c}.

Finally, U is a union of at most $\sqrt{n}$ sets of clones, and any one set of clones can send non-triangular edges to at most one vertex in K'. Therefore, there are at most $\sqrt{n}$ vertices in K' that have a non-triangular neighbour in U. $\square$

The proof of Proposition 28 is complete. Indeed, K' consists of $\Omega(n)$ vertices and all but $O(\sqrt{n})$ of them are triangular. $\square$

5.2 Structure

In this subsection we build on the fact that G has $\Omega(n)$ triangular vertices and prove that, in terms of structure, G has some similarities with a graph $G(a,b,c)$. In particular, we prove that the vertices of G can be partitioned into three linearly sized sets A, B, C such that A is a clique and all edges between A and B are present in G, while all edges between A and C are missing. We do not yet prove that the sets B, C are independent, but we show that both of them can be partitioned into a small number of independent sets (see Figure 6). Our main tool in this subsection is the assumption that G is compressed, and we also use Lemma 13.

Proposition 29. *There exists a partition $\{A, B, C\}$ of $V(G)$ such that all parts have size $\Omega(n)$ and the following properties are satisfied.*

1. *A is the set of triangular vertices in G, it spans a clique and its vertices are adjacent to all of B and none of C.*

2. *B may be partitioned into $O(1)$ sets of clones and a remainder of size $O(\sqrt{n}\log n)$.*

3. *C may be partitioned into $O(1)$ sets of clones, each having $\Omega(n)$ non-triangular neighbours in B, and a remainder of size $O(n^{1/3}\log n)$.*

Proof. Denote by A the set of triangular vertices in G. Since G is compressed, A induces a clique and the vertices of A have the same neighbourhood outside of A. Denote this neighbourhood by B and let $C = V(G) \setminus (A \cup B)$. Property 1 follows.

Note that the graph $G(a, b, c)$, where $c = \delta n/2$, $b = \sqrt{\delta}\, n$ and $a = n - b - c$, has at least $(1/2 - \delta)n^2$ edges and $\delta^{3/2}n^2/2$ non-triangular edges. Hence, since G is optimal and $e(G) \leq (1/2 - \delta)n^2$, it follows that $t(G) = \Omega(n^2)$.

By Proposition 28 we have $|A| = \Omega(n)$. Note that there are no non-triangular edges with both ends in $A \cup B$, and so the number of non-triangular edges in G is at most $|C|n$. Since $t(G) = \Omega(n^2)$, it follows that $|C| = \Omega(n)$. We will deduce that $|B| = \Omega(n)$ from a stronger statement that almost all vertices in C have $\Omega(n)$ non-triangular neighbours in B.

Claim 35. *All but $O(1)$ vertices of C have $\Omega(n)$ non-triangular neighbours in B.*

Proof. Let $c > 0$ and $k \in \mathbb{N}$ be constants. Suppose that there is a set $Z \subseteq C$ of size k whose every vertex has at most cn non-triangular neighbours in B. Our aim is to show that if c is sufficiently small and k is sufficiently large, then the existence of such a set Z would lead to a contradiction.

Consider the graph G', obtained from G by adding the edges between Z and A. Then $e(G') = e(G) + k|A|$ and $t(G') \geq t(G) - ckn - \binom{k}{2} \geq t(G) - 2ckn$. Provided that k is sufficiently large, Lemma 13 implies that $t(G') \leq t(G) - \zeta k|A|$ for some constant $\zeta > 0$ that does not depend on c or k. Therefore, $\zeta|A| \leq 2cn$ must hold. However, we may choose c small enough to make this false, thus obtaining a contradiction. □

The previous claim provides us with a set $C' \subseteq C$ such that $|C \setminus C'| = O(1)$ and every vertex in C' has $\Omega(n)$ non-triangular neighbours in B. The following claim implies that C' may be partitioned into $O(1)$ independent sets.

Claim 36. *There exists a set $S \subseteq B$ of size $O(1)$ such that every vertex in C' has a non-triangular neighbour in S.*

Proof. We construct $S = \{u_1, \ldots, u_k\}$ by choosing the elements $u_1, \ldots, u_k \in B$ and certain corresponding subsets $I_1, \ldots, I_k \subseteq B$ in the following way. Suppose that $u_1, \ldots, u_j$ and $I_1, \ldots, I_j$ have been chosen, where $j \geq 0$. Let U_j be the set of vertices in C' that have a non-triangular neighbour in $\{u_1, \ldots, u_j\}$ (so, in particular, $U_0 = \emptyset$). If $U_j = C'$, we stop the process. Otherwise, pick a vertex $v \in C' \setminus U_j$ and consider the set N consisting of the non-triangular neighbours of v in B. By the

definition of C', we have $|N| = \Omega(n)$. Moreover, since N is independent and G is compressed, N contains a set of clones of size at least $|N|/5$. Denote this set of clones by I_{j+1} and pick $u_{j+1} \in I_{j+1}$ arbitrarily.

It is clear that when the process terminates, every vertex in C' has a non-triangular neighbour in the resulting set S. It remains to check that the process stops after $O(1)$ steps. Indeed, suppose that it ran for k steps. The sets $I_1, \ldots, I_k$ are pairwise disjoint and have size at least $\Omega(n)$ each, whence $k = O(1)$. □

The non-triangular neighbourhoods of the vertices in S cover C'. Therefore, C' can be partitioned into $O(1)$ independent sets. Since G is compressed, each independent set can be partitioned into $O(\log n)$ sets of clones, all but at most four of which have size $O(n^{1/3})$. By combining the sets of clones of size $O(n^{1/3})$ into one set, we get a partition of C' into $O(1)$ sets of clones and a remainder of size $O(n^{1/3} \log n)$. Note that, by definition, every vertex in C' has $\Omega(n)$ non-triangular neighbours in B. Now, throw all of the $O(1)$ vertices of $C \setminus C'$ into the remainder to get a partition of C that satisfies Property 3.

It remains to prove Property 2. Partition C into sets Z, C'' where $|Z| = O(n^{1/3} \log n)$ and C'' is a union of $O(1)$ sets of clones. Let Y be the set of vertices in B that do not have non-triangular neighbours in C'' and denote $B' = B \setminus Y$. First, we will show that B' can be partitioned into $O(1)$ independent sets. Indeed, B' is covered by the non-triangular neighbourhoods of vertices in C'', and each of them is an independent set. Moreover, C'' is a union of $O(1)$ sets of clones, and so there are $O(1)$ distinct such neighbourhoods. Second, we will prove that $|Y| = O(\sqrt{n} \log n)$.

Claim 37. $|Y| = O(\sqrt{n} \log n)$.

Proof. Recall that A is the set of triangular vertices in G. Since Y is disjoint from A, every vertex in Y has a non-triangular neighbour, and that neighbour must be in Z. That is, the non-triangular neighbourhoods of vertices in Z cover Y. Since Z is a union of $O(\log n)$ sets of clones, Y can be partitioned into $O(\log n)$ independent sets. In particular, Y contains an independent set I of size $\Omega(|Y|/\log n)$.

Let G' be the graph obtained from G by adding all possible edges spanned by $|I|$. Then $e(G') = e(G) + \binom{|I|}{2}$ and $t(G') \geq t(G) - |I||Z| \geq t(G) - O(|I|n^{1/3} \log n)$. This is a contradiction to Lemma 13 unless $\binom{|I|}{2} = O(n)$ or $\binom{|I|}{2} = O(|I|n^{1/3} \log n)$. In either case $|I| = O(\sqrt{n})$, and so $|Y| = O(\sqrt{n} \log n)$, as required. □

The proof of Proposition 29 is now complete. We have already proved Properties 1 and 3 and that A, B, C are all of size $\Omega(n)$. To prove Property 2, recall that B is partitioned into a set Y of size $O(\sqrt{n} \log n)$ and a set B' which is a union of $O(1)$ independent sets. It follows from a similar argument as earlier that B' can be partitioned into $O(1)$ sets of clones and a remainder of size $O(n^{1/3} \log n)$. Assigning Y to this remainder gives the desired partition of B. □

5.3 Sizes

In the previous subsection we proved that $V(G)$ can be partitioned into sets A, B, C that correspond to the three parts of the graph $G(|A|, |B|, |C|)$. In this subsection we consider the sizes of the sets A, B, C. We show that the number of edges (and non-triangular edges) of G is very close to the number of edges (and non-triangular edges) of $G(|A|, |B|, |C|)$.

Proposition 30. *Let A, B, C be as in Proposition 29 and denote $a = |A|, b = |B|, c = |C|$. Then $e(G) = a^2/2 + ab + bc + O(n^{7/4}\sqrt{\log n})$ and $t(G) = bc + O(n^{7/4}\sqrt{\log n})$.*

In the proof of this proposition we revisit Füredi and Maleki's [4] proof of Theorem 2 which is an approximate version of our main theorem. In their proof, Füredi and Maleki repeatedly apply Lemma 11, which eliminates one vertex at a time from any independent set of size 3. Here, we will do the same thing, but we will keep tight control on the independent sets to which we apply this lemma.

Proof of Proposition 30. Recall that by Proposition 29 both sets B and C can be partitioned into $O(1)$ sets of clones and a remainder of size $O(\sqrt{n}\log n)$. Let G' be the graph obtained by removing the edges incident with vertices in this remainder. Then $e(G') \geq e(G) - O(n^{3/2}\log n)$ and $t(G') \geq t(G) - O(n^{3/2}\log n)$.

The following claim is a variation of Lemma 10. It allows us to approximate G' by a weighted subgraph whose intersection with C induces a clique.

Claim 38. *There is a weighted subgraph H of G' such that $|H| = n$, $e(H) \geq e(G')$ and $t(H) \geq t(G')$, which has the following properties.*

- *At least two vertices in A are present in H. Moreover, with at most one exception, the vertices in A that are present in H have weight 1.*
- *All vertices in B are present in H and have weight 1.*
- *The vertices in C that are present in H induce a clique.*

Proof. We perform the following process to obtain the weighted graph H. Initially, we set H to be G' with every vertex having weight 1. Then we perform multiple steps, during which we modify the weights of the vertices in $A \cup C$ (and remove some of these vertices) so that, at any given time, A has at most one vertex with weight not equal to 1. At each step we select vertices $u \in A$ and $v, w \in C$. We take u to be the unique vertex in A of weight not equal to 1, and if there is no such vertex, then we take it to be an arbitrary vertex remaining in A. We take v and w to be any pair of non-adjacent (in G') vertices in C. If choosing u, v, w according to these rules is impossible, then we terminate the process.

Suppose that we successfully selected the vertices u, v, w. They form an independent set, and so by Lemma 11 it is possible to remove one or two of these vertices and redistribute their weight on the remaining ones so that the new weights are positive, the total weight does not change and $e(H), t(H)$ do not decrease.

It is clear that this process terminates, because each step decreases the number of vertices remaining in H. Let us consider the resulting weighted graph H. Since the process terminated, either no vertices of A are present in H, or the remaining vertices of C induce a clique. We show that, in fact, at least two vertices remain in A, and so the latter condition must hold.

Suppose that fewer than two vertices in A remain in H. Denote by m the size of the largest clique that can be formed from vertices remaining in H. Since the vertex set of G' can be partitioned into A and $O(1)$ independent sets, we have $m = O(1)$. Apply Lemma 10 to obtain a good weighted subgraph F of H, with xy being its only non-triangular edge, such that $|F| = n$, $e(F) \geq e(G')$ and $t(F) \geq t(G')$. Let β and γ be the weights of x and y in F and suppose that $\beta \geq \gamma$. Then $\alpha = n-\beta-\gamma$ is the sum of the weights of the other vertices in F. We have $t(F) = \beta\gamma$ and, as in Inequality (9) from Claim 32, $e(F) \leq (1 - 1/m)\alpha^2/2 + \alpha\beta + \beta\gamma$. It follows that $t(G) \leq \beta\gamma + O(n^{3/2} \log n)$ and $e(G) \leq \alpha^2/2 + \alpha\beta + \beta\gamma - \Omega(n^2)$. Consider the graph $G'' = G(n - \lceil\beta\rceil - \lceil\gamma\rceil, \lceil\beta\rceil, \lceil\gamma\rceil)$. It is easy to check that $t(G'') \geq t(G) - O(n^{3/2} \log n)$ and $e(G'') \geq e(G) + \Omega(n^2)$. This is a contradiction to Lemma 13, since G is optimal. Therefore, at least two vertices in A are present in H.

It follows that the set of vertices in C that are present in H induces a clique. Hence, the weighted graph H satisfies the requirements of Claim 38. □

Let H be a weighted graph as given by Claim 38, so in particular, $e(H) \geq e(G) - O(n^{3/2} \log n)$ and $t(H) \geq t(G) - O(n^{3/2} \log n)$. By Lemma 13, since G is optimal,

$$\begin{aligned} e(H) &= e(G) + O\big(n^{3/2} \log n\big), \\ t(H) &= t(G) + O\big(n^{3/2} \log n\big). \end{aligned} \tag{10}$$

We remark that these two lines express both upper and lower bounds for the quantities $e(H)$ and $t(H)$. In the following claim we prove that, in fact, only one vertex of C is present in H.

Claim 39. *Exactly one vertex of C is present in H. Moreover, all but at most $O(n^{3/4}\sqrt{\log n})$ vertices in B are non-triangular neighbours of that vertex.*

Proof. Denote by $u_1, \ldots, u_m$ the vertices in C that are present in H, and let $N_1, \ldots, N_m$ be their non-triangular neighbourhoods in B. Since the set $\{u_1, \ldots, u_m\}$ forms a clique, there are no edges between u_i and N_j for $i \neq j$. In particular, the sets $N_1, \ldots, N_m$ are pairwise disjoint.

Let $Z = B \setminus (N_1 \cup \cdots \cup N_m)$. Since the intersection of H with A induces a clique on at least two vertices and since all edges between B and the intersection of H with A are present in H,

the vertices in Z are not incident with any non-triangular edges in H. We will show that $|Z| = O(n^{3/4}\sqrt{\log n})$. Indeed, recall that B is the union of $O(1)$ independent sets and a remainder of size at most $O(\sqrt{n}\log n)$. Thus, provided that $|Z| \geq C\sqrt{n}\log n$ for a sufficiently large constant C, there exists an independent set $I \subseteq Z$ of size $\Omega(|Z|)$. Consider the weighted graph H' obtained from H by adding the edges spanned by I. Then $e(H') = e(H) + \Omega(|Z|^2) \geq e(G) - O(n^{3/2}\log n) + \Omega(|Z|^2)$ and $t(H') = t(H) \geq t(G) - O(n^{3/2}\log n)$. It follows from Lemma 13 that $|Z| = O(n^{3/4}\sqrt{\log n})$.

Our aim is to prove that $m = 1$. We assume for contradiction that $m \geq 2$. In particular, since by Proposition 29 and the definition of G', in G' every vertex in C is either isolated or has $\Omega(n)$ non-triangular neighbours in B, the vertices $u_1, \ldots, u_m$ have the latter property. In other words, $|N_i| = \Omega(n)$ for every i.

For each i, let γ_i denote the weight of u_i in H. We will show that $\gamma_i = \Omega(n)$ for every i. Indeed, fix any i. Denote by H_i the weighted graph obtained from H by adding all edges spanned by N_i. Since N_i is an independent set in H, we have $e(H_i) \geq e(G) + \Omega(n^2)$ and $t(H_i) \geq t(G) - |N_i|\gamma_i - O(n^{3/2}\log n)$. By Lemma 13, $\gamma_i = \Omega(n)$.

Write $\beta_i = |N_i|$. Construct a weighted graph F, starting from H and carrying out the following steps. Firstly, remove all edges with an end in Z. Secondly, replace each set N_i by a vertex v_i of weight β_i. Finally, connect each vertex v_i to all of the vertices in A (that are present in H) as well as to u_i and v_j for every $j \neq i$ (see Figure 8). We have $e(F) \geq e(H) - |Z|n \geq e(G) - O(n^{7/4}\sqrt{\log n})$ and $t(F) \geq t(H) \geq t(G) - O(n^{3/2}\log n)$.

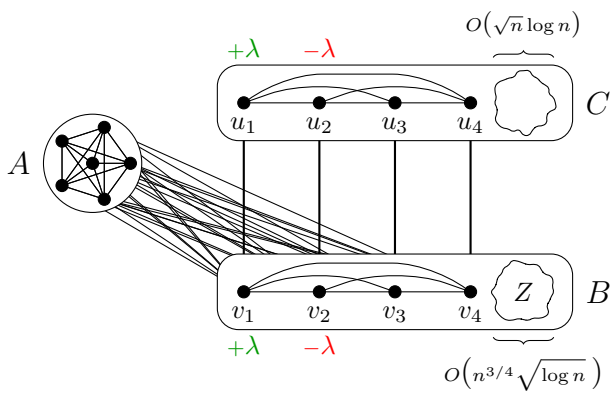

Figure 8: The graph F.

Pick any real λ such that $|\lambda| \leq \min\{\beta_1, \beta_2, \gamma_1, \gamma_2\}$. Let F_λ be the weighted graph obtained from F by adding λ to the weights of u_1 and v_1 and subtracting λ from the weights of u_2 and v_2. Clearly, $|F_\lambda| = |F| = n$ and it is easy to check that $e(F_\lambda) = e(F)$.

If $m \geq 3$, then the only non-triangular edges in F are $u_i v_i$. Hence, in this case,

$$\begin{aligned} t(F_\lambda) &= t(F) - (\beta_1\gamma_1 + \beta_2\gamma_2) + (\beta_1 + \lambda)(\gamma_1 + \lambda) + (\beta_2 - \lambda)(\gamma_2 - \lambda) \\ &= t(F) + (\beta_1 + \gamma_1 - \beta_2 - \gamma_2)\lambda + 2\lambda^2. \end{aligned}$$

If $\beta_1 + \gamma_1 \geq \beta_2 + \gamma_2$, then take $\lambda = \min\{\beta_1, \gamma_1, \beta_2, \gamma_2\}$. Otherwise, take $\lambda = -\min\{\beta_1, \gamma_1, \beta_2, \gamma_2\}$. In either case, $|\lambda| = \Omega(n)$ and $t(F_\lambda) \geq t(F) + \Omega(n^2) \geq t(G) + \Omega(n^2)$, contradicting Lemma 13.

This calculation is slightly different in the case when $m = 2$, because then we have to account for the edge u_1u_2, which is also non-triangular. In this case

$$\begin{aligned} t(F_\lambda) &= t(F) - (\beta_1\gamma_1 + \beta_2\gamma_2 + \gamma_1\gamma_2) + (\beta_1 + \lambda)(\gamma_1 + \lambda) + (\beta_2 - \lambda)(\gamma_2 - \lambda) + (\gamma_1 + \lambda)(\gamma_2 - \lambda) \\ &= t(F) + (\beta_1 - \beta_2)\lambda + \lambda^2. \end{aligned}$$

We may reach a contradiction to Lemma 13 by choosing λ of the same sign as $\beta_1 - \beta_2$ and with $|\lambda| = \min\{\beta_1, \beta_2, \gamma_1, \gamma_2\}$. We conclude that $m = 1$, completing the proof of the claim. □

Recall that, among the vertices in A that are present in H, at most one has weight not equal to 1. In the following claim we show that this weight cannot be very large.

Claim 40. *The weight in H of any vertex in A is $O(n^{3/4}\sqrt{\log n})$.*

Proof. Let u be a vertex of A of maximal weight in H, and let ω be its weight. Suppose that $\omega > 1$, in which case all other vertices in A have weight 1 in H.

Replace the vertex u by a clique of size $\lfloor\omega\rfloor$ whose vertices have weight $\omega/\lfloor\omega\rfloor$ and are adjacent to all of $(A \setminus \{u\}) \cup B$ and denote the resulting weighted graph by H'. We have to check the technical condition that the average weight of a vertex in H' is at least 1. However, this can be easily verified, since the total weight of H' is an integer and H' has at most one vertex whose weight is smaller than 1 (namely, the only vertex of C that remains in H').

By replacing u with a clique, we create new edges inside the clique, and these edges contribute $\binom{\lfloor\omega\rfloor}{2}(\omega/\lfloor\omega\rfloor)^2 = \Omega(\omega^2)$ towards $e(H')$. Therefore, we have $t(H') = t(H) \geq t(G) - O(n^{3/2}\log n)$ and $e(H') = e(H) + \Omega(\omega^2) \geq e(G) - O(n^{3/2}\log n) + \Omega(\omega^2)$. It follows from Lemma 13 that $\omega = O(n^{3/4}\sqrt{\log n})$. □

Recall that a, b, c are the sizes of the sets A, B, C in the original graph G. Let α, β, γ be the sums of weights (in H) of the vertices in these sets, summing over vertices present in H. So, for example, $\beta = b$ and γ is the weight of the single vertex in C that is present in H. Clearly, $\alpha + \gamma = a + c$, because both sides are equal to $n - b$. Now, we use the properties of H that we have proved to get good bounds on $e(H)$ and $t(H)$ in terms of α, β, γ.

Recall that the set A induces a clique in G, so its remainder induces a clique in H. Combined with Claim 40, this implies that the contribution of the edges within A to $e(H)$ is $\alpha^2/2 - O(n^{3/2}\log n)$. By Claim 39, the set B contains an independent set of size at least $|B| - O(n^{3/4}\sqrt{\log n})$. Therefore, the contribution of the edges within B to $e(H)$ (and in particular to $t(H)$) is $O(n^{7/4}\sqrt{\log n})$. Moreover, Claim 39 implies that the edges between B and C contribute $\beta\gamma - O(n^{7/4}\sqrt{\log n})$ to both $e(H)$ and $t(H)$. Putting this together, we get

$$\begin{aligned} e(H) &= \alpha^2/2 + \alpha\beta + \beta\gamma + O\big(n^{7/4}\sqrt{\log n}\big) \\ t(H) &= \beta\gamma + O\big(n^{7/4}\sqrt{\log n}\big). \end{aligned} \tag{11}$$

Again, we remark that these are both upper and lower bounds for the quantities $e(H)$ and $t(H)$. We deduce that α almost equals a and γ almost equals c.

Claim 41. *$\alpha = a + O(n^{3/4}\sqrt{\log n})$ and $\gamma = c + O(n^{3/4}\sqrt{\log n})$.*

Proof. We can read the inequality $\alpha \le a + O(n^{3/4}\sqrt{\log n})$ off Claim 40. To get the corresponding lower bound on α, we consider the quantity $e(H) - t(H)$. On one hand, the inequalities in (11) give

$$e(H) - t(H) = \alpha^2/2 + \alpha\beta + O\big(n^{7/4}\sqrt{\log n}\big).$$

On the other hand, we can use the inequalities in (10) to get

$$\begin{aligned} e(H) - t(H) &= e(G) - t(G) + O(n^{3/2}\log n) \\ &\ge a^2/2 + ab + O(n^{3/2}\log n), \end{aligned}$$

where the latter inequality comes from the fact that the quantity $e(G) - t(G)$ counts the triangular edges in G, and all vertices in A are triangular. Recall that $b = \beta$. Combining the two inequalities for $e(H) - t(H)$ we get $\alpha \ge a - O(n^{3/4}\sqrt{\log n})$, so $\alpha = a + O(n^{3/4}\sqrt{\log n})$. To complete the proof of the claim, note that $a + c = \alpha + \gamma$. □

Proposition 30 follows from the last claim and the inequalities in (10) and (11). □

5.4 End of the proof

We are now ready to complete the proof of Theorem 5.

Proposition 31. *$G \cong G(a, b, c)$ for some a, b, c.*

We gradually get closer to proving that $G \cong G(a, b, c)$. We start by showing that b is much bigger than c, which leads to the conclusion that C spans no non-triangular edges. This implies that almost all possible edges between B and C are present in G and are non-triangular, by Proposition 30. In

fact, using the fact that G is compressed, we deduce that there are large subsets of B and of C that span a complete bipartite graph consisting of non-triangular edges. With some more effort, using the optimality of G, we conclude that B and C themselves induce a complete bipartite graph, thus completing the proof.

Proof of Proposition 31. Let $\{A, B, C\}$ be the partition of $V(G)$ given by Proposition 29. As in the statement of Proposition 30, write $a = |A|, b = |B|, c = |C|$. We start by showing that b is significantly larger than c.

Claim 42. *We have $b \geq c + \Omega(n)$.*

Proof. Suppose to the contrary that $b \leq c + o(n)$. Then we have $n = a + b + c \geq 2b + a - o(n)$, and hence $b \leq (n - a + o(n))/2$. Since $a = \Omega(n)$, we can conclude that $b \leq n/2 - \Omega(n)$.

Consider the graph $H = G(a, b, c)$. Proposition 30 implies that $e(H) = e(G) + O(n^{7/4}\sqrt{\log n})$ and $t(H) = t(G) + O(n^{7/4}\sqrt{\log n})$. Consider also the graph $H' = G(a, b + d, c - d)$, where $d = n^{0.999}$. Note that $b + d \leq n/2 - \Omega(n)$. Therefore, from the expression $e(H) = \binom{a}{2} + (n - b)b$ and the corresponding expression for $e(H')$, we can see that $e(H') \geq e(H) + \Omega(dn) = e(G) + \Omega(dn)$. Similarly, $t(H') = (b + d)(c - d) \geq bc - o(dn)$, and so $t(H') \geq t(H) - o(dn) = t(G) - o(dn)$. However, this contradicts Lemma 13. Therefore, $b \geq c + \Omega(n)$. □

In the following claim we conclude that C spans no non-triangular edges.

Claim 43. *There are no non-triangular edges with both ends in C.*

Proof. By Proposition 30 there are $bc + o(n^2)$ non-triangular edges in G, and each one of them is incident with a vertex in C. Therefore, some vertex in C has at least $b - o(n)$ non-triangular neighbours. Thus, by Observation 8, every vertex in C has degree at least $b - o(n)$, and so the sum of the degrees of any two vertices in C is at least $2b - o(n) > b + c = |B \cup C|$, where the latter inequality comes from the previous claim. Since the vertices in C have neighbours only in $B \cup C$, it follows that any pair of vertices in C have a common neighbour, and hence they cannot be joined by a non-triangular edge. □

Recall that by Proposition 29 both sets B and C can be partitioned into $O(1)$ sets of clones and a remainder of size $O(\sqrt{n}\log n)$. In such a partition of B consider the sets of clones of size at least $n^{9/10}$ and let B' be their union. Similarly, let C' be the union of the sets of clones in the partition of C that have size at least $n^{9/10}$ and denote $Z = C \setminus C'$ and $Y = B \setminus B'$. Then $|Y| = O(n^{9/10})$ and $|Z| = O(n^{9/10})$. We show that all possible edges between B' and C' are present in G and are non-triangular.

Claim 44. *All possible edges between B' and C' are present in G and are non-triangular. In particular, B' and C' are independent sets.*

Proof. From the previous claim we know that every non-triangular edge in G has one end in B and one end in C. Suppose that there exists a pair of vertices, one in B' and one in C', that are not joined by a non-triangular edge. Then there are two sets of clones of size at least $n^{9/10}$, one contained in B' and the other in C', between which there are no non-triangular edges. But then $t(G) \leq bc - n^{9/5}$, contradicting Proposition 30. $\square$

In the following claim we obtain additional information about Y and Z, which brings us closer to showing that B and C induce a complete bipartite graph.

Claim 45. *There are no edges between B' and Y and between C' and Z. Moreover, every vertex in B has at least $c - o(n)$ neighbours in C, and every vertex in C has at least $b - o(n)$ neighbours in B.*

Proof. Any vertex in C' has at least $|B'| = b - o(n)$ non-triangular neighbours, so Observation 8 implies that every vertex in G has degree at least $b - o(n)$. Claim 42 implies that $|C| + |B \setminus B'|$, which does not exceed $c + o(n)$, is smaller than this quantity, and hence every vertex in C has a neighbour in B'. Therefore, because all possible edges between B' and C' are present in G and are non-triangular (by Claim 44), there are no edges between C' and Z. In particular, every vertex in C has at most $o(n)$ neighbours in C, so it has at least $b - o(n)$ neighbours in B.

Pick any vertex $u \in Y$. Since u is not in A, u has a non-triangular neighbour $v \in C$. We have just proved that v has at least $b - o(n)$ neighbours in B'. Therefore, u has at most $o(n)$ neighbours in B. Now suppose that u is adjacent to a vertex in B'. Then u has no neighbours in C'. Hence, u has at most $a + o(n)$ neighbours, of which at most $o(n)$ are non-triangular. However, any vertex of B' has at least $a + c - o(n)$ neighbours and at least $c - o(n)$ of them are non-triangular. This contradicts Observation 8. Therefore, u has no neighbours in B'.

It remains to verify that every vertex in Y has at least $c - o(n)$ neighbours in C. Let us again consider $u \in Y$ and denote by d the number of its neighbours in C. Then the degree of u is at most $a + d + o(n)$ and its non-triangular degree is at most d. By Observation 8, applied to u and any vertex in B', we know that $a + d \geq a + c - o(n)$ or $d \geq c - o(n)$. In either case $d \geq c - o(n)$. $\square$

In the following claim we prove that no edges are spanned by Z. We use a trick that we have used several times before, replacing a pair of adjacent vertices in Z by copies of vertices in A and B', increasing the number of both edges and non-triangular edges.

Claim 46. *The set Z is independent.*

Proof. Suppose to the contrary that Z contains a pair of adjacent vertices u, v. Then the non-triangular neighbours of u are all in B and they are not adjacent to v. By Claim 45, v has $b - o(n)$

neighbours in B, and hence u has at most $o(n)$ non-triangular neighbours. Likewise, v has at most $o(n)$ non-triangular neighbours.

Consider the graph G' obtained from G by removing u, v and adding new vertices x and y, where x is joined by edges to all vertices in $A \cup B$, and y is joined to all vertices in B'. We have $e(G') \geq e(G) + a - o(n)$ and $t(G') \geq t(G) + b - o(n)$. This contradicts the optimality of G, because $a = \Omega(n)$ and $b = \Omega(n)$. □

A similar trick enables us to conclude that Y spans no edges. Here we replace two adjacent vertices in Y by copies of vertices in A and B'.

Claim 47. *The set Y is independent.*

Proof. Suppose that there exists a pair of adjacent vertices $u, v \in Y$. Let G' be the graph obtained from G by removing u, v and adding new vertices x, y with x joined to all of $A \cup B'$ and y joined to all of $A \cup C'$.

Let us compare $e(G')$ and $t(G')$ with $e(G)$ and $t(G)$. By Claim 45, there are no edges in G between $\{u, v\}$ and B'. Therefore, the removal of u, v removes at most $2(a + c + o(n))$ edges. On the other hand, the addition of x, y creates $2a+b+c-o(n)$ new edges. Therefore, $e(G') \geq e(G)+b-c-o(n) > e(G)$. Furthermore, since u, v have at least $c - o(n)$ neighbours in C each, there are at most $o(n)$ vertices in C that are adjacent to precisely one of u, v. As a result, u, v are incident with at most $o(n)$ non-triangular edges in G. Since the addition of x, y creates $c-o(n)$ new non-triangular edges, we have $t(G') \geq t(G)+c-o(n) > t(G)$. This contradicts the optimality of G, because $c = \Omega(n)$. □

Proposition 31 easily follows from Claims 44 to 47. Indeed, these claims together imply that B and C are independent sets in G. Therefore, if there were any missing edges between B and C, we could add them to G without creating new triangles. Since G is an optimal graph, all possible edges between B and C are present. It follows that $G \cong G(|A|, |B|, |C|)$. □

6 Almost complete

In this section we prove Theorem 6.

Theorem 6. *There exist n_0 and $\delta > 0$ such that the following holds. Let G be a graph with $n \geq n_0$ vertices and e edges, where $e \geq (1/2 - \delta)n^2$. Then there exists a graph $H = G(a, b, c)$ such that $|H| = n$, $e(H) \geq e$ and $t(H) \geq t(G)$.*

The proof in this range is easier than in the other two ranges, though far from immediate. We start by making the usual assumption that G is an optimal and compressed graph, even though we do not use the full strength of the latter assumption: we only need Condition 2 from Definition 14.

If very few (namely, $2n-8$ or fewer) edges are missing from G, then we directly prove that $G \cong G(a,b,c)$ for some a,b,c. For the remaining range, we partition the vertices of G, according to their degrees, into sets A, B, C with the aim of showing that $G \cong G(|A|,|B|,|C|)$. We first prove that the sets have the correct orders of magnitude using a rough lower bound on $t(G)$. We are then able to prove better estimates for the sizes of the sets, and, finally, we deduce that G has the required structure.

Proof of Theorem 6. Fix a sufficiently small constant $\delta > 0$ (whose value can be determined from the proof) and let G be an optimal and compressed graph with n vertices and $\binom{n}{2} - \varepsilon n^2$ edges, where $0 \le \varepsilon \le \delta$. We first consider the case $e(G) \ge \binom{n}{2} - (2n-9)$.

Claim 48. *If $e(G) \ge \binom{n}{2} - (2n-9)$, then $G \cong G(a,b,c)$ for some a,b,c.*

Proof. If G has no non-triangular edges, then it is a clique by optimality, so we are done. We claim that G does not have two independent non-triangular edges. Indeed, if uv and xy are such edges, then for any other vertex w one of the two possible edges uw and vw is missing, as well as one of xw and yw. Therefore, G has at least $2n-8$ missing edges, contradicting our assumption. Therefore, since the triangle-free edges cannot form a triangle, they form a star. Let $uv_1, \dots, uv_k$ be the non-triangular edges. Then the set $A = V(G) \setminus \{u, v_1, \dots, v_k\}$ is the set of triangular vertices in G, so A induces a clique and all of the vertices in A have the same neighbourhood in $V(G) \setminus A$. Now, there are two possibilities: either u is adjacent to all of A, or u is not adjacent to any vertex in A. In the former case, there are no edges between A and $\{u_1, \dots, u_k\}$, and so $G \cong G(|A|, 1, k)$. In the latter case, optimality of G implies that all possible edges between A and $\{u_1, \dots, u_k\}$ are present in G, and so $G \cong G(|A|, k, 1)$. $\square$

From this point onwards we assume that $e(G) \le \binom{n}{2} - (2n-8)$. In particular, $\varepsilon \ge (2 - o(1))/n$. We wish to prove that G is isomorphic to the graph $G(a,b,c)$ for some parameters a,b,c. To get some idea on how large these parameters should be, we observe that $a \approx n$, because the number of missing edges is small. Now, the number of missing edges, $ac + \binom{b}{2} + \binom{c}{2}$, can be reasonably approximated by $cn + b^2/2$. Subject to b, c being non-negative reals such that $cn + b^2/2 \le \varepsilon n$, the quantity bc is maximised when $b = \sqrt{2\varepsilon/3}\, n, c = (2\varepsilon/3)n$. Therefore, we expect G to be isomorphic to $G(a,b,c)$ with $b \approx \sqrt{2\varepsilon/3}\, n$ and $c \approx (2\varepsilon/3)n$. We can use this conclusion to get a lower bound on $t(G)$.

Claim 49. $t(G) = \Omega(\varepsilon^{3/2} n^2)$.

Proof. Let $G' = G(a,b,c)$, where $b = \lfloor \sqrt{2\varepsilon/3}\, n \rfloor$, $c = \lfloor (2\varepsilon/3)n \rfloor$ and $a = n - b - c$. There at most $cn + b^2/2 \le \varepsilon n$ edges missing from G', so $e(G') \ge e(G)$. We now find a lower bound for $t(G')$ by a simple computation, but we have to be careful with rounding errors.

We have $\varepsilon n^2 \geq 2n - 9$, implying that $(2\varepsilon/3)n > 1$, and hence $c = \lfloor (2\varepsilon/3)n \rfloor = \Omega(\varepsilon n)$. Similarly, $b = \Omega(\sqrt{\varepsilon} n)$. It follows that $t(G') = bc = \Omega(\varepsilon^{3/2} n^2)$. Since G is optimal, we have $t(G) \geq t(G') = \Omega(\varepsilon^{3/2} n^2)$. □

We now define three sets $A, B, C \subseteq V(G)$ that correspond to the three parts of a graph $G(a, b, c)$. Let C be the set of vertices of degree at most $3n/4$, let B be the set of vertices in $V(G) \setminus C$ that have a non-triangular neighbour in C, and let $A = V(G) \setminus (B \cup C)$. Since any two vertices in $A \cup B$ have at least $n/2$ common neighbours, there are no non-triangular edges with both ends in $A \cup B$. Therefore, all vertices in A are triangular, so A induces a clique and its vertices have the same neighbourhood in $V(G) \setminus A$.

The next step is to obtain tight bounds for the sizes of A, B, C. First, we determine the order of magnitude of $|B|$ and $|C|$.

Claim 50. *$|B| = \Theta(\sqrt{\varepsilon} n)$ and $|C| = \Theta(\varepsilon n)$. Moreover, every vertex of B is an end of $\Omega(\sqrt{\varepsilon} n)$ missing edges.*

Proof. By definition, every vertex in C is an end of at least $n/4$ non-edges. Since there are εn^2 non-edges in total, we have $|C| = O(\varepsilon n)$. We know from the previous claim that there are at least $\Omega(\varepsilon^{3/2} n^2)$ non-triangular edges. All of these edges have at least one end in C, and so some vertex in C has at least $\Omega(\sqrt{\varepsilon} n)$ non-triangular neighbours. Therefore, by Observation 8, every vertex in G has at least $\Omega(\sqrt{\varepsilon} n)$ neighbours.

Pick any $v \in B$. By the definition of B, v has a non-triangular neighbour $u \in C$. This means that v is not adjacent to any neighbours of u, and so v is an end of at least $\Omega(\sqrt{\varepsilon} n)$ non-edges. Therefore, $|B| = O(\sqrt{\varepsilon} n)$. Moreover, since every non-triangular edge has both ends in C, or one in B and one in C, we have $|B||C| + |C|^2/2 \geq \Omega(\varepsilon^{3/2} n^2)$, which implies that $|B| = \Omega(\sqrt{\varepsilon} n)$ and $|C| = \Omega(\varepsilon n)$. □

An immediate consequence of the previous claim is that $|A| = (1 - O(\sqrt{\varepsilon}))n$. Recall that all vertices in A have the same neighbourhood in $V(G) \setminus A$. In particular, each vertex in $B \cup C$ is adjacent either to all vertices in A or to none of them. Since the vertices in B have degree at least $3n/4$, they are all adjacent to all of A, and, similarly, there are no edges between A and C. We can use this fact to give a better upper bound on $|C|$.

Claim 51. *There exists a constant $\xi > 0$ such that $|C| \leq (1 - \xi)\varepsilon n$.*

Proof. Every vertex in B is an end of $\Omega(\sqrt{\varepsilon} n)$ missing edges and $|B| = \Theta(\sqrt{\varepsilon} n)$, so there are $\Omega(\varepsilon n^2)$ missing edges with an end in B. Since all edges between A and C are missing, we have $(1 - O(\sqrt{\varepsilon}))n|C| + \Omega(\varepsilon n^2) \leq \varepsilon n^2$. Therefore, $|C| \leq (1 - \Omega(1))\varepsilon n/(1 - O(\sqrt{\varepsilon}))$, and the claim follows provided that ε is sufficiently small. □

It is now possible to accurately relate the sizes of B and C. Write $|C| = \gamma\varepsilon n$, where $\Omega(1) = \gamma \leq 1-\xi$. Define $\beta = \sqrt{2(1-\gamma)}$ and note that $\beta = \Theta(1)$.

Claim 52. *$|B| = \beta\sqrt{\varepsilon}n + O(\varepsilon n)$. Moreover, there are at least $|B||C| - O(\varepsilon^2 n^2)$ non-triangular edges between B and C.*

Proof. Let G' be the graph $G(a,b,c)$, where $c = |C|$, $b = \lfloor \beta\sqrt{\varepsilon}n \rfloor$ and $a = n-b-c$. It is easy to see that $b^2/2 + cn \leq \varepsilon n^2$. In particular, we have $e(G') \geq \binom{n}{2} - \varepsilon n^2 = e(G)$. Therefore, since G is optimal, $t(G) \geq t(G') = bc$.

Let us come back to the graph G. Since every non-triangular edge has an end in C, some vertex in C has at least b non-triangular neighbours. It follows from Observation 8 that every vertex in G has degree at least $b-1$. Moreover, since vertices in C are adjacent only to vertices in $B \cup C$, we have $|B| \geq b - c - 1 = b - O(\varepsilon n)$.

Every vertex in B has a non-triangular neighbour, and therefore is an end of at least $b-1$ missing edges. Hence, there are at least $|B|(b-1)/2$ missing edges with an end in B. Since there are no edges between A and C, we have

$$\frac{1}{2}|B|(b-1) + \big(1 - O(\sqrt{\varepsilon})\big)\, cn \;\leq\; \varepsilon n^2 \;\leq\; \frac{1}{2}b^2 + cn + O(\sqrt{\varepsilon}n),$$

where the latter inequality follows from the definition of b. It follows that $|B|b \leq b^2 + O(\sqrt{\varepsilon}cn)$, and hence $|B| \leq b + O(\varepsilon n)$. To finish the proof, observe that $t(G) \geq bc = |B||C| - O(\varepsilon^2 n^2)$ and recall that the non-triangular edges of G are either spanned by C (there are $O(\varepsilon^2 n^2)$ such edges) or they have one end in B and the other in C. □

A standard trick of replacing two vertices by copies of other vertices, which we have been using throughout the paper, allows us to conclude that C is an independent set.

Claim 53. *The set C is independent. Moreover, every vertex in C is adjacent to all but at most $O(\varepsilon n)$ vertices in B.*

Proof. The second conclusion of Claim 52 implies that some vertex in C has at least $|B| - O(\varepsilon n)$ non-triangular neighbours in B. As a consequence, B contains an independent set I of size $|B| - O(\varepsilon n)$. Moreover, Observation 8 implies that every vertex in C is adjacent to all but at most $O(\varepsilon n)$ vertices in $B \cup C$.

Suppose that C contains a pair of adjacent vertices u, v. Let G' be the graph obtained from G by removing the vertices u and v and adding new vertices x and y where x is adjacent to all of $A \cup B$, and y is adjacent to all of I. The removal of u and v decreases the total number of edges by at most $2(|B| + |C|) = O(\sqrt{\varepsilon}n)$, while the addition of x and y increases this number by at least $|A| = (1 - O(\sqrt{\varepsilon}))n$. Therefore, $e(G') > e(G)$. Moreover, since u and v are adjacent, they do

not form non-triangular edges with their common neighbours. Hence, u and v have at most $O(\varepsilon n)$ non-triangular neighbours in total. On the other hand, the addition of x and y adds $|I| = \Omega(\sqrt{\varepsilon}n)$ non-triangular edges. Therefore, $t(G') > t(G)$, a contradiction to the optimality of G. □

Finally, we prove that B is an independent set.

Claim 54. *The set B is independent.*

Proof. Suppose that $u, v \in B$ are adjacent. There are at most $|C|$ non-triangular edges with an end in $\{u, v\}$, because every vertex can only be a non-triangular neighbour of at most one of u and v. Moreover, by definition, every vertex in B has a non-triangular neighbour. Let $w \in C$ be a non-triangular neighbour of u. Since the edge uw is non-triangular, it follows that u is not adjacent to any of the neighbours of w. By Claim 53, w is adjacent to all but at most $O(\varepsilon n)$ vertices in B. Therefore, u has at most $O(\varepsilon n)$ neighbours in B and, likewise, so does v.

Let G' be the graph obtained by replacing u and v with new vertices x and y where x is adjacent to all of $A \cup C$ and y is adjacent to all of $(A \cup B) \setminus \{u, v\}$. We have $t(G') \geq t(G)$ and $e(G') \geq e(G)+|B|-2-2|C|-O(\varepsilon n) > e(G)$, contradicting the optimality of G. Therefore, B is independent, as required. □

We have proved that B and C are independent, A is complete, and its vertices are adjacent to all of B and none of C. We may add any missing edges between B and C without creating new triangles, so by the optimality of G, there are in fact no missing edges between B and C. Therefore, G is isomorphic to $G(|A|, |B|, |C|)$, completing the proof of Theorem 6. □

7 Concluding remarks

We note that we have not fully resolved Conjecture 1.

Conjecture 1. *Let n and $e > \lfloor n^2/4 \rfloor$ be integers and let G be an n-vertex graph with e edges that minimises the number of triangular edges. Then G is isomorphic to a subgraph of a graph $G(a, b, c)$ for some a, b, c.*

Theorem 3 shows that the minimum number of triangular edges among n-vertex graphs with e is attained by (a subgraph of) a graph $G(a, b, c)$. However, we have not shown that such graphs are the only minimisers. Nevertheless, we believe that this fact can be proved (for sufficiently large n) by retracing our proofs. Since our paper is already quite long, we spare the reader any further details. In any case, our result only holds for sufficiently large n, and it would be interesting to extend it to work for all n.

We have not specified explicitly how large n should be in order for our proof to work, mainly because, due to the complexity of the proof, it is quite hard to find such an explicit bound. Nevertheless, we expect this bound to be 'reasonably small': say, much smaller than a bound that may arise from the use of the regularity lemma, because the inequalities we need to hold are polynomial in n.

The following question arises from Conjecture 1, by considering edges on K_r for $r \geq 4$. To simplify the notation, for any graph H and an edge e of some other graph G, we say that e is an *H-edge* if it is contained in a subgraph of G isomorphic to H.

Problem 55. What is the smallest number of K_r-edges that a graph with n vertices and e edges may have? Which graphs with n vertices and e edges minimise this quantity?

It seems reasonable to believe that the extremal examples are analogues of graphs $G(a, b, c)$, namely, they may be formed by adding a clique to one of the parts of a complete $(r-1)$-partite graph with n vertices. We believe that the methods used in this paper may be useful when tackling this more general problem.

There is another natural generalisation, where we consider odd cycles instead of cliques.

Problem 56. What is the smallest number of C_{2k+1}-edges that a graph with n vertices and e edges may have? Which graphs minimise this quantity?

It turns out that the case $k \geq 2$ is quite different from $k = 1$ (that is, where the odd cycle is a triangle). Erdős, Faudree and Rousseau [3] proved that, for any fixed $k \geq 2$, any graph with n vertices and $\lfloor n^2/4 \rfloor + 1$ edges has at least $11n^2/144 + O(n)$ C_{2k+1}-edges. In contrast, the number of triangular edges can be as small as $2\lfloor n/2 \rfloor + 1$, as mentioned in the introduction. So, the jump in the number of C_{2k+1}-edges (for $k \geq 2$) is very sharp, while the jump in the number of triangular edges is much smoother.

In the same paper, Erdős, Faudree and Rousseau conjectured a stronger statement: they conjectured that, for any fixed $k \geq 2$, any graph with n vertices and $\lfloor n^2/4 \rfloor + 1$ edges has at least $2n^2/9 + O(n)$ C_{2k+1}-edges. This bound can be attained by (a subgraph of) the union of a complete graph on roughly $2n/3$ vertices and a balanced complete bipartite graph on the remaining vertices. However, an example by Füredi and Maleki [4] shows that the conjecture is false: they constructed n-vertex graphs with $\lfloor n^2/4 \rfloor + 1$ edges and $(0.213\ldots + o(1))n^2$ C_5-edges. The example is somewhat similar to a graph $G(a, b, c)$: here we have four sets A, B, C, D such that A induces a clique and all possible $A - B$, $B - C$ and $C - D$ edges are present. The $C - D$ edges are not C_5-edges, but all other edges are. The aforementioned bound is obtained by optimising the sizes of A, B, C, D. Füredi and Maleki also calculated, asymptotically, the minimum possible number of C_{2k+1}-edges (for $k \geq 2$) in n-vertex graphs with e edges, where $e = \gamma n^2$ for any fixed constant $1/4 < \gamma < 1/2$. Their findings provided supporting evidence that the conjecture of Erdős, Faudree and Rousseau should be true for $k \geq 3$.

Very recently, more progress on Problem 56 was made by Grzesik, Hu and Volec [6]. For any fixed $k \geq 2$, they obtained asymptotically sharp bounds for the smallest possible number of C_{2k+1}-edges in a graph with n vertices and at least $\lfloor n^2/4 \rfloor + 1$ edges, using the method of flag algebras. In particular, they almost confirmed the conjecture of Erdős, Faudree and Rousseau for $k \geq 3$ (with an error term of $o(n^2)$ instead of $O(n)$) and proved that the construction of Füredi and Maleki is asymptotically best for $k = 2$.

We believe that the method of Grzesik, Hu and Volec [6] should be sufficient to give the exact smallest number of C_{2k+1}-edges in a graph with n vertices and e edges, for any fixed $k \geq 2$, provided that n is sufficiently large. Furthermore, their stability result should be sufficient to establish that, for sufficiently large n, the construction described earlier is the unique extremal construction. However, Grzesik, Hu and Volec do not claim these results in their paper and many technical details would have to be checked to make sure that these results could indeed be proved. Answering these questions without the assumption that n is large is an interesting problem, which is still open.

Finally, we mention that all aforementioned problems are special cases of the following very general problem.

Problem 57. Fix any graph F. What is the smallest possible number of F-edges in a graph with n vertices and e edges? What are the extremal examples?

Füredi and Maleki [5] calculated this minimum, asymptotically, for 3-chromatic graphs F and for $e = \gamma n^2$ where γ is fixed and satisfies $1/4 < \gamma < 1/2$. For any other F, this problem is wide open. Finally, we note that it is possible to go even further and generalise the problem to the context of hypergraphs.

Acknowledgements

We would like to thank Béla Bollobás for bringing this problem to our attention. The work on this project started during our stay in IMT Lucca, and we would like to thank the institute for their hospitality. We would also like to thank the referee for carefully reading our manuscript.

References

[1] P. Erdős, *Some theorems on graphs*, Riveon Lematematika **9** (1955), 13–17, in Hebrew.

[2] ———, *On a theorem of Rademacher-Turán*, Illinois J. Math **6** (1962), 122–127.

[3] P. Erdős, R. J. Faudree, and C. C. Rousseau, *Extremal problems involving vertices and edges on odd cycles*, Discr. Math. **101** (1992), no. 1–3, 23–31.

[4] Z. Füredi and Z. Maleki, *The minimum number of triangular edges and a symmetrization for multiple graphs*, preprint, arXiv:1411.0771.

[5] _____, *A proof and a counterexample for a conjecture of Erdős concerning the minimum number of edges on odd cycles*, manuscript.

[6] A. Grzesik, P. Hu, and J. Volec, *Minimum number of edges that occur in odd cycles*, preprint, arXiv:1605.09055.

[7] L. Lovász and M. Simonovits, *On the number of complete subgraphs of a graph*, Proc. Fifth British Combinatorial Conference (Aberdeen) (1975), 431–442.

[8] _____, *On the number of complete subgraphs of a graph II*, Studies in pure mathematics, Birkhäuser, Basel, 1983, pp. 459–495.

[9] W. Mantel, *Problem 28*, Wiskundige Opaven **10** (1907), 60–61.

[10] T. S. Motzkin and E. G. Straus, *Maxima for graphs and a new proof of a theorem of Turán*, Canad. J. Math. **17** (1965), 533–540.

[11] H. Rademacher, (1941), unpublished.

[12] A. Razborov, *On the minimal density of triangles in graphs*, Combin. Probab. Comput. **17** (2008), 603–618.

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