Counting Steiner Triple Systems

Peter Keevash*

Abstract. We prove a conjecture of Wilson from 1974 on the number of Steiner Triple Systems. The proof illustrates our method of Randomised Algebraic Construction, which we developed recently to resolve a question of Steiner from 1853 on the existence of combinatorial designs.

2010 Mathematics Subject Classification. Primary 05B07; Secondary 05D40
Keywords. Designs, Steiner Systems, Randomised Algebraic Construction

1. Introduction

A *Steiner system* with parameters (n, q, r) is a set S of q-subsets of an n-set[1] X, such that every r-subset of X belongs to exactly one element of S. The question of whether there is a Steiner system with given parameters is one of the oldest problems in combinatorics, dating back to work of Plücker (1835), Kirkman (1846) and Steiner (1853); see [16] for a historical account.

More generally, we say that a set S of q-subsets of an n-set X is a *design* with parameters (n, q, r, λ) if every r-subset of X belongs to exactly λ elements of S. There are some obvious necessary 'divisibility conditions' for the existence of such S, namely that $\binom{q-i}{r-i}$ divides $\lambda\binom{n-i}{r-i}$ for every $0 \leq i \leq r-1$ (fix any i-subset I of X and consider the sets in S that contain I). It is not known who first advanced the 'Existence Conjecture' that the divisibility conditions are also sufficient, apart from a finite number of exceptional n given fixed q, r and λ.

The case $r = 2$ has received particular attention because of its connections to statistics, under the name of 'balanced incomplete block designs'. The first result in this direction was obtained by Kirkman in 1847, who proved the Existence Conjecture for objects now known as Steiner Triple Systems, namely Steiner systems with parameters $(n, 3, 2)$. We refer the reader to [3] for a summary of the large literature and applications of this field. The Existence Conjecture for $r = 2$ was a long-standing open problem, eventually resolved by Wilson [17, 18, 19] in a series of papers that revolutionised Design Theory, and had a major impact in Combinatorics. In [6] we proved the Existence Conjecture in general, via a new method which we call Randomised Algebraic Constructions.

Our inductive proof requires us to prove a more general result, which can be roughly stated that dense pseudorandom simplicial complexes have clique decompositions. We will illustrate the method by considering the case of triangle decom-

*Research supported in part by ERC grant 239696.
[1] i.e. $|X| = n$ and S consists of subsets of X each having size q

positions of dense pseudorandom graphs. This case admits many simplifications, to the extent that we will be able to give much of the proof in this short article (the remaining technical details can be found in the full version [7]).

We will also use our result on triangle decompositions to prove the following estimate on the number $STS(n)$ of Steiner Triple Systems on n points, conjectured by Wilson [20].

Theorem 1.1. *$STS(n) = (n/e^2 + o(n))^{n^2/6}$ if n is 1 or 3 mod 6, otherwise 0.*

Our expository goal in this article is to provide an informal introduction to the ideas of [6], so we will be rather imprecise in places, leaving the reader who desires more formality to consult [7] or [6]. Furthermore, there are some additional simplifications available in the case of triangle decompositions (many of which were pointed out by an anonymous referee and other readers of an earlier version of this paper), which will be remarked on during the paper, but not implemented as they are specific to triangle (or graph) decompositions, and so not very helpful for understanding the general (hypergraph) case.

2. Triangle decompositions

Next we will state our result on triangle decompositions of dense pseudorandom graphs. In this case, the necessary divisibility conditions mentioned above show that the number of edges must be divisible by three, and the degree of any vertex must be even. We say that G is *tridivisible* if it satisfies these divisibility conditions. The pseudorandomness condition is as follows. Let G be a graph on n vertices. The *density* of G is $d(G) = |G|/\binom{n}{2}$. We say that G is *c-typical* if every vertex has $(1 \pm c)d(G)n$ neighbours and every pair of vertices have $(1 \pm c)d(G)^2 n$ common neighbours. (We write $b \pm c$ for any real between $b - c$ and $b + c$.)

Theorem 2.1. *There exist $0 < c_0 < 1$ and $n_0 \in \mathbb{N}$ so that if $n \geq n_0$ and G is a c-typical tridivisible graph on n vertices with $d(G) > n^{-10^{-7}}$ and $c < c_0 d(G)^{10^6}$ then G has a triangle decomposition.*

In this paper, we will sketch the proof of a slightly weaker theorem, using the following stronger notion of typicality from [6], from which it is not difficult to deduce Theorem 2.1 via standard 'Szemerédi Regularity' methods (see [7] for more remarks on this).

We say that G is *(c, h)-typical* if

$$|\cap_{x \in S} G(x)| = (1 \pm |S|c)d(G)^{|S|}n \quad \text{for any } S \subseteq V(G) \text{ with } |S| \leq h.$$

Note that being c-typical is essentially the same as being $(c, 2)$-typical (up to a factor of 2 in c).

Henceforth, we will assume that G is $(c, 16)$-typical.

2.1. The number of Steiner Triple Systems. The upper bound in Theorem 1.1 was recently proved by Linial and Luria [10], who showed that $STS(n) \leq (n/e^2 + O(\sqrt{n}))^{n^2/6}$ Our lower bound will be $STS(n) \geq (n/e^2 + O(n^{1-a}))^{n^2/6}$ for some small $a > 0$. The idea is to prove a lower bound on the number of 'almost' Steiner Triple Systems S such that Theorem 2.1 can be applied to the graph $K_n \setminus \bigcup S$ of uncovered edges, thus completing S to a (genuine) Steiner Triple System. It has been known since the pioneering work of Rödl [14] that almost Steiner Triple Systems (and almost designs) can be constructed by the semirandom method (nibble). Rather than using the classical nibble, it will be most convenient for us to apply the recent analysis of the triangle removal process by Bohman, Frieze and Lubetzky [2].

We will say that an event E holds *with high probability* (whp) if $\mathbb{P}(E) = 1 - e^{-\Omega(n^b)}$ for some $b > 0$ as $n \to \infty$. Note that when n is sufficiently large, by union bounds we can assume that any specified polynomial number of such events all occur (this point is not important in this section but will be used later in the paper).

In the triangle removal process, we start with the complete graph K_n, and at each step we delete the edges of a uniformly random triangle in the current graph. It is shown in [2] that whp the process persists until only $O(n^{3/2+o(1)})$ edges remain, but we will stop at $n^{2-10^{-7}}$ edges (i.e. at the nearest multiple of 3 to this number) so that we can apply Theorem 2.1. We need the following additional facts from [2] about this stopped process: whp the final graph is $n^{-1/3}$-typical, and when $pn^2/2$ edges remain the number of choices for the deleted triangle is $(1 \pm n^{-2/3})(pn)^3/6$.

Proof of Theorem 1.1. Consider the following procedure for constructing a Steiner Triple System on n vertices: run the triangle removal process until $n^{2-10^{-7}}$ edges remain, then apply Theorem 2.1 (if its hypotheses are satisfied, which occurs in $1-o(1)$ proportion of all instances of the process). Writing m for the number of steps and $p(i) = 1-6i/n^2$, the logarithm of the number of choices in this procedure is

$$L_1 = \sum_{i=1}^{m} (\log(p(i)^3 n^3/6) \pm 2n^{-2/3}) = (n^2/6)(\log(n^3/6) - 3 \pm n^{-10^{-8}}),$$

since $\sum_{i=1}^{m} \log p(i) = (1+O(n^{-10^{-7}} \log n))(n^2/6) \int_0^1 \log p \, dp$ and $\int_0^1 \log p \, dp = -1$. Also, for any fixed Steiner Triple System, the logarithm of the number of times it is counted by this procedure is at most

$$L_2 = \sum_{i=1}^{m} \log(p(i)n^2/6) = (n^2/6)(\log(n^2/6) - 1 \pm n^{-10^{-8}}).$$

Therefore $\log(STS(n)) \geq L_1 - L_2 = (n^2/6)(\log(n) - 2 \pm 2n^{-10^{-8}})$, which implies the stated bound on $STS(n)$. □

2.2. Strategy. Our strategy for obtaining a triangle decomposition of G can be thought of as variant of the well-known Absorbing Method (see the survey [15]). We begin by creating an 'absorbing' set of edge-disjoint triangles, which we call the *template* T. Next we extend T to an almost-perfect triangle decomposition of G by standard random greedy methods. Finally, the 'absorbing' property of the template allows us to rearrange its triangles in order to complete this to the desired perfect triangle decomposition (the 'absorption').

However, there is an important difference from standard applications that makes our setting more difficult. To explain this, we first note that the triangle decomposition problem can be reformulated as the perfect matching problem in an auxiliary 3-graph H, whose vertices are all edges of G, and whose edges are all $\{xy, yz, zx\}$ such that xyz is a triangle of G. In a typical application of the Absorbing Method to find a perfect matching in a k-graph H, the strategy is to show that any k-set $S \subseteq V(H)$ has many 'absorbers' A in H, meaning that A is a small matching such that there is another matching B with $\bigcup B = S \cup \bigcup A$. One then shows that if T is a random matching covering a small constant proportion of $V(H)$ then whp every k-set has many absorbers in H.

This sketch is plausible in dense settings, where for any S we typically have $\Theta(|V(H)|^{ak})$ absorbers in H with a edges, each of which appears in T with probability $\Theta(|V(H)|^{-a(k-1)})$, so whp $\Theta(|V(H)|^a)$ appear in T for any S. However, the auxiliary hypergraph for triangle decompositions is very sparse: it has $\Theta(n^2)$ vertices but only $\Theta(n^3)$ edges. If we were to choose T randomly then the probability for any fixed triangle to appear would be $O(n^{-1})$. On the other hand, to absorb some fixed (tridivisible) $S \subseteq E(G)$, we need T to contain a set A of a edge-disjoint triangles such that $S \cup A$ has a triangle decomposition B, so we need $\omega(n^a)$ such A in G.

To see that this is impossible, we imagine selecting the triangles of A one at a time and keeping track of the number E of edges that belong to a unique triangle of $S \cup A$. If a triangle uses a vertex that has not been used previously then it increases E, and otherwise it decreases E by at most 3. We can assume that no triangle is used in both A and B, so we terminate with $E = 0$. Thus there can be at most $3a/4$ steps in which E increases, so there are only $O(n^{3a/4})$ such A in G.

The key idea for circumventing this obstacle is to instead define T by randomly embedding $V(G)$ into a field and taking those triangles defined by a certain algebraic condition; this is the method of Randomised Algebraic Construction introduced in [6]. We let $G^* = \bigcup T$ be the underlying graph of T. We will show that G^* is typical with respect to G (it behaves like a random subgraph of G in a way we will define later) and also that G^* is 'linearly typical' (it has certain algebraic properties which we will define later). Obtaining this combination of random-like and algebraic properties is the key to the proof.

We will now sketch in more detail the various steps in constructing a triangle decomposition of G. Firstly, in the *Nibble* step, the typicality properties of G and G^* and the semirandom method give a set N of edge-disjoint triangles in $G \setminus G^*$ such that the *leave* $L := (G \setminus G^*) \setminus \bigcup N$ has 'small' maximum degree; in particular, $N \cup T$ is an almost-perfect triangle decomposition of G and L is a small fraction

of G.

More precisely, L will have maximum degree at most $c_1 n$, where $c \ll c_1 \ll c_2 \ll c_3 \ll d(G)$ are parameters defined below (recall that $d(G)$ is the density of G and G is $(c, 16)$-typical). It may be helpful at first to think of these parameters as absolute constants, although in our application to the proof of Theorem 1.1 we allow them to decay polynomially with n.

The remaining steps of the proof work towards absorbing L. Our goal is to find $A \subseteq T$ and a set B of edge-disjoint triangles such that $\bigcup B = L \cup \bigcup A$. Then $N \cup (T \setminus A) \cup B$ will be a triangle decomposition of G.

In the *Cover* step, we apply a random greedy algorithm to find a set M^c of edge-disjoint triangles which cover the leave L and whose edges are in $G^* \cup L$. Thus $N \cup M^c \cup T$ is a set of triangles which covers every edge of G, but which covers some edges twice, namely the *spill* $S = G^* \cap \bigcup M^c$. We use the typicality properties of G and G^* and the bounded maximum degree of L to show that S has maximum degree at most $c_2 n$.

In an ideal world, we would have $S = \bigcup A$ for some $A \subseteq T$; then $B = M^c$ would satisfy $\bigcup B = L \cup \bigcup A$, giving the triangle decomposition $N \cup (T \setminus A) \cup B$. However, this is too much to hope for, because of the sparsity of T (as discussed above). Instead, in the *Hole* step, we will find a set M^i of edge-disjoint triangles in G^*, such that $\bigcup M^i$ is edge-disjoint from S and $S \cup \bigcup M^i$ has a triangle decomposition M^o (and has maximum degree at most $c_3 n$). We think of M^i as the 'inner' set and M^o as the 'outer' set. We can also think of (M^o, M^i) as a decomposition of S in which we allow triangles to have 'signs' (positive for M^o, negative for M^i). This line of thought is helpful for understanding the proof, as a preliminary step is to obtain an even weaker decomposition in which we allow each triangle to have any integer weight (an idea introduced in [5] and [21]).

Again, in an ideal world, if we had $M^o \subseteq T$, we could take $A = M^o$ and $B = M^c \cup M^i$ to obtain a triangle decomposition $N \cup (T \setminus A) \cup B$. However, this is again too much to hope for, so now the algebraic properties of T will come into play, enabling us to make local rearrangements to include the triangles of M^o, one by one. For each triangle of M^o, we find a copy of $K_{8,8,8}$ in G^* which contains it and whose edges are decomposed into 64 triangles of T. Then we *shuffle*: we remove these 64 triangles of T and replace them with 63 triangles disjoint from the triangle from M^o (it is not hard to see that this is possible). Repeating this, we have a way of rearranging T in order to cover exactly the edges of G not covered by $N \cup M^c \cup M^i$. Note that it is in finding these special copies of $K_{8,8,8}$ that the algebraic structure is critical: typical sets of triangles (chosen, for instance, at random from G) of a similar density do not contain any such dense structures.

There is one final complication: we have to be able to find these copies of $K_{8,8,8}$, and we have to be able to do so edge-disjointly, otherwise earlier shuffles might affect our ability to perform later ones. This is possible if the triangles M^o happen to have certain algebraic properties, but this need not be the case. Thus, before looking for shuffles, we modify $M^c \cup M^i$ to find a 'nicer' set M_2 of edge-disjoint triangles: M_2 still covers the leave L and some edges of G^*, but now $G^* \cap \bigcup M_2$ has a triangle decomposition. This modification idea echoes that used in the *Hole*

step, in which we started with a weaker form of decomposition (integral) and then modified it to obtain a better decomposition (signed). Both these modifications are performed in the *Completion* step: which replaces M^c, M^o and M^i by other sets of triangles with the same properties, where M_1 plays the role of $M^c \cup M^i$, M_2 of M^o, and each triangle f of M_2 can be embedded in a small subgraph that has one triangle decomposition (part of M_4) using f and another triangle decomposition (part of M_3) contained in T.

We can encapsulate the above discussion of the proof strategy as follows. We say that $J \subseteq G$ is *c-bounded* if $|J(v)| < c|V(G)|$ for every $v \in V(G)$, where $J(v) = \{u \in V(G) : uv \in J\}$ is the *neighbourhood* of v in J.

Strategy 2.2. Suppose we have $G^* \subseteq G$ with a 'template' triangle decomposition T such that

Nibble $G \setminus G^*$ contains a set N of edge-disjoint triangles with 'leave' $L := (G \setminus G^*) \setminus \bigcup N$ that is c_1-bounded,

Cover For any $L \subseteq G \setminus G^*$ that is c_1-bounded, there is a set M^c of edge-disjoint triangles in G such that $L = (G \setminus G^*) \cap (\bigcup M^c)$ and the 'spill' $S := G^* \cap (\bigcup M^c)$ is c_2-bounded,

Hole For any tridivisible $S \subseteq G^*$ that is c_2-bounded, there are 'outer' and 'inner' sets M^o, M^i of edge-disjoint triangles in G^* such that $\bigcup M^o$ is c_3-bounded and $(S, \bigcup M^i)$ is a partition of $\bigcup M^o$,

Completion We can modify L, M^c, M^o and M^i to obtain sets M_1, M_2, M_3, M_4 of edge-disjoint triangles in G^* such that $(L, \bigcup M_2)$ is a partition of $\bigcup M_1$, $\bigcup M_3 = \bigcup M_4$, $M_3 \subseteq T$ and $M_2 \subseteq M_4$.

The key step is choosing T (which determines G^*). To motivate the construction, suppose that $V(G)$ is an abelian group, and consider the set Σ of triples xyz such that $x + y + z = 0$. We note that Σ is a good 'model' for a triangle decomposition, as for any xy there is a unique z such that $x + y + z = 0$. However, we cannot simply take Σ, as not all such xyz are triangles of G; moreover, x, y, z may not even be pairwise distinct. The idea is that a suitable random subset of Σ can act as a template, which covers a large fraction of G (more precisely, it has density $\Theta(d(G)^2)$ in G).

It is not hard to see that G contains a triangle decomposition under the assumptions of Strategy 2.2. Indeed, we start by taking the sets N provided by **Nibble** and then the sets M^c and S provided by **Cover**. Now we note that $S = \bigcup T + \bigcup N + \bigcup M^c - G$ is tridivisible, as any integer linear combination of tridivisible graphs is tridivisible. So we can apply **Hole** to obtain M^o and M^i. Then we can apply **Completion** to obtain M_1, M_2, M_3, M_4. Finally, $M = N \cup M_1 \cup (M_4 \setminus M_2) \cup (T \setminus M_3)$ is a triangle decomposition of G. In the remainder of the paper we will sketch the various steps of Strategy 2.2.

2.3. Template. We choose the template as follows.

Construction 2.3. Let $a \in \mathbb{N}$ be such that $2^{a-2} < |V(G)| \leq 2^{a-1}$. Let $\pi : V(G) \to \mathbb{F}_{2^a} \setminus \{0\}$ be a uniformly random injection. Let

$$T = \{xyz \in K_3(G) : \pi(x) + \pi(y) + \pi(z) = 0\} \text{ and } G^* = \bigcup T.$$

In this subsection we establish the typicality properties of G^*, deferring discussion of the algebraic properties of T until they are needed in Section 4. We adopt this organisation for expository purposes, but note that we could equally well have proved all properties of T that we need later hold whp before proceeding to the other steps of Strategy 2.2, so one can imagine that T is fixed from the start with these properties.

We start with some notation and preliminary observations. Throughout we write $n = |V(G)|$. We identify G with its edge set $E(G)$, so that $|G|$ denotes the number of edges of G (rather than the number of vertices, as is used by some authors). We write $K_s(G)$ for the set of copies of K_s in G. We write $[n] = \{1, \dots, n\}$. We define

$$\gamma = 2^{-a}n,$$

and note that $1/4 < \gamma < 1/2$. We observe that if $x, y, z \in \mathbb{F}_{2^a} \setminus \{0\}$ and $x+y+z = 0$ then x, y, z are pairwise distinct. We note that $+1 = -1$ in $\mathbb{F}_{2^a}$, so we can use $+$ and $-$ interchangeably in $\mathbb{F}_{2^a}$-arithmetic. We consider $\mathbb{F}_{2^a}$ as a vector space over $\mathbb{F}_2$, and observe that any two nonzero elements span a subspace of dimension two.

Now we define the typicality condition for (G, G^*) and show that it holds whp. Let G^* be a subgraph of G. We say that (G, G^*) is (c, h)-*typical* if

$$|\bigcap_{x \in S^*} G^*(x) \cap \bigcap_{x \in S \setminus S^*} G(x)| = (1 \pm |S|c)d(G^*)^{|S^*|}d(G)^{|S|-|S^*|}n$$

for any $S^* \subseteq S \subseteq V(G)$ with $|S| \leq h$.

Lemma 2.4. *whp $d(G^*) = (1 \pm 3c)\gamma d(G)^3$ and (G, G^*) is $(6c, 16)$-typical.*

The proof uses the following consequence of Azuma's inequality.

Definition 2.5. Let S_n be the symmetric group, $f : S_n \to \mathbb{R}$ and $b \geq 0$. We say that f is *b-Lipschitz* if for any $\sigma, \sigma' \in S_n$ such that $\sigma = \tau \circ \sigma'$ for some transposition $\tau \in S_n$ we have $|f(\sigma) - f(\sigma')| \leq b$.

For a proof of the following lemma, see e.g. the discussion after Theorem 3.7 in [13].

Lemma 2.6. *Suppose $f : S_n \to \mathbb{R}$ is b-Lipschitz, $\sigma \in S_n$ is uniformly random and $X = f(\sigma)$. Then*

$$\mathbb{P}(|X - \mathbb{E}X| > t) \leq 2e^{-t^2/2nb^2}.$$

Proof of Lemma 2.4. We start by estimating $\mathbb{E}|G^*| = \sum_{e \in G} \mathbb{P}(e \in G^*)$. For any $e = xy$, given $\pi(x)$ and $\pi(y)$, we have $e \in G^*$ if and only if $\pi(z) = \pi(x) + \pi(y)$ for some z such that $xyz \in K_3(G)$. Since G is $(c, 16)$-typical, there are $(1 \pm 2c)d(G)^2n$ choices for z. Each satisfies $\pi(z) = \pi(x) + \pi(y)$ with probability

$(2^a-3)^{-1}$, so $\mathbb{E}|G^*| = |G|(1\pm 2c)d(G)^2n(2^a-3)^{-1}$. We can view π as $\sigma\circ\pi_0$, where $\pi_0 : V(G)\to \mathbb{F}_{2^a}\setminus\{0\}$ is any fixed injection and σ is a random permutation of $\mathbb{F}_{2^a}\setminus\{0\}$. Any transposition of σ affects $|G^*|$ by $O(n)$, so by Lemma 2.6 whp $d(G^*) = (1\pm 2.1c)\gamma d(G)^3$.

Similarly, if $S^*\subseteq S\subseteq V(G)$ with $|S|\leq 16$, we write $Y = \bigcap_{x\in S^*} G^*(x)\cap \bigcap_{x\in S\setminus S^*} G(x)$, and estimate $\mathbb{E}|Y| = \sum_{y\in V(G)}\mathbb{P}(y\in Y)$. For any $y\in \cap_{x\in S}G(x)$, given $\pi(y)$ and $\pi(x)$ for all $x\in S$, we have $y\in Y$ if and only if for all $x\in S^*$ there is $xyz_x\in K_3(G)$ such that $\pi(z_x) = \pi(x)+\pi(y)$. Since G is $(c,16)$-typical, there are $(1\pm|S|c)d(G)^{|S|}n$ choices for y. By excluding $O(1)$ choices of y we can assume $\pi(x)+\pi(y)\neq\pi(x')$ for all $x,x'\in S$. Then for each $x\in S^*$ there are $(1\pm 2c)d(G)^2n$ choices for z_x, and for any set of choices, with probability $(1+O(1/n))2^{-a|S^*|}$ they all satisfy $\pi(z_x) = \pi(x)+\pi(y)$. This gives

$$\mathbb{E}|Y| = O(1) + (1\pm|S|c)d(G)^{|S|}n\cdot((1\pm 2c)d(G)^2n)^{|S^*|}\cdot(1+O(1/n))2^{-a|S^*|}.$$

Any transposition of σ fixing every element of S affects $|Y|$ by $O(1)$, so by Lemma 2.6 applied on $[n]\setminus S$ whp $|Y| = (1\pm(3|S|+1)c)d(G)^{|S|}(\gamma d(G)^2)^{|S^*|}n = (1\pm 6|S|c)d(G^*)^{|S|}d(G)^{|S|-|S^*|}n$. □

Henceforth, we assume that G^* has been chosen to satisfy the conclusion of Lemma 2.4.

2.4. Nibble. To implement the *Nibble* step, we will show that the following theorem can be applied with $H = G\setminus G^*$.

Theorem 2.7. *There are $b_0>0$ and $n_0\in\mathbb{N}$ so that if $n>n_0$, $n^{-0.1}<b<b_0$ and H is a b-typical graph on n vertices with $d(H)>b$, then there is a set N of edge-disjoint triangles in H such that $L = H\setminus\bigcup N$ is $b^{1/4}$-bounded.*

We remark that the parameters in Theorem 2.7 are not very sharp: we have just fixed some convenient values that suffice for our purposes. Similar results are well-known, but we are not aware of any reference that implies the theorem as stated, so we sketch a proof in [7].

To apply the theorem we show that $G\setminus G^*$ is $50c$-typical. First we recall that (G,G^*) is $(6c,16)$-typical and note that as $d(G^*) = (1\pm 3c)\gamma d(G)^3$ and $1/4<\gamma<1/2$ we have $0.24d(G)^3<d(G^*)<0.51d(G)$ for small c. Now, for any $v\in V(G)$ we have

$$\begin{aligned}|(G\setminus G^*)(v)| &= (1\pm c)d(G)n - (1\pm 6c)d(G^*)n\\ &= (d(G)-d(G^*))n\pm 6c(d(G)+d(G^*))n = (1\pm 20c)d(G\setminus G^*)n.\end{aligned}$$

Furthermore, for any $u,v\in V(G)$ we estimate $|(G\setminus G^*)(u)\cap(G\setminus G^*)(v)|$ as

$$\begin{aligned}&|G(u)\cap G(v)| - |G^*(u)\cap G(v)| - |G(u)\cap G^*(v)| + |G^*(u)\cap G^*(v)|\\ &= (1\pm 2c)d(G)^2n - 2(1\pm 12c)d(G)d(G^*)n + (1\pm 12c)d(G^*)^2n\\ &= (d(G)-d(G^*))^2n\pm 12c(d(G)+d(G^*))^2n = (1\pm 50c)d(G\setminus G^*)^2n.\end{aligned}$$

Thus $G \setminus G^*$ is $50c$-typical, so we obtain *Nibble* with $c_1 = (50c)^{1/4}$.

We give here the values of some other parameters that will be used in the paper (we have already mentioned c_1, c_2, c_3 in Strategy 2.2; c_4 and c_5 will be used in the *Completion* step):

$$c_1 = (50c)^{1/4}, \quad c_2 = 10^2 c_1 d(G)^{-6}, \quad c_3 = 10^{20} c_2 d(G)^{-50},$$
$$c_4 = 10^{20} c_3 d(G)^{-100}, \quad c_5 = 10^{10} c_4 d(G)^{-180}.$$

The tightest constraint on c that will be required in our calculations is $100c_5 = 10^{54}(50c)^{1/4} d(G)^{-336} < 10^{-6} d(G)^{180}$; this holds for small c_0 if $c < c_0 d(G)^{3000}$. (This is the bound we need if G is $(c, 16)$-typical, but if G is c-typical we need the stronger bound in Theorem 2.1.)

2.5. Cover. Recall that in the *Cover* step we want to choose a set M^c of edge-disjoint triangles in $G^* \cup L$ which cover the leave L, such that $S = G^* \cap \bigcup M^c$ is c_2-bounded. This can be most easily achieved by a deterministic greedy algorithm, but in keeping with our goal of illustrating the ideas of [6], we will use the following random greedy algorithm that can be applied in more general settings.

Algorithm. Let $L = \{e_i : i \in [t]\}$ (with edges ordered arbitrarily). Let $M^c = \{T_i : i \in [t]\}$ be triangles such that T_i consists of e_i and two edges of G^*, and is chosen uniformly at random from all such triangles that are edge-disjoint from all previous choices; if there is no available choice for T_i then the algorithm aborts.

To analyse the algorithm we will use a concentration inequality. We say that a random variable Y is (μ, C)-*dominated*, if there are constants $\mu_1, \dots, \mu_m$ with $\sum_{i=1}^m \mu_i < \mu$, and we can write $Y = \sum_{i=1}^m Y_i$, such that $|Y_i| \le C$ for all i, and conditional on any given values of Y_j for $j < i$ we have $\mathbb{E}|Y_i| < \mu_i$. The following lemma follows easily (see [6, Lemma 2.7]) from Freedman's inequality [4] (or from Hoeffding's inequality and a coupling argument, as noted by a referee).

Lemma 2.8. *If Y is (C, μ)-dominated then*

$$\mathbb{P}(|Y| > (1+c)\mu) < 2e^{-\mu c^2/2(1+2c)C}.$$

For the following lemma, we recall that G^* satisfies the conclusion of Lemma 2.4, L is c_1-bounded, where $c_1 = (50c)^{1/4}$, and $c_2 = 10^2 c_1 d(G)^{-6}$.

Lemma 2.9. *whp the algorithm to choose M^c does not abort, and $S := G^* \cap (\bigcup M^c)$ is c_2-bounded.*

Proof. For $i \in [t]$ we let $\mathcal{B}_i$ be the bad event that $S_i := G^* \cap (\cup_{j<i} T_j)$ is not c_2-bounded. We define a stopping time τ be the smallest i for which $\mathcal{B}_i$ holds or the algorithm aborts, or ∞ if there is no such i. It suffices to show whp $\tau = \infty$.

We fix $t_0 \in [t]$ and bound $\mathbb{P}(\tau = t_0)$ as follows. For any $i < t_0$, since $\mathcal{B}_i$ does not hold, S_i is c_2-bounded. Writing $e_i = v_i v_i'$, we can bound the number of excluded choices for T_i by $c_2 n < |G^*(v_i) \cap G^*(v_i')|/2$, so at most one half of the triangles on e_i are excluded.

Next we fix $e = vv' \in G^*$, and estimate $r_e := \sum_{i \le t_0} \mathbb{P}'(e \subseteq T_i)$, where $\mathbb{P}'$ denotes the conditional probability given the choices made before step i. We compare r_e to the expected number of times that e would be covered if we chose all triangles independently. To be precise, we let

$$E_e := \sum_{i \le t_0} \mathbb{P}(e \subseteq T_i'),$$

where each T_i' is a uniform random triangle consisting of e_i and two edges of G^*, and $(T_i' : i \in [t])$ are independent. By the bound on excluded choices, $\mathbb{P}'(e \subseteq T_i) < 2\mathbb{P}(e \subseteq T_i')$, so $r_e < 2E_e$.

The ith summand in E_e is only nonzero when $e_i \cap e \neq \emptyset$. As L is c_1-bounded, the number of such i is at most $|L(v)| + |L(v')| < 2c_1 n$. Also, for each i such that $e_i \cup e$ spans a triangle, we have

$$\mathbb{P}(e \subseteq T_i') = |G^*(v_i) \cap G^*(v_i')|^{-1} < 2d(G^*)^{-2}n^{-1}.$$

Therefore $E_e < 4c_1 d(G^*)^{-2} < c_2/4$.

Finally, fix $v \in V(G)$ and consider $X = |S_{t_0}(v)| = \sum_{i \le t_0} X_i$, where $X_i = \sum_{v \in e \in G^*} 1_{e \subseteq T_i}$. We have $|X_i| \le 2$ and

$$\sum_{i \le t_0} \mathbb{E}'(X_i) = \sum_{i \le t_0} \sum_{v \in e \in G^*} \mathbb{P}'(e \subseteq T_i) = \sum_{v \in e \in G^*} r_e < c_2 n/2.$$

By Lemma 2.8 we have $\mathbb{P}(X \ge c_2 n) < 2e^{-c_2 n/24}$. Taking a union bound over $i \le t_0 \le t$, whp $|S(v)| < c_2 n$, i.e. S is c_2-bounded and $\tau = \infty$. □

2.6. Random greedy algorithms. Below we will require several more random greedy algorithms similar to that used in *Cover*, so for future reference we now make some further comments on the proof of Lemma 2.9. One could formulate an abstract general lemma to cover all cases (see [6, Lemma 4.11]), but here we will prefer the more intuitive approach of identifying the key principles of the proof, so that it will be clear how it may be adapted to future instances. For a general random greedy algorithm, we identify some desired boundedness conclusion, then at each step of the algorithm, assuming that boundedness has not failed, we show that at most one half (say) of the choices of the required configuration have been excluded. Then for each edge e in the underlying graph H we estimate the expected number E_e of times that e would be covered if we chose all configurations independently. If $E_e < b/4$ and the configurations have constant size (not depending on n) then the graph of all covered edges is whp b-bounded.

We also note for future reference some estimates that are useful for such arguments. Suppose H is a small fixed graph ($|H| \le 500$ say), $F \subseteq V(H)$ and ϕ is an embedding of $H[F]$ in G^*. We call $E = (\phi, F, H)$ an *extension*. Let $X_E(G^*)$ be the number of embeddings ϕ^* of H in G^* that restrict to ϕ on F. We suppose that E is 16-*degenerate*, meaning that we can construct the embedding one vertex at a time, so that at each step we add a vertex adjacent to at most 16 existing vertices. As (G, G^*) is $(6c, 16)$-typical, when we add a vertex adjacent to $t \le 16$

existing vertices, there are $(1 \pm 6tc)d(G^*)^t n$ choices. Multiplying these estimates, we obtain the following estimate for $X_E(G^*)$.

Lemma 2.10. *Suppose $E = (\phi, F, H)$ is a 16-degenerate extension with $|H| \leq 500$. Then*

$$X_E(G^*) = (1 \pm 7|H|c)d(G^*)^{|H \setminus H[F]|} n^{|V(H)|-|F|}.$$

Now suppose that we wish to exclude embeddings ϕ^* that use some edge in J, which is c-bounded. Fix $e \in H \setminus H[F]$ and consider the embeddings ϕ^* with $\phi^*(e) \in J$. If $e \cap F \neq \emptyset$ there are at most cn choices for the embedding of e then at most $n^{|V(H)|-|F|-1}$ choices for the remainder of ϕ^*. If $e \cap F = \emptyset$ there are at most cn^2 choices for the embedding of e then at most $n^{|V(H)|-|F|-2}$ choices for the remainder of ϕ^*. Thus at most $|H|cn^{|V(H)|-|F|}$ choices of ϕ^* are excluded, which is a negligible fraction of $X_E(G^*)$.

3. Hole

In this section we establish **Hole**. Our first step is to consider an integral relaxation, in the following sense. Instead of thinking of $(S, \bigcup M^i)$ as a partition of $\bigcup M^o$, we think of S as a weighted sum of edge sets of triangles, where triangles in M^o have weight 1 and triangles in M^i have weight -1. We can express this by the equation $\Phi A = S$, where Φ is the corresponding ± 1-vector indexed by triangles, and A is the inclusion matrix of triangles against edges, i.e. $A_{fe} = 1_{e \subseteq f}$ for any edge e and triangle f. It is straightforward to show that this equation has a solution if we allow Φ to have any integer weights on triangles (see [5, 21, 22] for more general results). However, we also need to control the maximum degree of the multigraph formed by these triangles, so that we can continue to apply random greedy algorithms as discussed in the previous section; in particular, we will apply such an algorithm later in this section to convert the integral decomposition into the signed decomposition required by **Hole**.

First we set up some more notation. It will be more convenient to work with linear maps rather than matrices. For any graph H we define $\mathbb{Z}$-linear boundary/shadow maps $\partial_j : \mathbb{Z}^{K_i(H)} \to \mathbb{Z}^{K_j(H)}$ for $i \geq j \geq 0$ by $\partial_j(e) = \sum \binom{e}{j}$ for $e \in K_i(H)$, i.e. for $J \in \mathbb{Z}^{K_i(H)}$ and $f \in K_j(H)$ we define $\partial_j(J)_f = \sum_{f \subseteq e \in K_i(H)} J_e$. For example, if $J \in \mathbb{Z}^H$ then $\partial_1(J) \in \mathbb{Z}^{V(H)}$ (identifying H with $K_2(H)$ and $V(H)$ with $K_1(H)$) is defined by $\partial_1(J)_v = \sum_{v \in e \in H} J_e$.

It will also be convenient to identify vectors with (generalised) sets. It is standard to identify $v \in \{0,1\}^X$ with the set $\{x \in X : v_x = 1\}$. Similarly, we can identify $v \in \mathbb{N}^X$ with the multiset in X in which x has multiplicity v_x (for our purposes $0 \in \mathbb{N}$). We also apply similar notation and terminology as for multisets to vectors $v \in \mathbb{Z}^X$ ('intsets'). Here our convention is that 'for each $x \in v$' means that x is considered $|v_x|$ times in any statement or algorithm, and has a sign attached to it (the same as that of v_x); we also refer to x as a 'signed element' of v. For $v \in \mathbb{Z}^X$ we write $v = v^+ - v^-$, where $v^+_x = \max\{v_x, 0\}$ and $v^-_x = \max\{-v_x, 0\}$ for $x \in X$. Given $J \in \mathbb{N}^G$ and $v \in V(G)$, we define $J(v) \in \mathbb{N}^{V(G)}$ by $J(v)_u = 1_{uv \in G} J_{uv}$. Then

we can extend the definition of boundedness to multigraphs: J is c-bounded if $|J(v)| < cn$ for every $v \in V(G)$.

3.1. Integral decomposition. Our integral relaxation of **Hole** is expressed by the following lemma (in which K_n denotes the complete graph on $V(G)$, and we recall that S is c_2-bounded); for **Hole** we will need the additional properties that $\Phi(f) = 0$ for any $f \in K_3(K_n) \setminus K_3(G^*)$, and $\Phi(f) \in \{0, 1, -1\}$ for all $f \in K_3(G^*)$, as then we can write $\Phi = M^o - M^i$.

Lemma 3.1. *There is $\Phi \in \mathbb{Z}^{K_3(K_n)}$ with $\partial_2\Phi = S$ such that $\partial_2\Phi^+$ is $100c_2$-bounded.*

Proof. We will construct $\Phi = \Phi_0 + \Phi_1 + \Phi_2$ such that $J^0 = S - \partial_2\Phi_0$, $J^1 = J^0 - \partial_2\Phi_1$, $J^2 = J^1 - \partial_2\Phi_2$ satisfy $\partial_i J^i = 0$ for $i = 0, 1, 2$. In words, we reduce to zero the sum of all values, then the sum of all values at any given vertex, and finally the sum of all values (i.e. the value) at any given edge. Recalling that S is tridivisible, each J^i will be tridivisible, in the 'intgraph' sense: i.e. $\sum_e J^i_e$ is divisible by 3 and $\sum_u J^i_{uv}$ is divisible by 2 for all v.

Step 0: For Φ_0, we choose $|S|/3$ independent uniformly random triangles in K_n; then $J^0 = S - \partial_2\Phi_0$ satisfies $\partial_0 J^0 = 0$. (Note that $\partial_0 J^0 = \sum_e J^0_e$.) For each vertex v, the number of these triangles containing v is binomial with mean $|S|/n < c_2 n/2$, so by the Chernoff bound whp $\partial_2\Phi_0$ is $1.1c_2$-bounded.

Step 1: We let $J^* = \partial_1 J^0$, so $\partial_0 J^* = 2\partial_0 J^0 = 0$, i.e. $|J^{*+}| = |J^{*-}|$. Note for all $x \in V(G)$ that J^*_x is even, as J^0 is tridivisible, and $|J^*_x| < 1.1c_2 n$. We fix an arbitrary sequence $((x_i^+, x_i^-) : i \in [|J^{*+}|/2])$ so that each $x \in V(G)$ occurs $J^{*+}_x/2$ times as some x_i^+ and $J^{*-}_x/2$ times as some x_i^-. For each i we choose $a_i b_i \subseteq V(G) \setminus \{x_i^+, x_i^-\}$ independently uniformly at random, and let $\Phi_1 = \sum_{i \in [|J^{*+}|/2]} (\{x_i^+ a_i b_i\} - \{x_i^- a_i b_i\})$; then $J^1 = J^0 - \partial_2\Phi_1$ satisfies $\partial_1 J^1 = 0$.

We claim that whp $\partial_2\Phi_1^\pm$ are $8c_2$-bounded. To see this, we first fix any $e \in K_n$ and estimate the expected contributions to e from each step i, according to whether e contains x_i^+, x_i^-, or neither. Each endpoint of e occurs at most $0.6c_2 n$ times as $x_i^\pm$, and for such i we cover e with probability $2/(n-2)$, so the expected contribution to $(\partial_2\Phi_1^\pm)_e$ from all such i is at most $2.5c_2$. At any other step, we cover e with probability $\binom{n-2}{2}^{-1}$, so the total expected contribution to $(\partial_2\Phi_1^\pm)_e$ from these steps is at most $1.1c_2$. Now, for each vertex v, summing over its incident edges, $|\partial_2\Phi_1^\pm(v)|$ are both $(4c_2 n, 1)$-dominated, so the claim holds by Lemma 2.8.

Step 2: We start by fixing an arbitrary expression $J^1 = \sum_{C \in \mathcal{C}_0} C$, where each C is a closed walk in G^* with edge weights alternating between 1 and -1, and there are no cancellations, i.e. every edge appears in the sum only with weight 1 or only with weight -1. As is well-known, such an expression may be found by a greedy algorithm: each C can be obtained by following an arbitrary alternating walk on the signed elements of J^1 until we return to our starting point using an edge with the opposite sign to that of the first edge, whereupon we add $-C$ to J^1 and repeat the procedure.

Next we express each $C \in \mathcal{C}_0$ as a sum of signed four-cycles in the complete graph K_n on $V(G)$, where we write each closed walk of length $2m$ as a chain of $m-1$ signed four-cycles, using the identity

$$\sum_{i=1}^{m-1}(-1)^i(\{x_ix_{i+1}\}-\{x_{i+1}y_{i+1}\}+\{y_{i+1}y_i\}-\{y_ix_i\})$$
$$=\{x_1y_1\}+(-1)^m\{x_my_m\}+\sum_{i=1}^{m-1}(-1)^i\{x_ix_{i+1}\}+\sum_{i=1}^{m-1}(-1)^i\{y_iy_{i+1}\}.$$

This identity can be used as is if $x_i \neq y_i$ for $i \in [m]$. For each i such that $x_i = y_i$, we note that $1 < i < m$, $x_{i-1} \neq y_{i-1}$, $x_{i+1} \neq y_{i+1}$, and $x_{i+1} \neq y_{i-1}$, so we can replace the four-cycles for summands $i-1$ and i by

$$(-1)^{i-1}(\{x_{i-1}x_i\}-\{x_ix_{i+1}\}+\{x_{i+1}y_{i-1}\}-\{y_{i-1}x_{i-1}\}, \text{ and}$$
$$(-1)^i(\{x_{i+1}y_{i-1}\}-\{y_{i-1}y_i\}+\{y_iy_{i+1}\}-\{y_{i+1}x_{i+1}\}).$$

Thus we can write $J^1 = \sum_{C\in\mathcal{C}} C$, where each summand is a signed four-cycle in K_n. Furthermore, the above construction has the property that for each $v \in V(G)$ and $w \in \{-1,1\}$ we use at most $3|J^{1+}(v)| < 24c_2n$ edges at v with weight w.

For each $C = \{ab\}-\{bc\}+\{cd\}-\{da\} \in \mathcal{C}$ we choose $x \in V(G)\setminus\{a,b,c,d\}$ independently uniformly at random, and add $\{xab\}-\{xbc\}+\{xcd\}-\{xda\}$ to Φ_2; then $\partial_2\Phi_2 = \sum_{C\in\mathcal{C}} C = J^1$. Let Γ denote the multigraph formed by summing $\{xa, xb, xc, xd\}$ over all such C. For any $e \in K_n$, at most $48c_2n$ elements of $\mathcal{C}$ can contribute to Γ_e, so $\mathbb{E}\Gamma_e < 49c_2n$. Then for any v, summing over its incident edges, $|\Gamma(v)|$ is $(49c_2n, 4)$-dominated, so by Lemma 2.8 (modified) whp Γ is $50c_2$-bounded. Defining $\Phi = \Phi_0 + \Phi_1 + \Phi_2$, we have $\partial_2\Phi = S$ and $\partial_2\Phi^+$ is $100c_2$-bounded. □

We note that the argument given in Step 2 does not generalise to hypergraph decompositions, where the corresponding arguments in [6] are considerably more difficult. As noted by a referee, we could have proved Lemma 3.1 more directly by repeatedly using triangles to shortcut walks in S, but we prefer the proof given here, as it at least shares some features of the arguments in [6].

3.2. Signed decomposition. To obtain the signed decomposition required by **Hole**, we will modify the integral decomposition Φ obtained in the previous subsection using the following 'octahedral' configurations. Consider a copy of $K_{2,2,2}$, the complete tripartite graph with 2 points in each part, with parts $\{(j,0),(j,1)\}$ for $j \in [3]$. We denote its triangles by $\{f_x : x \in \{0,1\}^3\}$, where $f_x = \{(j,x_j) : j \in [3]\}$. The sign of f_x is $s(f_x) = (-1)^{\sum x}$. Thus each edge is in one triangle of each sign. Defining $\Omega = \sum_{x\{0,1\}^3} s(f_x)\{f_x\} \in \mathbb{Z}^{K_3(K_{2,2,2})}$, we see that $\partial_2\Omega = 0$. This gives a method to eliminate any signed triangle f from Φ without altering $\partial_2\Phi$: we add some copy of Ω with the opposite sign to f in which (say) $f_{000} = f$, thus replacing f by seven other signed triangles that have the same total 2-shadow. Similarly (and more importantly), we can eliminate any pair of triangles f, f' that have opposite sign and share an edge e, replacing f, f' by six other signed triangles that have

the same total 2-shadow and do not use e. We apply this method in the following two-phase algorithm.

Octahedral Elimination Algorithm (Phase I). We eliminate all triangles in Φ, according to a random greedy algorithm, where in each step we consider some original signed element f of Φ, and choose an octahedral configuration Ω_f to replace f. We refer to edges of Ω_f not in f as *new* edges, and choose Ω_f uniformly at random subject to the new edges belonging to G^* and being disjoint from $\partial_2\Phi^+$ and all new edges from previous steps.

Let Φ' denote the result of Phase I (if it does not abort). Then $\partial_2\Phi' = \partial_2\Phi = S$, and we can write $\partial_2\Phi'^+ = \partial_2\Phi^+ + \Gamma$, where Γ is the graph of new edges, and every signed element of Φ' contains at most one edge of $\partial_2\Phi^+$.

Octahedral Elimination Algorithm (Phase II). We replace all signed edges apart from those in S and Γ. To do this, we fix a sequence $\mathcal{S}$ of pairs of signed elements of Φ', so that (i) for each $ff' \in \mathcal{S}$, there is some $e \in \partial_2\Phi^+$ such that f and f' both contain e, and f and f' have opposite signs, and (ii) the multiset consisting of all e as in (i) is $\partial_2\Phi^-$. Now we eliminate each $ff' \in \mathcal{S}$, according to a random greedy algorithm, by subtracting some copy $\Omega_{ff'}$ of Ω with $f_{000} = f$ and $f_{001} = f'$, or vice versa, depending on the signs. We refer to edges of $\Omega_{ff'}$ not in f or f' as *new* edges, and choose $\Omega_{ff'}$ uniformly at random subject to the new edges belonging to G^* and being distinct from $\partial_2\Phi^+ \cup \Gamma$ and all new edges from previous steps.

Let Ψ denote the result of this algorithm (if it does not abort) and Γ' the graph of new edges for Phase II. Then $\partial_2\Psi = S$ and $\partial_2\Psi^- = \Gamma \cup \Gamma' \subseteq G^*$. This implies $\Psi(f) = 0$ for any $f \in K_3(K_n) \setminus K_3(G^*)$, and $\Psi(f) \in \{0, 1, -1\}$ for all $f \in K_3(G^*)$, so $\Psi = M^o - M^i$, where M^o and M^i are as in **Hole**, once we have verified the boundedness condition.

Lemma 3.2. *whp the Octahedral Elimination Algorithm produces M^o and M^i as in **Hole**.*

Sketch proof. We will indicate how to analyse the algorithm in a similar way to the proof of Lemma 2.9, following the discussion after the proof of that Lemma 2.9. Recall that at each step of the algorithm, assuming that boundedness has not failed, we want to show that the number of excluded configurations is less than half of the total; then it suffices to estimate the expected number E_e of times that any given edge e would be covered if we chose all configurations independently.

We first show that whp Γ is c_2'-bounded, where $c_2' = 10^5 c_2 d(G^*)^{-9}$. Here a configuration for f consists of the new edges of some Ω_f. By Lemma 2.10, at each step, the number of choices of Ω_f with all new edges belonging to G^* (with no excluded configurations) is $(1 \pm 60c)d(G^*)^9 n^3$. Assuming that the graph of previous new edges is c_2'-bounded, as $\partial_2\Phi^+$ is $100c_2$-bounded, the number of excluded configurations is at most $10c_2' n^3$, which is less than half of the total. Next, for each $e \in G^*$, we consider separately the contributions to E_e, according to whether e intersects f in 0 or 1 vertex (there is no contribution to new edges from triangles containing e). There are at most $600c_2 n$ signed elements of Φ that

intersect e in 1 vertex. For each of these, a random configuration covers e with probability at most $3n^2/(1-60c)d(G^*)^9n^3$, so the total contribution to E_e from such elements is at most $2000c_2d(G^*)^{-9}$. Also, Φ has at most $100c_2n^2$ signed elements, and for each one that is disjoint from e the contribution to E_e is at most $6n/(1-60c)d(G^*)^9n^3$, so the total contribution from such elements is at most $1000c_2d(G^*)^{-9}$. We obtain $E_e < 3000c_2d(G^*)^{-9}$, which implies the claimed bound on Γ.

Next we claim that whp Γ' is c_2''-bounded, where $c_2'' = 20c_2'd(G^*)^{-7}$. The argument is very similar to that given for Γ. Now a configuration for ff' consists of the new edges of some $\Omega_{ff'}$. By Lemma 2.10, at each step, the number of choices of $\Omega_{ff'}$ with all new edges belonging to G^* (with no excluded configurations) is $(1\pm 50c)d(G^*)^7n^2$. Assuming that the graph of previous new edges is c_2''-bounded, as $\partial_2\Phi^+ \cup \Gamma$ is $2c_2'$-bounded, the number of excluded configurations is at most $10c_2''n^2$, which is less than half of the total. Next, for each $e \in G^*$, we consider separately the contributions to E_e according to whether e intersects $f \cup f'$ in 0 or 1 vertex (there is no contribution to new edges if $e \subseteq f \cup f'$).

First we consider those $ff' \in \mathcal{S}$ that intersect e in 1 vertex x. There are two choices for $x \in e$. If $x \in f \cap f'$ then there are at most $200c_2n$ choices for $f \cap f' \in \partial_2\Phi^+ \cup \partial_2\Phi^-$, which determines f and f'. If $\{x\} = f \setminus f'$ then there are at most $|\Gamma(x)| < c_2'n$ choices for f, and so f'. The same bound applies if $\{x\} = f' \setminus f$, so there are at most $5c_2'n$ such ff'. Each contributes at most $2n/(1-50c)d(G^*)^7n^2$ to E_e, so the total contribution from such ff' is at most $11c_2'd(G^*)^{-7}$. Also, $|\mathcal{S}| = |\partial_2\Phi^-| < 100c_2n^2$, and for each $ff' \in \mathcal{S}$ with $e \cap (f \cup f') = \emptyset$ the contribution to E_e is at most $2/(1-50c)d(G^*)^7n^2$, so the total contribution from such elements is at most $300c_2d(G^*)^{-7}$. We obtain $E_e < 12c_2'd(G^*)^{-7}$, which implies the claimed bound on Γ'. Recalling that $d(G^*) > 0.24d(G)^3$ and $c_3 = 10^{20}c_2d(G)^{-50}$ we see that $\bigcup M^o = \partial_2\Psi^+ = S \cup \Gamma \cup \Gamma'$ is c_3-bounded, so we have the required properties for **Hole**. □

4. Completion

For the **Completion** step, we divide the analysis into two parts. Firstly, we will determine what conditions on M_1 and M_2 enable us to find M_3 and M_4. Secondly, we will show that the sets M^c, M^o and M^i from **Cover** and **Hole** can be modified to give M_1 and M_2 satisfying the required conditions. For convenient notation we suppress the embedding $\pi : V(G) \to \mathbb{F}_{2^a}$ whenever we do not need to refer to it, instead thinking of $V(G)$ as a subset of $\mathbb{F}_{2^a}$.

4.1. Shuffles. Suppose we have a set M_2 of edge-disjoint triangles in G^*, and we want to find sets M_3 and M_4 of edge-disjoint triangles in G^* such that $\bigcup M_3 = \bigcup M_4$, $M_3 \subseteq T$ and $M_2 \subseteq M_4$. Our basic building blocks ('shuffles') will be edge-disjoint subgraphs of G^*, each having two different triangle decompositions, one only using triangles in T, and the other including any specified triangle of M_2. Then the unions over all blocks of the two triangle decompositions will give M_3 and M_4 as required.

We define the shuffles as follows. Fix $x = (x_1, x_2, x_3) \in \mathbb{F}_{2^a}^3$ and $t = (t_1, t_2) \in \mathbb{F}_{2^a}^2$ such that $\{x_1, x_2, x_3, t_1, t_2\}$ is linearly independent over $\mathbb{F}_2$. Let $\langle x \rangle$ be the subspace of $\mathbb{F}_{2^a}$ generated by $\{x_1, x_2, x_3\}$. The xt-shuffle S_{xt} is the complete tripartite graph with parts $t_i + \langle x \rangle = \{t_i + y : y \in \langle x \rangle\}$, $i \in [3]$, where $t_3 := t_1 + t_2$. If $S_{xt} \subseteq G^*$ then it has a triangle decomposition M_{3xt} only using triangles in T: take all triangles $y_1y_2y_3$ where each $y_i \in t_i + \langle x \rangle$ and $y_1 + y_2 + y_3 = 0$. We define another triangle decomposition M_{4xt} of S_{xt} by translating each triangle of M_{3xt} by (x_1, x_2, x_3), i.e. M_{4xt} consists of all triangles $y_1y_2y_3$ where each $y_i \in t_i + \langle x \rangle$ and $x_1 + x_2 + x_3 + y_1 + y_2 + y_3 = 0$.

To construct M_3 and M_4, we choose shuffles according to a random greedy algorithm, where in each step we consider some $z_1z_2z_3 \in M_2$, and choose some shuffle $S_{xt} \subseteq G^*$ such that $z_i = t_i + x_i$ for all $i \in [3]$. We will see in Lemma 4.1 that the Randomised Algebraic Construction is whp such that there are many choices for such a shuffle. This is the most important property of the construction, and it would not hold if we had chosen the template to be a uniformly random set of edge-disjoint triangles; in fact the expected number of shuffles (or any 'shuffle-like' configuration) would be $o(1)$. First we identify a property that we need for triangles in M_2 so that the required shuffles exist and can be chosen to be edge-disjoint. We say that $z_1z_2z_3$ is *octahedral* if $z_1 + z_2 + z_3 \neq 0$ and there is a copy K' of $K_{2,2,2}$ in G such that $\pi(K')$ has parts $\{z_1, z_2 + z_3\}$, $\{z_2, z_1 + z_3\}$ and $\{z_3, z_1 + z_2\}$; we call K' the *associated octahedron* of $z_1z_2z_3$. We assume

(P1) all triangles in M_2 are octahedral, with edge-disjoint associated octahedra.

The associated octahedron has all the properties that we require for the construction of M_3 and M_4, so we could implement our algorithm without using shuffles. (This remark was communicated to the author by Yang, and independently by Glebov and Luria.) We have opted to keep the shuffle argument in this paper, as it indicates how to treat general (hyper)graphs (we only see how to dispense with it for triangles).

Lemma 4.1. *Under the random choice of π used in the definition of T, whp for any octahedral $z_1z_2z_3$ there are $(1 \pm 200c)d(G)^{180}\gamma^{18}2^{2a}$ shuffles $S_{xt} \subseteq G^*$ such that $t_i + x_i = z_i$ for $i \in [3]$.*

Proof. We can write the number of such shuffles as a sum of indicator variables $X = \sum 1_{E(K,\ell,x,t)}$, where the sum ranges over all (K, ℓ, x, t) such that K is a copy of $K_{8,8,8}$ in G containing the associated octahedron K' of $z_1z_2z_3$, ℓ is a bijective labelling of each part of K by $\mathbb{F}_2^3$, we let $E(K, \ell, x, t)$ be the event that $\pi(w) = t_i + \ell(w) \cdot x$ for all $i \in [3]$ and w in the ith part of K, and we assume ℓ is consistent with K', in that $\ell(\pi^{-1}(z_i)) = e_i$ and $\ell(\pi^{-1}(z_i + z_j)) = e_i + e_j$ for $\{i, j\} \subseteq [3]$.

As G is $(c, 16)$-typical, there are $(1 \pm 181c)d(G)^{180}n^{18}$ choices of (K, ℓ). There are $2^{2a} - O(n)$ choices of t, which determines x given z, as only $O(n)$ choices of t are excluded by the condition that $\{x_1, x_2, x_3, t_1, t_2\}$ is linearly independent over $\mathbb{F}_2$: there are $O(1)$ possible linear relations between them, and each such relation is linearly independent or contradictory to the system $t_i + x_i = z_i$ for $i \in [3]$ (as

$z_1 + z_2 + z_3 \neq 0$), so is satisfied by at most 2^a choices of t. Given (K, ℓ, x, t), conditional on $\pi|_{K'}$, we have $\mathbb{P}(E(K, \ell, x, t)) = (1 + O(1/n))2^{-18a}$. Therefore $\mathbb{E}X = (1 \pm 182c)d(G)^{180}\gamma^{18}2^{2a}$.

Also, any transposition τ of π affects X by at most $100 \cdot 2^a$. To see this, we estimate the number of shuffles containing $z_1z_2z_3$ and any fixed $v \in \mathbb{F}_{2^a} \setminus \{z_1, z_2, z_3, z_1+z_2, z_1+z_3, z_2+z_3\}$. Consider any $j \in [3]$, $b \in \mathbb{F}_2^3 \setminus \{e_j, (1,1,1) - e_j\}$, and the equations $t_i + b \cdot x = v$ and $t_i + x_i = z_i$ for $i \in [3]$ in (t, x). We have four linearly independent constraints, so there are at most 2^a solutions. Including multiplicative factors for i, b and τ gives the required bound. Now by Lemma 2.6 whp $X = (1 \pm 200c)d(G)^{180}\gamma^{18}2^{2a}$. □

4.2. Linear extensions. We digress to note a more general estimate for future reference. Suppose H is a graph, $y = (y_i : i \in [g])$ are variables, and for all $v \in V(H)$ we have distinct linear forms $L_v(y) = c_v + \sum_{i \in S_v} y_i$ for some $c_v \in \mathbb{F}_{2^a}$ and $S_v \subseteq [g]$. We call $E = (L, H)$ a *linear extension* with *base* $F = \{v \in V(H) : S_v = \emptyset\}$. Let $X_E(G^*)$ be the number of *L-embeddings* of H, i.e. embeddings ϕ of H in G^* such that for some $y \in \mathbb{F}_{2^a}^g$ we have $\phi(v) = L_v(y)$ for all $v \in V(H)$. The above argument (see also [6, Lemma 5.15]) gives the following formula analogous to that obtained for shuffles.

Lemma 4.2. *Let $E = (L, H)$ be a 16-degenerate linear extension with $|H| \leq 500$. Suppose*

- *H has a triangle decomposition M such that for each $xyz \in M$ we have $L_x + L_y = L_z$,*
- *The incidence matrix of $\{S_v : v \in V(H)\}$ has full column rank $g \geq 1$.*

Then

$$X_E(G^*) = (1 \pm 1.1|H|c)d(G)^{|H \setminus H[F]|}\gamma^{|V(H)\setminus F|}2^{ga}.$$

4.3. Shuffle algorithm. Recalling our general framework for random greedy algorithms, we want to show that, of the potential shuffles S_{xt} with $t_i + x_i = z_i$ for $i \in [3]$, at most half are excluded due to sharing an edge with a previous shuffle, assuming some boundedness condition on the graph Γ of new edges from previous shuffles. We classify the potential restrictions according to the label of the shuffle edge involved, which is specified by some $\{j, k\} \subseteq [3]$ and $b_j, b_k \in \mathbb{F}_2^3$ such that $b_j \notin \{(e_j, (1,1,1) - e_j)$ or $b_k \notin \{(e_k, (1,1,1) - e_k)$ (here we do not consider edges of the associated octahedra: these are already determined, and edge-disjoint by (P1).) For any $v_jv_k \in G^*$, the shuffles excluded because of mapping the given labelled shuffle edge to v_jv_k are given by the (x, t)-solutions of the system $\mathcal{S}$ of equations $t_j + b_j \cdot x = v_j$, $t_k + b_k \cdot x = v_k$ and $t_i + x_i = z_i$ for $i \in [3]$. There may be 0, 1 or 2^a solutions. We can ignore the case of 0 solutions, as it does not exclude anything. For the cases with 1 solution, we can bound the number of excluded choices by the number of edges covered by all shuffles, which is $192|M_2|$.

It remains to consider the case that $\mathcal{S}$ has 2^a solutions, which occurs when one of the equations is redundant, due to being a linear combination of the other

equations. There are a constant number of linear combinations, and each constrains (v_j, v_k) to lie on a line, as may be seen from general considerations of linear algebra, or simply by enumerating the possibilities: wlog $t_k + b_k \cdot x = v_k$ is redundant, due to

(i) $b_k = e_k$ and $v_k = z_k$,

(ii) $b_k = (1,1,1) - e_k$ and $v_k = z_1 + z_2 + z_3 - z_k$,

(iii) $b_j + b_k = e_j + e_k$ and $v_j + v_k = z_j + z_k$,

(iv) $b_j + b_k = e_i$ and $v_j + v_k = z_i$, where $[3] = \{i, j, k\}$.

In cases (i) and (ii) where v_k is fixed, assuming that Γ is c_5-bounded, there are at most $c_5 n$ choices for v_j such that $v_j v_k \in \Gamma$. In cases (iii) and (iv) we need an additional boundedness condition:

We say that Γ is *linearly c_5-bounded* if Γ is c_5-bounded and also contains at most $c_5 2^a$ edges from any line of the form $\{(x_1 + \mu, x_2 + \mu) : \mu \in \mathbb{F}_{2^a}\}$.

We also need similar conditions so that we can avoid the associated octahedra; writing Δ for the union of all associated octahedra of triangles in M_2, we will ensure that

(P2) Δ is linearly c_4-bounded.

Then the total number of excluded shuffles is at most $192(|M_2| + (c_4 + c_5)2^{2a}) < 200c_5 2^{2a}$, which is less than half of the total.

Next we fix $e \in G^*$ and estimate E_e. To do so, we fix b_j, b_k as above, write $e = v_j v_k$ and estimate the sum over $z_1 z_2 z_3 \in M_2$ of the probability p that a random shuffle S_{xt} with $t_i + x_i = z_i$ for $i \in [3]$ satisfies $t_j + b_j \cdot x = v_j$ and $t_k + b_k \cdot x = v_k$. For fixed $z_1 z_2 z_3$, if the system $\mathcal{S}$ as above has N solutions then $p = N/(1 \pm 200c)d(G)^{180}\gamma^{18}2^{2a}$. When $N = 1$ the total contribution is at most $|M_2|/(1 - 200c)d(G)^{180}\gamma^{18}2^{2a} < 1.1c_4 d(G)^{-180}\gamma^{-18}$. If $N = 2^a$ then (z_1, z_2, z_3) is constrained to lie in a certain plane (this can be seen by linear algebra, or by considering each possibiity as above: e.g. in case (iii) the plane is $v_j + v_k = z_j + z_k$). Thus we see the final property that we need from M_2:

(P3) M_2 contains at most $c_4 2^a$ elements $z_1 z_2 z_3$ from any *basic plane* of the form $b \cdot z = v$ where $b \in \mathbb{F}_2^3 \setminus \{0\}$.

(Note that by (P1) we can assume $v \neq 0$ in (P3).) Then the total contribution is at most $c_4 2^a \cdot 2^a/(1 - 200c)d(G)^{180}\gamma^{18}2^{2a}$. Summing over $\{j, k\}$, b_j and b_k, we can estimate $E_e < 250c_4 d(G)^{-180}\gamma^{-18} = c_5/4$. Applying Lemma 2.8 as in the proof of Lemma 2.9, we deduce that whp the boundedness assumptions on Γ used above do not fail (linear boundedness follows in the same way as boundedness), and so the algorithm does not abort. This completes the analysis of the first part of **Completion**: given M_1 and M_2 as in **Completion**, under the conditions (P1–P3) on M_2, we can find M_3 and M_4 as in **Completion**.

4.4. Octahedral Elimination Algorithm. To complete the proof of **Completion**, and so of the theorems, it remains to show that we can find M_1 and M_2 satisfying the conditions (P1–P3). We apply a similar two-phase algorithm to that used in **Hole**.

Phase I. We start with $\Phi = M^c + M^i - M^o$, so $\partial_2\Phi = L$, $\partial_2\Phi^+ = \bigcup(M^c \cup M^i)$, $\partial_2\Phi^- = \bigcup M^o$. Next we eliminate all triangles in Φ according to a random greedy algorithm, where in each step we consider some original signed element f of Φ, and choose an octahedral configuration Ω_f to replace f. We say that a triangle f' of Ω_f is *far* if $|f' \cap f| \leq 1$, and that Ω_f is *valid* if (i) none of its triangles are template triangles, with the possible exception of f, and (ii) all of its far triangles are octahedral, and their associated octahedra share edges only in Ω_f, in which case we denote their union by the extended configuration Ω_f^+. We say that an edge of Ω_f^+ not in f is *new*, and choose a valid Ω_f uniformly at random subject to the new edges being distinct from all new edges from previous steps.

Let Φ' denote the result of Phase I (if it does not abort). We have $\partial_2\Phi' = \partial_2\Phi = L$, and writing Γ for the graph of new edges, every signed element of Φ' is either a far triangle consisting of three edges of Γ, or is not far and consists of two edges of Γ and one edge of $\partial_2\Phi^+$.

Phase II. Now we will eliminate all triangles of Φ' apart from those that contain an edge of L or were far in the previous modification procedure. We partition all such triangles into a sequence $\mathcal{S}$ of pairs of signed elements of Φ', so that for each $ff' \in \mathcal{S}$, there is some $e \in \partial_2\Phi^+$ such that f and f' both contain e, and f and f' have opposite signs. We eliminate each $ff' \in \mathcal{S}$, according to a random greedy algorithm, by subtracting some copy $\Omega_{ff'}$ of Ω with $f_{000} = f$ and $f_{001} = f'$, or vice versa, depending on the signs. Now we say that $\Omega_{ff'}$ is *valid* if all of its triangles apart from f and f' are octahedral, and their associated octahedra share edges only in $\Omega_{ff'}$, in which case we denote their union by the extended configuration $\Omega_{ff'}^+$. We refer to edges of $\Omega_{ff'}^+$ not in f or f' as *new* edges, and choose a valid $\Omega_{ff'}$ uniformly at random subject to the new edges being distinct from Γ and all new edges from previous steps.

Let Ψ denote the result of this algorithm (if it does not abort) and Γ' the graph of new edges for Phase II. Since $\partial_2\Psi = \partial_2\Phi = L$, defining $M_1 = \Psi^+$ and $M_2 = \Psi^-$, we see that $\bigcup M_2 = \Gamma \cup \Gamma'$ and $\bigcup M_1 = L \cup \Gamma \cup \Gamma'$, so $(L, \bigcup M_2)$ is a partition of $\bigcup M_1$. The following lemma completes the proof of **Completion**, and so of the theorems, under the assumption that G is $(c, 16)$-typical.

Lemma 4.3. *whp M_2 satisfies (P1), (P2) and (P3).*

Sketch Proof. To analyse Phase I, we first estimate the number of choices for an extended configuration on a triangle f. This can be described by the linear extension (Ω_f^+, L), where Ω_f^+ is as above, we have variables $z = (z_1, z_2, z_3)$, which we also use to label the vertices of $\Omega_f \setminus f$, we define $L_x = x$ for all $x \in \Omega_f$, and define L_x for all other x as required for the far triangles in Ω_f to be octahedral, i.e. in the associated octahedron for a triangle abc, the linear forms on the two vertices in each

of the three parts are $\{L_a, L_b + L_c\}$, $\{L_b, L_c + L_a\}$ and $\{L_c, L_a + L_b\}$. By Lemma 4.2 whp G^* is such that for any triangle f in Φ, there are $(1 \pm 60c)d(G)^{45}\gamma^{15}2^{3a}$ valid choices of Ω_f. Here we also use the fact that for any triangle abc of Ω_f other than f there are only 2^{2a} solutions to $L_a(z)+L_b(z)+L_c(z) = 0$. The precise exponents of $d(G)$ and γ (which are not important for the argument) may be easily calculated from the observation that adding an octahedron to a triangle adds 3 new vertices and 9 new edges, and Ω_f^+ is the composition of 5 such extensions.

Next we claim that whp the graph Γ of new edges is linearly c_3'-bounded, where $c_3' = 400c_3d(G)^{-45}\gamma^{-15}$. We assume this bound on the current graph of new edges and estimate how many configurations are excluded. Consider any edge uu' of the extended configuration. Suppose first that $uu' \cap f = \emptyset$. If $L_u(y) + L_{u'}(y)$ is not constant, then for any $vv' \in G^*$ the number of L-embeddings with $L_u(y) = v$ and $L_{u'}(y) = v'$ is at most 2^a. There are at most $45(|M^c| + |M^i| + |M^o|) < 100c_3n^2$ choices for a previous new edge vv', so this excludes at most $100c_3n^2 2^a$ configurations. On the other hand, if $L_u(y) + L_{u'}(y)$ is constant, then $L_u(y)$ and $L_{u'}(y)$ are constrained to lie on a basic line; there are at most $c_3'2^a$ choices for vv' by linear boundedness, and each such vv' excludes at most 2^{2a} configurations. The latter estimate also applies to the case when one of u or u' is in f. Summing these bounds over all uu', we see that fewer than half of the total configurations are excluded.

Next we fix any edge uu' of the extended configuration, any $vv' \in G^*$, and estimate the sum over $f \in \Phi$ of the probability p that a random configuration satisfies $L_u(y) = v$ and $L_{u'}(y) = v'$. If $uu' \cap f = \emptyset$ and $L_u(y) + L_{u'}(y)$ is not constant, then $p < 2^a/(1 - 60c)d(G)^{45}\gamma^{15}2^{3a}$ for any $f \in \Phi$. There are at most c_3n^2 choices for f, so the total contribution is at most $2c_3d(G)^{-45}\gamma^{-15}$. Otherwise, if $L_u(y)+L_{u'}(y)$ is constant or one of u or u' is in f, then one vertex of f is specified by $(L_u(y), L_{u'}(y))$. For example, writing $f = abc$, in the associated octahedron for az_2z_3, if $u = z_2$ and $u' = a + z_2$ then a is specified by $(L_u(y), L_{u'}(y))$. Then there are at most $2c_3n$ choices for f (as $\bigcup M^o$ is c_3-bounded). For each such f we have a contribution of at most $2^{2a}/(1 - 60c)d(G)^{45}\gamma^{15}2^{3a}$, so again the total contribution is at most $2c_3d(G)^{-45}\gamma^{-15}$. Summing these bounds over all uu' we can estimate $E_{vv'} < 100c_3d(G)^{-45}\gamma^{-15} = c_3'/4$. Applying Lemma 2.8 as in the proof of Lemma 2.9, we deduce the claimed bound on Γ.

We also claim that whp there are at most $2c_3'2^a$ far triangles in any basic plane $\Pi = \{z : b\cdot z = v\}$. To see this, we first consider the contribution from the template triangles $\Pi^* = \Pi \cap T$. Since $z_1+z_2+z_3 = 0$ is linearly independent or contradictory to the defining equation of Π we have $|\Pi^*| \leq 2^a$. Summing $E_{vv'} < c_3'/4$ over an edge vv' in each triangle of Π^*, by Lemma 2.8 whp Π contains at most $c_3'2^a$ template triangles. Now fix any far non-template triangle f' of the extended configuration, any $g \in K_3(G^*)$, and estimate the sum over $f \in \Phi$ of the probability p that a random configuration satisfies $L_{f'}(y) = g$. If $f' \cap f = \emptyset$ then as f' is non-template it determines the configuration, so $p < 1/(1 - 60c)d(G)^{45}\gamma^{15}2^{3a}$, giving a total contribution of at most $2c_3d(G)^{-45}\gamma^{-15}n^{-1}$. Otherwise, f' determines one of the associated octahedra, so specifies one vertex of f, for example, writing $f = abc$, if $f' = \{z_2, a + z_2, a + z_3\}$ then a is specified. Then there are at most $2c_3n$

choices for f; for each such f we have $p < 2^a/(1-60c)d(G)^{45}\gamma^{15}2^{3a}$, so again the total contribution is at most $2c_3 d(G)^{-45}\gamma^{-15}n^{-1}$. Summing over f' and applying Lemma 2.8 as in the proof of Lemma 2.9, we deduce the claimed bound on Π. This completes the analysis of Phase I.

We omit the similar analysis of Phase II (see [7] for more details). Finally, M_2 satisfies the conditions (P1–P3): indeed, (P1) holds by definition of the extended configurations and random greedy algorithms, (P2) holds as $\Delta \subseteq \Gamma \cup \Gamma'$, and (P3) holds as whp Ψ has at most $c_4 n$ triangles in any basic plane: this holds for the new triangles in this algorithm by the same argument as for Φ', and we may include the far triangles from the previous algorithm in this estimate. □

5. Concluding remarks

Although we have proved Wilson's conjecture, one may still ask for more precise estimates (even an asymptotic formula) for the number of Steiner Triple Systems, and more generally designs. Such results have been obtained by Kuperberg, Lovett and Peled [9], using very different methods to ours, but only for designs within a certain range of parameters. One open case of particular interest (drawn to my attention by Ron Peled) is the problem of estimating the number $G(n,d)$ of d-regular graphs on n vertices. These may be viewed as designs with parameters $(n,2,1,d)$, for which our methods give $G(n,d) = d!^{-n}(dn/e + o(dn))^{dn/2}$. Much more precise results have been obtained by McKay and Wormald, including asymptotic enumeration for $d = \omega(n/\log n)$ (see [11]) and $d = o(\sqrt{n})$ (see [12]); their conjecture in [11] regarding a general asymptotic formula remains open.

Acknowledgement. We thank an anonymous referee for an extremely detailed report with many helpful suggestions on exposition and alternative proofs.

References

[1] P. Bennett and T. Bohman, A natural barrier in random greedy hypergraph matching, arXiv:1210.3581.

[2] T. Bohman, A. Frieze, E. Lubetzky, Random triangle removal, arXiv:1203.4223.

[3] C. J. Colbourn and J. H. Dinitz, *Handbook of Combinatorial Designs*, 2nd ed. Chapman & Hall / CRC, Boca Raton, 2006.

[4] D. A. Freedman, On tail probabilities for martingales, *Ann. Probability* 3:100–118 (1975).

[5] J. E. Graver and W. B. Jurkat, The module structure of integral designs, *J. Combinatorial Theory Ser. A* 15:75–90 (1973).

[6] P. Keevash, The existence of designs, arXiv:1401.3665.

[7] P. Keevash, Counting designs, arXiv:1504.02909.

[8] P. Keevash, A hypergraph regularity method for generalised Turán problems, *Random Struct. Alg.* 34:123–164 (2009).

[9] G. Kuperberg, S. Lovett and R. Peled, Probabilistic existence of regular combinatorial objects, Proc. 44th ACM STOC (2012).

[10] N. Linial and Z. Luria, Upper bounds on the number of Steiner triple systems and 1-factorizations, *Random Struct. Alg.* 43:399–406 (2013).

[11] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, *Europ. J. Combin.* 11:565–580 (1990).

[12] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, *Combinatorica* 4:369–382 (1991).

[13] C. McDiarmid, Concentration, in: Probabilistic Methods for Algorithmic Discrete Mathematics, *Alg. Combin.* 16:195–248 (1998).

[14] V. Rödl, On a packing and covering problem, *European J. Combinatorics* 6:69–78 (1985).

[15] V. Rödl and A. Ruciński, Dirac-type questions for hypergraphs – a survey (or more problems for Endre to solve), *An Irregular Mind (Szemerédi is 70)* 21:1–30 (2010).

[16] R. Wilson, The early history of block designs, *Rend. del Sem. Mat. di Messina* 9:267–276 (2003).

[17] R. M. Wilson, An existence theory for pairwise balanced designs I. Composition theorems and morphisms, *J. Combinatorial Theory Ser. A* 13:220–245 (1972).

[18] R. M. Wilson, An existence theory for pairwise balanced designs II. The structure of PBD-closed sets and the existence conjectures, *J. Combinatorial Theory Ser. A* 13:246–273 (1972).

[19] R. M. Wilson, An existence theory for pairwise balanced designs III. Proof of the existence conjectures, *J. Combinatorial Theory Ser. A* 18:71–79 (1975).

[20] R. M. Wilson, Nonisomorphic Steiner Triple Systems, *Math. Zeit.* 135:303–313 (1974).

[21] R. M. Wilson, The necessary conditions for t-designs are sufficient for something, *Utilitas Math.* 4:207–215 (1973).

[22] R. M. Wilson, Signed hypergraph designs and diagonal forms for some incidence matrices, *Designs, Codes and Cryptography*, 17:289–297 (1999).

Mathematical Institute, University of Oxford, Oxford, UK.
E-mail: keevash@maths.ox.ac.uk