

**Original citation:**

Pikhurko, Oleg and Razborov, Alexander. (2016) Asymptotic structure of graphs with the minimum number of triangles. *Combinatorics, Probability and Computing* .

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/79463>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>

**A note on versions:**

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# Combinatorics, Probability and Computing

<http://journals.cambridge.org/CPC>

Additional services for ***Combinatorics, Probability and Computing***:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)

---

## Asymptotic Structure of Graphs with the Minimum Number of Triangles

OLEG PIKHURKO and ALEXANDER RAZBOROV

Combinatorics, Probability and Computing / *FirstView* Article / June 2016, pp 1 - 23

DOI: 10.1017/S0963548316000110, Published online: 04 May 2016

**Link to this article:** [http://journals.cambridge.org/abstract\\_S0963548316000110](http://journals.cambridge.org/abstract_S0963548316000110)

### How to cite this article:

OLEG PIKHURKO and ALEXANDER RAZBOROV Asymptotic Structure of Graphs with the Minimum Number of Triangles. *Combinatorics, Probability and Computing*, Available on CJO 2016  
doi:10.1017/S0963548316000110

**Request Permissions :** [Click here](#)

---

---

# Asymptotic Structure of Graphs with the Minimum Number of Triangles

---

OLEG PIKHURKO<sup>1†</sup> and ALEXANDER RAZBOROV<sup>2‡</sup>

<sup>1</sup>Mathematics Institute and DIMAP, University of Warwick, Coventry CV4 7AL, UK  
(e-mail: [O.Pikhurko@warwick.ac.uk](mailto:O.Pikhurko@warwick.ac.uk))

<sup>2</sup>Department of Computer Science, University of Chicago, Chicago, IL 60637, USA  
(e-mail: [razborov@cs.uchicago.edu](mailto:razborov@cs.uchicago.edu))

*Received 24 September 2014; revised 16 November 2015*

We consider the problem of minimizing the number of triangles in a graph of given order and size, and describe the asymptotic structure of extremal graphs. This is achieved by characterizing the set of flag algebra homomorphisms that minimize the triangle density.

2010 *Mathematics subject classification*: Primary 05C35

## 1. Introduction

The famous theorem of Turán [39] determines  $\text{ex}(n, K_r)$ , the maximum number of edges in a graph with  $n$  vertices that does not contain the  $r$ -clique  $K_r$  (the case  $r = 3$  was previously solved by Mantel [25]). The unique extremal graph is the *Turán graph*  $T_{r-1}(n)$ , the complete  $(r - 1)$ -partite graph of order  $n$  whose part sizes differ at most by 1. Thus, for fixed  $r$ , we have

$$\text{ex}(n, K_r) = \left(1 - \frac{1}{r-1}\right) \binom{n}{2} + O(1).$$

Rademacher (unpublished, 1941) proved that a graph with  $\text{ex}(n, K_3) + 1$  edges has at least  $\lfloor n/2 \rfloor$  triangles. This prompted Erdős [10] to pose the more general problem: What is  $g_r(m, n)$ , the smallest number of  $K_r$ -subgraphs in a graph with  $n$  vertices and  $m$  edges? Various results have been obtained by Erdős [11, 13], Moon and Moser [26], Nordhaus and Stewart [28], Bollobás [2], Fisher [15], Lovász and Simonovits [20, 21], Razborov [34, 35], Nikiforov [27], Reiher [36], and others.

<sup>†</sup> Supported by ERC grant 306493 and EPSRC grant EP/K012045/1.

<sup>‡</sup> Part of this work was done while the author was at Steklov Mathematical Institute, supported by the Russian Foundation for Basic Research, and at Toyota Technological Institute, Chicago.

Let us consider the asymptotic question, that is, what is the limit

$$g_r(a) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \frac{g_r(\lfloor a \binom{n}{2} \rfloor, n)}{\binom{n}{r}}$$

for any given  $a \in [0, 1]$  and  $r$ ? While it is not difficult to show that the limit exists, determining  $g_r(a)$  is a much harder task that was accomplished only relatively recently (for  $r = 3$  by Razborov [35], for  $r = 4$  by Nikiforov [27], and for  $r \geq 5$  by Reiher [36]).

The following construction gives the value of  $g_3(a)$  (as well as  $g_r(a)$  for every  $r \geq 4$ ). Given  $a \in (0, 1)$ , we choose integer  $t \geq 1$  and real  $c \in [1/(t + 1), 1/t]$  such that the complete  $(t + 1)$ -partite graph of order  $n \rightarrow \infty$  with  $t$  largest parts each of size  $(c + o(1))n$  has edge density  $a + o(1)$ . Formally, let the integer  $t \geq 1$  satisfy

$$a \in \left( 1 - \frac{1}{t}, 1 - \frac{1}{t + 1} \right] \tag{1.1}$$

and let

$$c = \frac{t + \sqrt{t(t - a(t + 1))}}{t(t + 1)} \tag{1.2}$$

be the (unique) root of the quadratic equation

$$2 \left( \binom{t}{2} c^2 + tc(1 - tc) \right) = a \tag{1.3}$$

with  $c \geq 1/(t + 1)$ . Since  $a > 1 - 1/t$ , it follows from (1.2) (or from (1.3)) that  $c < 1/t$ . Partition the vertex set  $[n] = \{1, \dots, n\}$  into  $t + 1$  non-empty parts  $V_1, \dots, V_{t+1}$  with  $|V_1| = \dots = |V_t| = \lfloor cn \rfloor$  for  $i \in [t]$ . Let  $G$  be obtained from the complete  $t$ -partite graph  $K(V_1, \dots, V_{t-1}, U)$ , where  $U = V_t \cup V_{t+1}$ , by adding an arbitrary triangle-free graph  $G[U]$  on  $U$  with  $|V_i||V_{t+1}|$  edges<sup>1</sup>. Clearly, the edge density of  $G$  is  $a + o(1)$ . Thus  $g_3(a) \leq h(a)$ , where

$$h(a) \stackrel{\text{def}}{=} 6 \left( \binom{t}{3} c^3 + \binom{t}{2} c^2(1 - tc) \right). \tag{1.4}$$

If  $a = 1$ , we let  $G$  be the complete graph  $K_n$  and define  $h(1) = 1$ . If  $a = 0$ , we take the empty graph and let  $h(0) = 0$ . For  $a \in [0, 1]$ , let  $\mathcal{H}_{a,n}$  be the set of all possible graphs  $G$  on  $[n]$  that arise in this way,  $\mathcal{H}_a \stackrel{\text{def}}{=} \cup_{n \in \mathbb{N}} \mathcal{H}_{a,n}$ , and  $\mathcal{H} \stackrel{\text{def}}{=} \cup_{a \in [0,1]} \mathcal{H}_a$ . In general,  $\mathcal{H}_{a,n}$  has many non-isomorphic graphs and this seems to be one of the reasons why this extremal problem is so difficult.

Although each of the papers [27, 35, 36] implies the lower bound  $g_3(a) \geq h(a)$ , it is not clear how to extract the structural information about extremal graphs from these proofs. Here we partially fill this gap by showing that, modulo changing a negligible proportion of adjacencies, the set  $\mathcal{H}$  consists of all almost extremal graphs for the  $g_3$ -problem. Here is the formal statement.

<sup>1</sup> One possible choice is to take  $G[U] = K(V_t, V_{t+1})$ , resulting in  $G = K(V_1, \dots, V_{t+1})$ . But since each edge of  $G[U]$  belongs to exactly  $|V_1| + \dots + |V_{t-1}|$  triangles, the choice of  $G[U]$ , due to its triangle-freeness, has no effect on the triangle density.

**Theorem 1.1.** For every  $\varepsilon > 0$  there are  $\delta > 0$  and  $n_0$  such that every graph  $G$  with  $n \geq n_0$  vertices and at most  $(g_3(a) + \delta) \binom{n}{3}$  triangles, where  $a = e(G) / \binom{n}{2}$ , can be made isomorphic to some graph in  $\mathcal{H}_{a,n}$  by changing at most  $\varepsilon \binom{n}{2}$  adjacencies.

We remark that although this statement resembles (and implies) the celebrated Triangle Removal Lemma, it does not say anything new in that direction since its proof relies on the lemma. What our Theorem 1.1 can and should be compared to, is the following old result due to Lovász and Simonovits.

**Theorem 1.2 ([21, Theorem 2]).** For any real  $\varepsilon > 0$  and integers  $t \geq r - 1 \geq 2$ , there are  $\delta > 0$  and  $n_0$  such that every graph  $G$  with  $n \geq n_0$  vertices,  $(1 - 1/t \pm \delta) \binom{n}{2}$  edges, and at most  $(g_r(1 - 1/t) + \delta) \binom{n}{r}$  copies of  $K_r$  can be made isomorphic to  $T_r(n)$  by changing at most  $\varepsilon \binom{n}{2}$  adjacencies. □

Note that  $T_r(n)$  is  $o(n^2)$ -close in the edit distance to every graph in  $\mathcal{H}_{1-1/t,n}$ , hence the difference between them is immaterial. Thus, comparing our Theorem 1.1 to Theorem 1.2, note that Theorem 1.1 covers all values of  $a$  (not only those that are close to *critical points*  $a = 1 - 1/t$  for an integer  $t \geq r - 1$ ) but it deals with the case  $r = 3$  only.

Theorem 1.1 is obtained by building upon the flag algebra approach from [35]. In order to prove it we have to characterize first the set of extremal flag algebra homomorphisms for the  $g_3$ -problem. This is done in Theorem 2.1 of Section 2, where the precise statement can be found. This task requires some extra work in addition to the arguments in [35] and is an example of how flag algebra calculations may lead to structural results about graphs. (For some other results of a similar type, see, e.g., [8, 9, 17, 29, 30].)

Theorem 1.1 (or more precisely Theorem 2.1) can be viewed as a small step towards the more general problem of understanding graph limits with given edge and triangle densities. The latter problem naturally appears in the study of exponential random graphs (see, e.g., [1, 6, 31, 32, 33]) and large deviation inequalities for the triangle density in Erdős–Rényi random graphs (see, e.g., [4, 5, 7, 23, 24]).

Let us now briefly review what is known (and conjectured) about exact results. As with any extremal problem, the two relevant and related questions here are the following (see [21, Problems 1, 2]).

**Question 1.** Determine  $g_r(m, n)$  as tightly as possible.

**Question 2.** Say as much as possible about the structure of extremal configurations.

Toward Question 1, it makes sense to compare  $g_r(m, m)$  with the function  $g_r(a)$ , now explicitly known due to [27, 35, 36]. A straightforward blow-up construction (see, e.g., [35, Theorem 4.1]) gives us

$$g_r(m, n) \geq \frac{n^r}{r!} g_r(2m/n^2).$$

In the reverse direction, an obvious calculation based on the graphs from  $\mathcal{H}_{a,n}$  gives the estimate

$$g_r(m, n) \leq \frac{n^r}{r!} g_r(2m/n^2) + O\left(\frac{n^{r+1}}{n^2 - 2m}\right).$$

Nikiforov [27, Theorem 1.3] improved this to

$$g_r(m, n) \leq \frac{n^r}{r!} g_r(2m/n^2) + \frac{n^r}{n^2 - 2m}.$$

Lovász and Simonovits made the following remarkable conjecture.

**Conjecture 1.3 ([20, Conjecture 1]).** *For every  $r \geq 3$  there is  $n_0$  such that for every  $n \geq n_0$  and  $m$  with  $0 \leq m \leq \binom{n}{2}$  at least one of  $g_r(m, n)$ -extremal graphs is obtained from a complete partite graph by adding a triangle-free graph inside one part.*

If Conjecture 1.3 is proved, then one may consider Question 1 combinatorially answered: the number of  $K_r$ -subgraphs in such a graph  $G$  is some explicit polynomial in  $m$ ,  $n$ , and part sizes, and the question reduces to its minimization over the integers. This task may be difficult but it involves no graph theory. In fact, it is not hard to show (see, e.g., [27, Section 3]) that the optimal part ratios are approximately those of the graphs in  $\mathcal{H}_a$ , where  $a = m/\binom{n}{2}$ . (However, our rounding  $|V_1| = \lfloor cn \rfloor$ , etc., was rather arbitrary: it was chosen just to have the family  $\mathcal{H}_a$  well-defined.)

Since the value of  $g_3(m, n)$  resulting from Conjecture 1.3 does not even have a nice analytical expression, it is conceivable that the only way of attacking Question 1 is via Question 2, using the so-called *stability approach*. This indeed turned out to be so in the only non-trivial intervals where the problem has been solved so far. Namely, assume that  $\text{ex}(n, K_{t+1}) \leq m \leq \text{ex}(n, K_{t+1}) + \epsilon(r, t)n^2$ , where  $\epsilon(r, t) > 0$  is a rather small constant; in other words, that  $a$  is in a small (upper) neighbourhood of a critical point  $1 - 1/t$ . Then for  $r \geq 4$  Lovász and Simonovits [21] proved Conjecture 1.3 in a much stronger universal form. Given recent developments, we would like to make the explicit conjecture that their result can be extended to arbitrary values of  $m$ .

**Conjecture 1.4.** *For every  $r \geq 4$  there exists  $n_0$  such that for every  $n \geq n_0$  and  $m$  with  $0 \leq m \leq \binom{n}{2}$  every  $g_r(m, n)$ -extremal graph is obtained from a complete partite graph by adding a triangle-free graph inside one part.*

For the case  $r = 3$  Lovász and Simonovits verified Conjecture 1.3 in the same neighbourhoods of critical points. Conjecture 1.4, however, is no longer true: for some pairs  $(m, n)$ , there are additional extremal graphs; see the families  $U_0$  and  $U_2$  in [21].

We hope that the techniques in our paper will turn out to be helpful in attacking Conjectures 1.3 and 1.4 for arbitrary  $m$ .

The paper is organized as follows. We outline the main ideas behind flag algebras and state some of the key inequalities from [35] in Section 2. There, we also state our result on the structure of  $g_3$ -extremal homomorphisms (Theorem 2.1) and show how this

implies Theorem 1.1. Section 3 contains a sketch of the proof from [35] that  $g_3(a) = h(a)$ . Theorem 2.1 is proved in Section 4.

### 2. Flag algebras

In order to understand this paper the reader should be familiar with the concepts introduced in [34]. We do not see any reasonable way of making this paper self-contained, without making it quite long and repeating large passages from [34]. Therefore, we restrict ourselves to sketching the proofs in [34, 35], during which we informally illustrate the main ideas by providing some analogues from the discrete world. This serves two purposes: to state the key inequalities from [34, 35] that we need here and to provide some guiding intuition for the reader who is about to start reading [34]. We stress that some flag algebra concepts do not have direct combinatorial analogues or require a plethora of constants to state them in terms of graphs. Here we just try to distil and present some motivational ideas. Besides, even if the theory was intentionally developed to cover arbitrary combinatorial structures, in our brief exposition we confine ourselves to the case of ordinary graphs, as the most intuitive one.

Many proofs in extremal graph theory proceed by considering possible densities of small subgraphs and deriving various inequalities between them. These calculations often become very cumbersome and difficult to keep track of ‘by hand’, especially since the number of non-isomorphic graphs increases very quickly with the number of vertices. One of the motivations behind introducing flag algebras was to develop a framework where the mechanical book-keeping part of the work is relegated to a computer.

So suppose that we have a graph  $G$ . Let  $n = |V(G)|$  be its order.

The *density* of a graph  $F$  in  $G$ , denoted by  $p(F, G)$ , is the probability that a random  $|V(F)|$ -subset of  $V(G)$  spans a subgraph isomorphic to  $F$ . The quantities that we are interested in are finite linear combinations  $\sum_{i=1}^s \alpha_i p(F_i, G)$ , where  $F_i$  is a graph and  $\alpha_i$  is a real constant. One can view a formal finite sum  $\sum_{i=1}^s \alpha_i F_i$  as a function that evaluates to  $\sum_{i=1}^s \alpha_i p(F_i, G)$  on input  $G$ . Since we would like to operate with these objects on computers, we try to keep redundancies to minimum. In particular, the graphs  $F_i$  are unlabelled and pairwise non-isomorphic. Let  $\mathcal{F}^0$  consist of all (unlabelled non-isomorphic) graphs and let  $\mathbb{R}\mathcal{F}^0$  be the vector space that has  $\mathcal{F}^0$  as a basis. (The meaning of the superscript 0 will be explained a bit later.)

There are some relations which are identically true when it comes to evaluations on input  $G$ : for example if  $n \geq \ell \geq |V(\tilde{F})|$  for some graph  $\tilde{F}$  and we know the densities of all subgraphs on  $\ell$  vertices, then the density of  $\tilde{F}$  can be easily determined:

$$p(\tilde{F}, G) = \sum_{F \in \mathcal{F}_\ell^0} p(\tilde{F}, F)p(F, G), \tag{2.1}$$

where  $\mathcal{F}_\ell^0 \subseteq \mathcal{F}^0$  consists of all graphs with exactly  $\ell$  vertices. So it makes sense to factor over  $\mathcal{K}^0$ , the subspace of  $\mathbb{R}\mathcal{F}^0$  generated by

$$\tilde{F} - \sum_{F \in \mathcal{F}_\ell^0} p(\tilde{F}, F)F,$$

over all choices of  $\tilde{F}$  and  $\ell \geq |V(\tilde{F})|$ . Let

$$\mathcal{A}^0 \stackrel{\text{def}}{=} \mathbb{R}\mathcal{F}^0 / \mathcal{K}^0.$$

By (2.1), any element of  $\mathcal{A}^0$  can still be identified with an evaluation on (sufficiently large) graphs.

Let some  $F_i \in \mathcal{F}_{\ell_i}^0$  for  $i = 1, 2$  be fixed. The product  $p(F_1, G)p(F_2, G)$  is the probability that two random subsets  $U_1, U_2 \subseteq V(G)$  of sizes  $\ell_1$  and  $\ell_2$ , drawn independently, induce copies of  $F_1$  and  $F_2$  respectively. With probability  $1 - O(1/n)$  (recall that  $n = |V(G)|$ ), the sets  $U_1$  and  $U_2$  are disjoint. Let us condition on this event. The conditional distribution can be generated as follows: first pick a random  $(\ell_1 + \ell_2)$ -set  $U$  and then take a random partition  $U = U_1 \cup U_2$  with  $|U_i| = \ell_i$ . Thus

$$p(F_1, G)p(F_2, G) = \sum_{F \in \mathcal{F}_{\ell_1 + \ell_2}^0} p(F_1, F_2; F)p(F, G) + O(1/n), \tag{2.2}$$

where  $p(F_1, F_2; F)$  denotes the probability that  $F[U_i] \cong F_i$  (i.e., the subgraph of  $F$  induced by  $U_i$  is isomorphic to  $F_i$ ) for both  $i = 1, 2$  when we take a random partition  $U_1 \cup U_2$  of the vertex set of  $F \in \mathcal{F}_{\ell_1 + \ell_2}^0$  with part sizes  $\ell_1$  and  $\ell_2$ . Since we are interested in the case when  $n \rightarrow \infty$ , we formally define the product  $F_1 \cdot F_2$  to be equal to

$$\sum_{F \in \mathcal{F}_{\ell_1 + \ell_2}^0} p(F_1, F_2; F) F \in \mathbb{R}\mathcal{F}^0$$

and extend this multiplication to  $\mathbb{R}\mathcal{F}^0$  by linearity. It is not surprising that this definition is compatible with the factorization by  $\mathcal{K}^0$ , making  $\mathcal{A}^0$  a commutative associate algebra with the empty graph being the multiplicative identity; see [34, Lemma 2.4].

Unfortunately, we do not have the property that graph evaluations preserve multiplication exactly. This can be rectified if we take as input not just a single graph  $G$  but a sequence of graphs  $\{G_n\}$  which is *convergent*, by which we mean that  $|V(G_1)| < |V(G_2)| < \dots$  (we call such sequences *increasing*) and for every graph  $F$  the limit

$$\phi(F) \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} p(F, G_n) \tag{2.3}$$

exists. Then the ‘value’ of  $\sum_{i=1}^s \alpha_i F_i \in \mathbb{R}\mathcal{F}^0$  on  $\{G_n\}$  is

$$\sum_{i=1}^s \alpha_i \phi(F_i).$$

One can take the dual point of view, considering  $\phi$  as a map from  $\mathbb{R}\mathcal{F}^0$  to  $\mathbb{R}$ ; it is routine to see that, for each convergent sequence  $\{G_n\}$ , the corresponding map  $\phi : \mathbb{R}\mathcal{F}^0 \rightarrow \mathbb{R}$  is compatible with the factorization by  $\mathcal{K}^0$  and, in fact, gives an algebra homomorphism from  $\mathcal{A}^0$  to  $\mathbb{R}$  (which we also denote by  $\phi$ ); see [34, Theorem 3.3]. We say that  $\phi$  is the *limit* of  $\{G_n\}$  and, following the notation in [34, Section 3.1], denote this as  $\phi = \lim_{n \rightarrow \infty} p^{G_n}$ , where

$$p^{G_n}(F) \stackrel{\text{def}}{=} p(F, G_n)$$

if  $|V(F)| \leq |V(G_n)|$ , and 0 otherwise.

Clearly,  $\phi$  is non-negative, that is,  $\phi(F) \geq 0$  for every graph  $F$ . Let  $\text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  be the set of all non-negative homomorphisms.

It turns out that every non-negative homomorphism  $\phi : \mathcal{A}^0 \rightarrow \mathbb{R}$  is the limit of some sequence of graphs. It is instructive to sketch a proof of this; see Lovász and Szegedy [22, Lemma 2.4] for details (or [34, Theorem 3.3] in a more general context). Take some integer  $n$ . Since the identity  $\sum_{F \in \mathcal{F}_n^0} F = 1$  holds in  $\mathcal{A}^0$ , we have that  $\sum_{F \in \mathcal{F}_n^0} \phi(F) = 1$ , that is,  $\phi$  defines some probability distribution on  $\mathcal{F}_n^0$ . Let  $G_{n,\phi} \in \mathcal{F}_n^0$  be drawn according to this distribution with the choices for different values of  $n$  being independent. Fix some  $F$  and  $\varepsilon > 0$ . Let  $n \geq |V(F)|$ . An easy calculation shows that the expectation of  $p(F, G_{n,\phi})$  is exactly  $\phi(F)$ . Also, the variance of  $p(F, G_{n,\phi})$ , which can be expressed via counting pairs of  $F$ -subgraphs versus two independent copies of  $F$ , is  $O(1/n)$ . Chebyshev's inequality implies that the probability of the 'bad' event  $|p(F, G_{n,\phi}) - \phi(F)| > \varepsilon$  is  $O(1/n)$  and the Borel–Cantelli Lemma shows that with probability 1 only finitely many bad events occur when  $n$  runs over, for example, all squares. Since there are only countably many choices of  $F$  and, for example,  $\varepsilon \in \{1, 1/2, 1/3, \dots\}$ , we conclude that  $\{G_{n^2,\phi}\}$  converges to  $\phi$  with probability 1. Thus the required convergent sequence exists.

If one wishes that the graph orders in the sequence span all natural numbers, one can pick some convergent sequence and fill all orders by uniformly 'blowing' up its members; see, e.g., [17, Section 2.3]. Alternatively, one can show that the sequence  $\{G_{n,\phi}\}$  itself converges with probability 1 via a stronger concentration result for  $p(F, G_{n,\phi})$  that considers its first four moments; see [19, Lemma 11.7].

How can these concepts be useful for proving that  $g_3(a) = h(a)$ ? Pick an increasing sequence of graphs  $\{G_n\}$  of edge density  $a + o(1)$  such that the limit of  $p(K_3, G_n)$  exists and is equal to  $g_3(a)$ . A standard diagonalization argument shows that  $\{G_n\}$  has a convergent subsequence; let  $\phi$  be its limit. Then  $\phi(K_2) = a$ . Now, if we can show that

$$\forall \phi \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R}) \quad (\phi(K_2) = a \implies \phi(K_3) \geq h(a)), \tag{2.4}$$

then we can conclude that indeed  $g_3(a) = h(a)$ , as was done in [35].

In this paper, we achieve more: we describe the set of all extremal homomorphisms, that is, those  $\phi \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  that achieve equality  $\phi(K_3) = g_3(\phi(K_2))$ .

Let  $\Phi \subseteq \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  consist of all possible limits of convergent sequences  $\{G_n\}$  for which there is  $a \in [0, 1]$  such that  $G_n \in \mathcal{H}_a$  for all  $n$ . Equivalently,  $\Phi$  can be defined as follows. Recall that the join  $G_1 \vee \dots \vee G_k$  of graphs  $G_1, \dots, G_k$  is obtained by taking their disjoint union and adding all edges in between. We define a similar operation on homomorphisms  $\phi_1, \dots, \phi_k \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$ . We need a more general construction where one specifies how much relative weight each  $\phi_i$  has, by giving non-negative reals  $\alpha_1, \dots, \alpha_k$  with sum 1. Let  $n \rightarrow \infty$  and, for  $i \in [k]$ , let  $G_{i,n}$  be a graph with  $\lfloor \alpha_i n \rfloor$  vertices such that the sequence  $\{G_{i,n}\}$  converges to  $\phi_i$ ; as we have already remarked, it exists. Let  $F_n = G_{1,n} \vee \dots \vee G_{k,n}$ . Let the join  $\phi = \vee(\phi_1, \dots, \phi_k; \alpha_1, \dots, \alpha_k)$  be the limit of  $\{F_n\}$  (it is easy to see that the limit exists).

Alternatively, we can define the join  $\phi$  without appealing to convergence. To this end, it is enough to define the density of each graph  $F \in \mathcal{F}^0$ , and we do it as follows. Let  $\text{aut}(F)$

denote the number of automorphisms of  $F$ . Let

$$\phi(F) \stackrel{\text{def}}{=} \frac{|V(F)|!}{\text{aut}(F)} \sum_{(V_1, \dots, V_k)} \prod_{i=1}^k \left( \alpha_i^{|V_i|} \phi_i(F[V_i]) \frac{\text{aut}(F_i)}{|V_i|!} \right), \tag{2.5}$$

where the summation runs over all possible ways (up to isomorphism) to partition  $V(F) = V_1 \cup \dots \cup V_k$  into  $k$  labelled parts (allowing empty parts) so that the induced bipartite subgraph  $F[V_i, V_j]$  is complete for all  $1 \leq i < j \leq k$ . The reader is welcome to formally check that the join is well-defined (with respect to the factorization by  $\mathcal{K}^0$ ) and belongs to  $\text{Hom}^+(\mathcal{A}^0, \mathbb{R})$ . (These facts are obvious from the first definition.)

Now,  $\Phi$  is exactly the set of all possible joins

$$\vee(\underbrace{0, \dots, 0}_{t-1 \text{ times}}; \underbrace{c, \dots, c}_{t-1 \text{ times}}, 1 - (t - 1)c),$$

where  $0$  denotes the (unique) non-negative homomorphism in  $\text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  of zero edge-density,  $\psi \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  is arbitrary with  $\psi(K_3) = 0$  and

$$\psi(K_2) = 2c(1 - tc)/(1 - (t - 1)c)^2,$$

and  $c$  is a real from the interval  $[1/(t + 1), 1/t]$ .

Our main result states that the set of  $g_3$ -extremal homomorphisms is exactly  $\Phi$ .

**Theorem 2.1.**

$$\Phi = \{ \phi \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R}) : \phi(K_3) = g_3(\phi(K_2)) \}.$$

Let us show that Theorem 2.1 implies Theorem 1.1. The shortest way is to refer to some known results about the so-called *cut-distance*  $\delta_{\square}$  that goes back to Frieze and Kannan [16]. We omit the definition of  $\delta_{\square}$  but refer the reader to [3, Definition 2.2] (see also [19, Chapter 8]).

Suppose for the sake of contradiction that Theorem 1.1 is false, which is witnessed by some  $\varepsilon > 0$ . Then we can find an increasing sequence  $\{G_n\}$  of graphs with

$$p(K_3, G_n) \leq g_3(p(K_2, G_n)) + o(1)$$

that violates the conclusion of Theorem 1.1. By passing to a subsequence, we can assume that  $\{G_n\}$  is convergent. Let  $\phi_0 \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  be its limit. Let  $a = \phi_0(K_2)$ . Clearly,  $\phi_0(K_3) = g_3(a)$ . By Theorem 2.1,  $\phi_0 \in \Phi$  and we can choose a sequence  $\{H_n\}$  in  $\mathcal{H}$  which converges to  $\phi_0$  with  $V(H_n) = V(G_n)$ .

This convergence means that asymptotically  $G_n$  and  $H_n$  have the same statistics of fixed subgraphs. This does not necessarily imply that  $G_n$  and  $H_n$  are close in the edit distance. (For example, two typical random graphs of edge density 1/2 have similar subgraph statistics but are far in the edit distance.) However, the presence of a spanning complete partite graph in  $H_n$  implies a similar conclusion about  $G_n$ , as follows.

Theorem 2.7 in Borgs, Chayes, Lovász, Sós and Vesztergombi [3] gives that  $\delta_{\square}(G_n, H_n) = o(1)$ , that is, the cut-distance between  $G_n$  and  $H_n$  tends to 0. (An important property of the

cut-distance is that an increasing sequence  $\{G_n\}$  is convergent if and only if it is Cauchy with respect to  $\delta_{\square}$ .)

By [3, Theorem 2.3], we can relabel  $V(H_n)$  so that for every disjoint  $S, T \subseteq V(G_n)$  we have

$$|e(G_n[S, T]) - e(H_n[S, T])| = o(v^2), \tag{2.6}$$

where  $v = v(n)$  is the number of vertices in  $G_n$ . Informally, this means that the graphs  $G_n$  and  $H_n$  have almost the same edge distribution with respect to cuts. Take the partition  $V(H_n) = V_1 \cup \dots \cup V_{t-1} \cup U$  that was used to define  $H_n$ . Let  $i \in [t - 1]$ . If we set  $S = V_i$  and  $T = V(G_n) \setminus V_i$  in (2.6), then we conclude that the number of  $S - T$  edges that are missing from  $G_n$  is  $o(v^2)$ . Also, the number of edges in  $G[V_i]$  is  $o(v^2)$ , for otherwise a random partition  $V_i = S \cup T$  would contradict (2.6). Thus, by changing  $o(v^2)$  adjacencies in  $G_n$ , we can assume that the graphs  $G_n$  and  $H_n$  coincide except for the subgraph induced by  $U$ . Suppose that  $|U| = \Omega(n)$  for otherwise we are done. We have

$$|e(G_n[U]) - e(H_n[U])| = |e(G_n) - e(H_n)| = o(v^2).$$

Of course, when we modify  $o(v^2)$  adjacencies in  $G_n$ , then the number of triangles changes by  $o(v^3)$ . Each edge of  $G_n[U]$  (and of  $H_n[U]$ ) is in the same number of triangles with the third vertex belonging to  $V(G_n) \setminus U$ . Since  $H_n[U]$  is triangle-free and  $G_n$  is asymptotically extremal, we conclude that  $G_n[U]$  spans  $o(v^3)$  triangles. By the Triangle Removal Lemma [14, 37] (see e.g., [18, Theorem 2.9]), we can make  $G_n[U]$  triangle-free by deleting  $o(v^2)$  edges.

If  $e(G_n[U]) \geq e(H_n[U])$ , then we just remove some edges from  $G_n[U]$  until exactly  $e(H_n[U])$  edges are left, in which case the obtained graph  $G_n$  belongs to  $\mathcal{H}_{a,n}$  and Theorem 1.1 is proved. Otherwise we obtain the same conclusion for all large  $n$  by applying the following lemma to  $G_n[U]$  and  $s = e(H_n[U])$ .

**Lemma 2.2.** *For every  $\varepsilon > 0$  there are  $\delta > 0$  and  $n_0$  such that for every  $K_3$ -free graph  $G$  on  $n \geq n_0$  vertices and every integer  $s$  with*

$$e(G) < s \leq \min(e(G) + \delta n^2, \lfloor n^2/4 \rfloor) \tag{2.7}$$

*one can change at most  $\varepsilon n^2$  adjacencies in  $G$  so that the new graph is still  $K_3$ -free and has exactly  $s$  edges.*

**Proof.** Clearly, it is enough to show how to ensure *at least*  $s$  edges in the final  $K_3$ -free graph. Given  $\varepsilon > 0$ , choose small positive constants  $c \gg \delta$ . Let  $n$  be large and let  $s$  satisfy (2.7). Let  $m = e(G)$ .

We can assume that, for example,  $m \geq \varepsilon n^2/3$ . Also, assume that  $m \leq \lfloor n^2/4 \rfloor - cn^2$ , for otherwise we are done by the Stability Theorem of Erdős [12] and Simonovits [38], which implies that  $G$  can be transformed into the Turán graph  $T_2(n)$  by changing at most  $\varepsilon n^2$  adjacencies.

The number  $p$  of paths of length 2 in  $G$  is

$$\sum_{x \in V(G)} \binom{d(x)}{2},$$

which is at least  $n \binom{2m/n}{2}$  by the convexity of the function  $\binom{x}{2}$ . By averaging, there is an edge  $xy \in E(G)$  that belongs to at least

$$\frac{2p}{m} \geq \frac{2n \binom{2m/n}{2}}{m} \geq \frac{4m}{n} - \delta n$$

such paths (which is just the number of edges between the set  $\{x, y\}$  and its complement).

Let  $G'$  be obtained from  $G$  by adding  $cn$  clones of  $x$  and  $cn$  clones of  $y$ . Thus  $G'$  has  $n' = (1 + 2c)n$  vertices and  $m' \geq m + cn(4m/n - \delta n) + (cn)^2$  edges. If we take a random  $n$ -subset  $U$  of  $V(G')$ , then each edge of  $G'$  is included with probability  $\binom{n}{2} / \binom{n'}{2}$ . Thus there is a choice of an  $n$ -set  $U$  such that the number of edges in  $H = G'[U]$  is at least the average, which in turn is at least

$$\frac{(m + cn(4m/n - \delta n) + (cn)^2) \binom{n}{2}}{\binom{(1+2c)n}{2}} \geq m + \frac{c^2(n^2 - 4m) - 2c\delta n^2}{(1 + 2c)^2}.$$

This is at least  $m + \delta n^2 \geq s$  by our assumption on  $m$ . Since  $G$  and  $H$  coincide on the set  $V(G) \cap V(H)$  of least  $n - 2cn$  vertices,  $G$  can be transformed into the  $K_3$ -free graph  $H$  by changing at most  $2cn^2 \leq \epsilon n^2$  adjacencies, as required. □

### 3. Sketch of proof of $\phi(K_3) \geq h(\phi(K_2))$

Let us sketch the proof of (2.4) from [34, 35], being consistent with the notation defined there. Let  $\rho \stackrel{\text{def}}{=} K_2 \in \mathcal{F}_2^0$ . Consider the ‘defect’ functional  $f(\phi) = \phi(K_3) - h(\phi(\rho))$ , where  $h$  is defined by (1.4). We can identify each homomorphism  $\phi \in \text{Hom}(\mathcal{A}^0, \mathbb{R})$  with the sequence

$$(\phi(F))_{F \in \mathcal{F}^0} \in \mathbb{R}^{\mathcal{F}^0}$$

of its values on graphs. Let us equip all products with the pointwise convergence (or product) topology. The set  $\text{Hom}(\mathcal{A}^0, \mathbb{R})$  is a closed subset of  $\mathbb{R}^{\mathcal{F}^0}$  as the intersection of closed subsets corresponding to the relations that an algebra homomorphism has to satisfy. Thus the set

$$\text{Hom}^+(\mathcal{A}^0, \mathbb{R}) = \bigcap_{F \in \mathcal{F}^0} \{\phi \in \text{Hom}(\mathcal{A}^0, \mathbb{R}) : \phi(F) \geq 0\}$$

is closed too. Moreover, it lies inside the compact space  $[0, 1]^{\mathcal{F}^0}$ , so it is compact as well. Since  $h(x)$  is a continuous function (including the special point  $x = 1$ ), our functional  $f$  is also continuous and achieves its smallest value on  $\text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  at some non-negative homomorphism  $\phi_0$ . Fix one such  $\phi_0$  for the rest of the proof. Let  $a = \phi_0(\rho)$ . Let  $t = t(a)$  and  $c = c(a)$  be defined as in the Introduction. Let  $b = \phi_0(K_3)$ . We have to show that  $b \geq h(a)$ .

If  $a = \phi(\rho) \leq 1/2$ , then  $h(a) = 0$  and there is nothing to do.

Let us write an explicit formula for the function  $h(x)$  defined in (1.4) when  $1 - 1/t \leq x \leq 1 - 1/(t + 1)$ :

$$h_t(x) \stackrel{\text{def}}{=} \frac{(t - 1)(t - 2\sqrt{t(t - x(t + 1))})(t + \sqrt{t(t - x(t + 1))})^2}{t^2(t + 1)^2}. \tag{3.1}$$

If  $a = 1 - 1/(t + 1)$ , then we are done by the well-known bound – proved independently by Moon and Moser [26] and Nordhaus and Stewart [28] – that for every  $0 \leq m \leq \binom{n}{2}$

$$g_3(m, n) \geq \frac{x(x - 1)(x - 2)}{6} \left(\frac{n}{x}\right)^3, \quad x \stackrel{\text{def}}{=} (1 - 2m/n^2)^{-1}. \tag{3.2}$$

So let us assume that  $a$  lies in the open interval  $(1 - 1/t, 1 - 1/(t + 1))$ . Here the function  $h_t(x)$  is differentiable, and it is routine to see that  $h'_t(a) = 3(t - 1)c$ . A calculation-free intuition is that if we add one edge to  $H \in \mathcal{H}_a$  then the number of triangles increases by  $((t - 1)c + o(1))n$  (while the effect of the change in the part sizes is relatively negligible); so we expect that

$$h'_t(a) \binom{n}{2}^{-1} \approx (t - 1)cn \binom{n}{3}^{-1}.$$

Let us see which properties  $\phi_0$  has. Let  $\{G_n\}$  converge to  $\phi_0$  with  $|V(G_n)| = n$ . Let  $\varepsilon > 0$  be a small constant.

It is impossible that at least  $\varepsilon n^2$  edges of  $G_n$  are each in more than  $((t - 1)c + \varepsilon)n$  triangles: by removing a uniformly spread subset of these edges we get a change that is noticeable in the limit and strictly decreases the defect functional  $f$ . Thus, if we pick a random edge from  $E(G_n)$ , then with probability  $1 - o(1)$  there are at most  $((t - 1)c + o(1))n$  triangles containing this edge. (Note that  $G_n$  has  $\Omega(n^2)$  edges by our assumption  $a \geq 1/2$ .) The corresponding flag algebra statement [35, (3.3)] reads

$$\phi_0^E(K_3^E) \leq \frac{1}{3}h'_t(a) \quad \text{a.e. (=almost everywhere)}. \tag{3.3}$$

Let us explain (3.3) informally. It involves counting triangles that contain a specified edge. Let  $\mathcal{F}^E$  consist of  $E$ -flags, by which we mean graphs with some two adjacent vertices being labelled as 1 and 2. Any isomorphism has to preserve the labels. We may represent elements of  $\mathcal{F}^E$  as  $(G; x_1, x_2)$ , where  $G \in \mathcal{F}^0$  is a graph and  $x_i \in V(G)$  is the vertex that gets label  $i$ . Suppose that we wish to keep track of various subgraph densities and their finite linear combinations for  $E$ -flags. We can view  $(F; y_1, y_2) \in \mathcal{F}^E$  as an evaluation on  $\mathcal{F}^E$  that on input  $(G; x_1, x_2)$  returns  $p((F; y_1, y_2), (G; x_1, x_2))$ , the probability that the  $E$ -subflag of  $G$  induced by a random  $|V(F)|$ -set  $X$  with  $\{x_1, x_2\} \subseteq X \subseteq V(G)$  is isomorphic to  $(F; y_1, y_2)$ .

Again, if we know the densities of all  $E$ -flags with  $\ell \geq |V(F)|$  vertices, then we can determine the density of  $(F; y_1, y_2)$  by the analogue of (2.1). So we can define the corresponding linear subspace  $\mathcal{K}^E$  and let  $\mathcal{A}^E \stackrel{\text{def}}{=} \mathbb{R}\mathcal{F}^E/\mathcal{K}^E$ . The obvious analogue of (2.2) holds, and the corresponding coefficients define a multiplication on  $\mathbb{R}\mathcal{F}^E$  that turns  $\mathcal{A}^E$  into a commutative algebra. The multiplicative identity is  $E \in \mathcal{F}^E$ , the unique  $E$ -flag on  $K_2$ . As in the unlabelled case, the limits of convergent sequences of  $E$ -flags are precisely non-negative algebra homomorphisms from  $\mathcal{A}^E$  to the reals ([34, Theorem 3.3]).

Now, we can turn  $G_n$  into an  $E$ -flag by taking a random edge uniformly from  $E(G_n)$  and randomly labelling its endpoints by 1 and 2. Thus for each  $n$  we have a probability distribution on  $E$ -flags which weakly converges to the distribution on  $\text{Hom}^+(\mathcal{A}^E, \mathbb{R})$ , and it is very important that this distribution can be uniquely retrieved from  $\phi_0$  only (see [35, Section 3.2]). In particular, it will not depend on the choice of the representing convergent sequence  $\{G_n\}$ . In (3.3),  $\phi_0^E$  denotes the *extension* of  $\phi_0$  (that is, a random homomorphism from  $\text{Hom}^+(\mathcal{A}^E, \mathbb{R})$  drawn according to this distribution), while  $K_3^E$  is the unique  $E$ -flag with the underlying graph being  $K_3$ .

Let us consider the effect of removing a vertex  $x$  from  $G_n$ . When we first remove  $d(x)$  edges at  $x$ , the edge density goes down by  $d(x)/\binom{n}{2}$ . Next, when we remove the (now isolated) vertex  $x$ , the edge density is multiplied by

$$\binom{n}{2} / \binom{n-1}{2} = 1 + \frac{2}{n} + O(n^{-2}).$$

Thus the edge density changes by

$$-d(x) / \binom{n}{2} + 2a/n + O(n^{-2}).$$

Likewise, the triangle density changes by

$$-K_3^1(x) / \binom{n}{3} + 3b/n + O(n^{-2}),$$

where  $K_3^1(x)$  is the number of triangles per  $x$ . Thus for all but at most  $\varepsilon n$  vertices  $x$  we have

$$(-2d(x)/n + 2a)h'_i(a) < -3K_3^1(x) / \binom{n}{3} + 3b + \varepsilon,$$

for otherwise by removing  $\varepsilon n$  such vertices (and taking the limit of a convergent subsequence of the resulting graphs) we can strictly decrease the defect functional  $f$ . In the flag algebra language this reads as

$$-2h'_i(a)\phi_0^1(K_2^1) + 2h'_i(a)a \leq -3\phi_0^1(K_3^1) + 3b, \quad \text{a.e.}, \tag{3.4}$$

where  $\mathcal{F}^1$  consists of all graphs with one vertex labelled 1,  $K_2^1, K_3^1 \in \mathcal{F}^1$  ‘evaluate’ the edge and triangle density at the labelled vertex, and  $\phi_0^1 \in \text{Hom}^+(\mathcal{A}^1, \mathbb{R})$  is the random extension of  $\phi_0$  constructed similarly<sup>2</sup> to  $\phi_0^E$ .

Note that if we take the expectation of each side of (3.4) with respect to the random  $\phi_0^1 \in \text{Hom}^+(\mathcal{A}^1, \mathbb{R})$ , then we get 0. (A calculation-free intuition is that the edge/triangle density of a graph  $G$  is equal to the average density of edges/triangles sitting on a random vertex of  $G$ .) Thus we conclude that (3.4) is in fact equality a.e. ([35, (3.2)]).

How can (3.3) and (3.4) be converted into statements about  $\phi_0$ ? If, for example, one applies the averaging operator  $[\dots]_1$  ([34, Section 2.2]) to (3.4), that is, taking the expected value of (3.4) over  $\phi_0^1$ , then one obtains the identity  $0 = 0$ , as we have just mentioned. However, one can multiply both sides of (3.4) by some 1-flag  $F$  and then average. (In

<sup>2</sup> Now it is an appropriate place to observe that the superscript in  $\mathcal{F}^0$  refers to the empty type 0.

terms of graphs this corresponds to weighting vertices of  $G_n$  proportionally to the density of  $F$ -subgraphs rooted at them.) What sufficed in [34, 35] was to take  $F = K_2^1$ . Denoting  $e = K_2^1$  for convenience and rearranging terms, we get ([35, (3.4)])

$$\phi_0(3\llbracket eK_3^1 \rrbracket_1 - 2h'_i(a)\llbracket e^2 \rrbracket_1) = a(3b - 2ah'_i(a)). \tag{3.5}$$

Applying the operator  $\llbracket \dots \rrbracket_E$  (averaging over  $\phi_0^E$  and multiplying by the probability that two random vertices induce the type  $E$ ) directly to (3.3) is not useful. Namely, if we take a graph  $G \in \mathcal{H}_a$ , then the graph analogue of (3.3) may have slack for edges that connect two larger parts; thus the obtained inequality will not be best possible. The trick in [34] was first to multiply (3.3) by the  $E$ -flag  $\bar{P}_3^E$  whose graph is the complement of the 3-vertex path. (Thus each edge of  $\mathcal{H}_a$  with slack gets weight 0.) We obtain ([35, (3.5)])

$$\phi_0(\llbracket \bar{P}_3^E K_3^E \rrbracket_E) \leq \frac{1}{3}h'_i(a)\phi_0(\llbracket \bar{P}_3^E \rrbracket_E) = \frac{1}{9}h'_i(a)\phi_0(\bar{P}_3). \tag{3.6}$$

We will also need the following identity, which may be routinely checked (compare with [35, Lemma 3.2]):

$$3\llbracket eK_3^1 \rrbracket_1 + 3\llbracket \bar{P}_3^E K_3^E \rrbracket_E = 2K_3 + K_4 + \frac{1}{4}\bar{K}_{1,3}, \tag{3.7}$$

where  $K_{s,t}$  is the complete bipartite graph with part sizes  $s$  and  $t$ . (Thus  $\bar{K}_{1,3}$  is a triangle plus an isolated vertex.) Also, we have

$$\frac{1}{3}\bar{P}_3 + 2\llbracket e^2 \rrbracket_1 = \rho + K_3. \tag{3.8}$$

Now, if we apply  $\phi_0$  to (3.7) and (3.8) and combine with (3.5) and (3.6), then we obtain the following inequality (see [35, (3.6)], where it is also proved that  $h'_i(a) + 3a - 2 > 0$ ):

$$b \geq \frac{a(2a - 1)h'_i(a) + \phi_0(K_4) + \frac{1}{4}\phi_0(\bar{K}_{1,3})}{h'_i(a) + 3a - 2}. \tag{3.9}$$

If  $\phi_0(\bar{K}_{1,3}) = 0$  and  $\phi_0(K_4)$  is equal to the limiting  $K_4$ -density in  $\mathcal{H}_a$ , then the right-hand side of (3.9) is exactly  $h(a)$ . Thus it remains to bound  $\phi_0(K_4)$  from below. In particular, we are already done if  $a \leq 2/3$  since every graph in  $\mathcal{H}_a$  has no (or very few) copies of  $K_4$ ; this is what was done in [34]. Of course, the result of Nikiforov [27] – who determined  $g_4(a)$  for all  $a$  – would suffice here, but in order to prove our new Theorem 2.1 we need to analyse the argument of [35] further.

Following [35, page 612] define

$$\begin{aligned} A &\stackrel{\text{def}}{=} \frac{2}{3}h'_i(a) = 2(t - 1)c, \\ B &\stackrel{\text{def}}{=} Aa - b = \frac{2}{3}ah'_i(a) - b. \end{aligned} \tag{3.10}$$

Then, for example, (3.4), which is an equality a.e., can be rewritten as

$$\phi_0^1(K_3^1) = A\phi_0^1(e) - B \quad \text{a.e.} \tag{3.11}$$

Also, let us apply the averaging operator  $\llbracket \dots \rrbracket_{E,1}$  to (3.3). Informally speaking, given the labelled vertex  $x_1 \in V(G_n)$ , we pick the second labelled vertex  $x_2$  uniformly at random

and take the expectation of (3.3) multiplied by the indicator function of  $x_1$  and  $x_2$  being adjacent. Since  $\llbracket K_3^E \rrbracket_{E,1} = K_3^1$  and  $\llbracket 1 \rrbracket_{E,1} = \llbracket E \rrbracket_{E,1} = e$ , we get ([35, (3.8)])

$$\phi_0^1(K_3^1) \leq \frac{1}{3} h'_t(a) \phi_0^1(e) = \frac{A}{2} \phi_0^1(e) \quad \text{a.e.} \tag{3.12}$$

The combinatorial meaning of the last step is very simple: if each edge is in at most  $(t - 1)cn$  triangles, then a given vertex  $x_1$  can belong to at most  $\frac{1}{2}d(x_1)(t - 1)cn$  triangles.

From (3.11) and (3.12) we obtain

$$0 < \frac{B}{A} \leq \phi_0^1(e) \leq \frac{2B}{A} \quad \text{a.e.} \tag{3.13}$$

Now let us take any individual  $\phi^1 \in \text{Hom}^+(\mathcal{A}^1, \mathbb{R})$  for which (3.11)–(3.13) hold. Let

$$\psi \stackrel{\text{def}}{=} \phi^1 \pi^e \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R}), \tag{3.14}$$

see [35, page 612]. Informally, we take an arbitrary vertex  $x$  of  $G_n$  and assume that the density of edges/triangles containing  $x$  satisfies (3.11)–(3.13). Then  $\psi$  corresponds to taking the subgraph  $H_n$  of  $G_n$  induced by the neighbourhood of  $x$ . For example, the edge density of  $H_n$  can be calculated by taking the triangle density at  $x$  and multiplying it by

$$\binom{n-1}{2} / \binom{d(x)}{2} \approx \left( \frac{n-1}{d(x)} \right)^2.$$

In the flag algebra formalism this reads ([35, (3.13)])

$$\psi(\rho) = \frac{\phi^1(K_3^1)}{(\phi^1(e))^2} = \frac{A\phi^1(e) - B}{(\phi^1(e))^2} = \frac{z - \mu}{z^2}, \tag{3.15}$$

where following [35, page 612] we define

$$z \stackrel{\text{def}}{=} \phi^1(e)/A \quad \text{and} \quad \mu \stackrel{\text{def}}{=} B/A^2. \tag{3.16}$$

Some calculations based on (3.2) show that ([35, (3.15)])

$$\psi(\rho) \leq 1 - \frac{1}{t}. \tag{3.17}$$

Summarizing (in the graph theory language): the degree of a typical  $x \in V(G_n)$  determines the edge density of  $G_n[N(x)]$ , the subgraph induced by the neighbourhood  $N(x)$  of  $x$ . Moreover, this density is at most  $1 - 1/t + o(1)$ . This gives us a strategy for bounding the number of  $K_4$ 's in  $G_n$  from below: use induction on  $t$  to bound the number of  $K_3$ 's in  $N(x)$  and then sum this over all  $x \in V(G_n)$  (and divide by 4). Unfortunately, this bound on  $\psi(K_3)$  involves radicals and it is not clear how to average it, since  $t(\psi(\rho))$  may assume different values for different choices of  $\phi^1$ . These difficulties are overcome by proving the following lower bound on  $\phi^1(K_4^1) = \psi(K_3)(\phi^1(e))^3$ , which is a linear function of  $\phi^1(e)$  that does not depend on  $t(\psi(\rho))$  ([35, (3.24)]):

$$\phi^1(K_4^1) \geq A^3 \left( \frac{3}{2}(1 - 2\mu) \left( \frac{\phi^1(e)}{A} - \eta_{t-1} \right) + \eta_{t-1}^3 \frac{(t-2)(t-3)}{(t-1)^2} \right), \tag{3.18}$$

where, for  $1 \leq s \leq t - 1$ ,  $\eta_s$  is the unique root of the equation

$$\frac{\eta_s - \mu}{\eta_s^2} = 1 - \frac{1}{s} \tag{3.19}$$

that lies in the interval  $[\mu, 2\mu]$ ; see [35, (3.17)]. Thus the random extension  $\phi_0^1$  satisfies (3.18) a.e. and we can average it, obtaining a lower bound on  $\phi_0(K_4)$ , which is [35, (3.25)]. (Note that the expectation of  $\phi_0^1(K_4^1)$  is  $\phi_0(K_4)$ .) It turns out that this lower bound, when substituted into (3.9), suffices for proving the desired conclusion  $b \geq h(a)$ . The derivations (also those of (3.18)) are rather messy, do not involve any genuine flag algebra calculations and are not needed for our proof. So we omit them and refer the reader to [35] for all details.

#### 4. Proof of Theorem 2.1

All notation here is compatible with that of [34, 35]. As before, let 0, 1, and  $E$  denote the (unique) types with respectively 0, 1 and 2 (adjacent) vertices. Also,

$$\rho \stackrel{\text{def}}{=} K_2 \in \mathcal{F}_2^0 \quad \text{and} \quad e \stackrel{\text{def}}{=} K_2^1 \in \mathcal{F}_2^1$$

are the (unique) 0- and 1-flags having two adjacent vertices. In the arXiv version of our paper (arXiv.org:1204.2846) we offer a *Mathematica* code that verifies some laborious flag algebra (in)equalities that are needed here.

Let  $\Phi \subseteq \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  be the set of the conjectured extremal homomorphisms defined in Section 2. Let  $\phi_0 \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  be arbitrary such that  $\phi_0(K_3) = h(\phi_0(\rho))$ . We have to show that  $\phi_0 \in \Phi$ . Let

$$a \stackrel{\text{def}}{=} \phi_0(\rho) \quad \text{and} \quad b \stackrel{\text{def}}{=} \phi_0(K_3).$$

We prove Theorem 2.1 (that is, the claim that  $\phi_0 \in \Phi$ ) by induction on the parameter  $t = t(a)$  that was defined by (1.1). If  $t = 1$ , then  $a \leq 1/2$ ,  $b = 0$ , and there is nothing to do: every non-negative homomorphism of triangle density 0 is in  $\Phi$  by definition. Let  $t \geq 2$  and assume that we have proved the theorem for all smaller  $t$ .

Suppose first that  $a = 1 - 1/s$  for some integer  $s$ . Apply Theorem 1.2 to any sequence  $\{G_n\}$  convergent to  $\phi_0$ , say with  $|V(G_n)| = n$ , to conclude that  $G_n$  is  $o(n^2)$ -close to the Turán graph  $T_s(n)$  in the edit distance. Clearly, when we change  $o(n^2)$  edges in  $G_n$ , then the density of any fixed graph  $F$  changes by  $o(1)$ , so  $\phi_0$  is still the limit of  $\{G_n\}$ . Since the limit of  $\{T_s(n)\}$  is in  $\Phi$ , we are done in this case.

So let  $a$  lie in the open interval

$$\left(1 - \frac{1}{t}, 1 - \frac{1}{t+1}\right).$$

Let  $c$  be defined by (1.2). We assume that the reader is familiar with the proof in [35]; part of it was sketched in Section 3, and we utilize the notation and facts established there.

Since  $\phi_0$  is extremal, we know that  $b = h(a)$ . This gives some noticeable simplifications to (3.10), (3.16) and (3.19):

$$\begin{aligned} B &= t(t-1)c^2, \\ \mu &= \frac{B}{A^2} = \frac{t}{4(t-1)}, \\ \eta_{t-1} &= 1/2. \end{aligned} \tag{4.1}$$

The *support* of the random extension  $\phi_0^\sigma$  discussed in the previous section is the smallest closed subset of  $\text{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$  of measure 1; it will be denoted by  $S^\sigma(\phi_0)$ . A useful property of the support is that if some closed property has measure 1, then *every* element of  $S^\sigma(\phi_0)$  has this property. We fix an arbitrary  $\phi^1 \in S^1(\phi_0)$ . Inequalities (3.11)–(3.13) hold a.e. and define a closed subset, thus  $\phi^1$  satisfies them. In particular, (3.13) simplifies to

$$0 < \frac{tc}{2} \leq \phi^1(e) \leq tc < 1. \tag{4.2}$$

So, we can define  $\psi$  by (3.14).

Let us prove that  $\psi$  is extremal (that is, has the smallest possible triangle density given its edge density). It is this part of our proof that most heavily relies upon [35]; it basically amounts to checking that the extremality assumption  $b = h(a)$  makes tight sufficiently many useful inequalities proved there.

**Claim 4.1.**  $\psi \in \Phi$  and

$$\psi(\rho) \in \left[ 1 - \frac{1}{t-1}, 1 - \frac{1}{t} \right].$$

**Proof.** Let  $s$  be such that

$$\psi(\rho) \in \left( 1 - \frac{1}{s}, 1 - \frac{1}{s+1} \right].$$

We know that the result of averaging (3.18) (which is [35, (3.25)]) is an equality. Hence (3.18) is equality a.e., and by the same token as before, it holds for every  $\phi^1 \in S^1(\phi_0)$ . The analysis of the calculations in [35] shows that [35, (3.16)] (which is equivalent to  $\psi(K_3) \geq h_s(\psi(\rho))$ ) is also equality. Thus the homomorphism  $\psi \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R})$  is extremal. By (3.17) we have that  $s \leq t - 1$ . The (global) induction assumption implies that  $\psi \in \Phi$ .

We still have to show the second part of the claim when  $t \geq 3$ . Recall that

$$\psi(\rho) = \frac{z - \mu}{z^2}$$

by (3.15). In view of (4.1), the quadratic equation

$$\frac{z - \mu}{z^2} = 1 - \frac{1}{t-1}$$

has two roots:  $z = 1/2$  and  $z = t/(2t - 4)$ . By (4.2), it is impossible that  $z \geq t/(2t - 4)$  (which is equivalent to  $\phi^1(e) \geq t(t - 1)c/(t - 2)$ ). Thus, if we assume that  $s \leq t - 2$ , then

$$\psi(\rho) \leq 1 - \frac{1}{t-1} \quad \text{and} \quad z \leq \frac{1}{2} = \eta_{t-1}.$$

Thus, when we apply the proof of [35, Claim 3.3], the case  $z \leq \eta_{t-1}$  takes place. This implies that [35, (3.21)] is tight. Then [35, (3.23)] is also tight. Its proof on page 615 of [35] shows that this is possible only if  $\mu = (s + 1)/4s$  is the largest element of

$$\left[ \frac{z}{2}, \frac{s+1}{4s} \right],$$

the admissible interval for  $\mu$ . By (4.1) we have that  $s = t - 1$ , as required. □

Claim 4.1 alone suffices to verify Theorem 2.1 in the toy-like case  $\phi_0(\bar{P}_3) = 0$ , where  $\bar{P}_3$  denotes the complement of the 3-vertex path; combinatorially this means that  $\phi_0$  is the limit of complete multipartite graphs. Indeed,  $\phi_0(\bar{P}_3) = 0$  obviously implies that the homomorphism  $\psi$  defined by (3.14) also satisfies  $\psi(\bar{P}_3) = 0$  and, moreover,  $\phi_0$  is equal to the join  $\vee(0, \psi; 1 - \phi^1(e), \phi^1(e))$ . The latter fact readily follows from definitions; combinatorially it means that every vertex  $x$  in a complete multipartite graph  $G_n$  defines its decomposition as the join  $G_n = I_n \vee H_n$ , where  $H_n$  is the subgraph induced by all neighbours of  $x$  and  $I_n$  is the independent set induced by all non-neighbours of  $x$ . Thus, applying Claim 4.1 inductively, we conclude that every  $\phi_0 \in \Phi$  with  $\phi_0(\bar{P}_3) = 0$  necessarily has the form

$$\vee(\underbrace{0, \dots, 0}_{k \text{ times}}; c_1, \dots, c_k),$$

where, say,  $0 < c_1 \leq \dots \leq c_k$ , for some *fixed finite*  $k$ . We are only left to prove that  $c_2 = \dots = c_k$ , and the simplest way of doing this is to invoke [27, Claim 2.13] used by Nikiforov for an essentially identical purpose.

**Claim 4.2.** *Let  $\gamma_3 \geq \gamma_2 \geq \gamma_1 > 0$  be real numbers satisfying*

$$\begin{aligned} \gamma_1 + \gamma_2 + \gamma_3 &= \alpha, \\ \gamma_1\gamma_2 + \gamma_2\gamma_3 + \gamma_3\gamma_1 &= \beta, \end{aligned}$$

*and let  $\gamma_1\gamma_2\gamma_3$  be minimized subject to these two constraints. Then  $\gamma_2 = \gamma_3$ .* □

The case  $\phi_0(\bar{P}_3) > 0$  is way more elaborate, and this is where the main novelty of our contribution lies. We begin with the following claim. The intuition behind it is as follows. Identity (3.11) gives a linear relation between triangle and edge densities via a vertex. By Claim 4.1 we know that (3.11) also holds for the subgraph induced by the neighbourhood of almost every vertex  $x \in V(G)$ . If we average this for all choices of  $x$ , then we get some linear relation between the densities of  $K_4$ ,  $K_3$ , and  $K_2$  that has to hold for all extremal homomorphisms. Repeating, we get a linear relation for  $K_5$ ,  $K_4$ , and  $K_3$ , and so on.

**Claim 4.3.** *For every  $r \geq 3$ , we have*

$$\phi_0(K_r) = 2(t - r + 2)c\phi_0(K_{r-1}) - (t - r + 3)(t - r + 2)c^2\phi_0(K_{r-2}). \tag{4.3}$$

**Proof.** We use induction on  $r$ . If  $r = 3$ , then the identity relates  $b = \phi_0(K_3)$  and  $a = \phi_0(\rho)$ . Both of these parameters have been explicitly expressed in terms of  $c$  and  $t$  and the desired identity (4.3) can be routinely checked.

Suppose that (4.3) is true (for all extremal  $\phi_0$ ). Let us prove it for  $r + 1$ . Let  $\phi^1 \in S^1(\phi_0)$  be arbitrary and let  $\psi = \phi^1\pi^e$ . By Claim 4.1 we know that

$$\psi(\rho) \in \left[ 1 - \frac{1}{t-1}, 1 - \frac{1}{t} \right].$$

Let  $\gamma = c(\psi(\rho))$ , where  $c(x)$  is defined by (1.3), that is,  $\gamma$  is the unique root of

$$2 \left( \binom{t-1}{2} \gamma^2 + (t-1)\gamma(1-(t-1)\gamma) \right) = \psi(\rho) \tag{4.4}$$

with  $\gamma \geq 1/t$ . We have that  $\gamma = c/\phi^1(e)$ . Indeed, this value satisfies (4.4) by (3.15) and is at least  $1/t$  by (4.2). (An informal reason is that all derived inequalities are sharp for  $\Phi$  and, if we pass to a neighbourhood of a vertex in some  $H \in \mathcal{H}_a$ , then its  $t-2$  largest parts have the same (absolute) sizes as the  $t-1$  largest parts of  $H$ .)

By Claim 4.1, we have that  $t(\psi(\rho)) = t-1$ . Thus, by the induction assumption,

$$\psi(K_r) = 2(t-r+1)\gamma\psi(K_{r-1}) - (t-r+2)(t-r+1)\gamma^2\psi(K_{r-2}).$$

If we now substitute

$$\gamma = c/\phi^1(e) \quad \text{and} \quad \psi(K_s) = \phi^1(K_{s+1}^1)/(\phi^1(e))^s,$$

cancel all occurrences of  $(\phi^1(e))^{-r}$ , and average the result, we obtain exactly what we need. □

Let us define  $h^{(r)}(1) = 1$  and, for  $0 \leq x < 1$ ,

$$h^{(r)}(x) \stackrel{\text{def}}{=} r! \left( \binom{t}{r} c^r + \binom{t}{r-1} c^{r-1}(1-tc) \right),$$

where  $c = c(x)$  is again defined by (1.3). In other words,  $h^{(r)}(x)$  is the limiting density of  $K_r$  in the graphs from  $\mathcal{H}_{x,n}$  as  $n \rightarrow \infty$ . (In particular,  $h^{(3)}$  is equal to our function  $h$ .) It is an upper bound on  $g_r(x)$  and, as it was recently shown by Reiher [36], they are in fact equal:  $g_r(x) = h^{(r)}(x)$ .

Claim 4.3 has the following useful corollary.

**Claim 4.4.** *Let  $r \geq 3$ . Then  $\phi_0(K_r) = h^{(r)}(a)$ , that is, each clique has the ‘right’ density. In particular,  $\phi_0(K_s) = 0$  for  $s \geq t+2$ .*

**Proof.** This is true for  $r = 3$  as  $\phi_0(K_3) = g_3(a)$ . The general case follows from Claim 4.3 by induction on  $r$ . □

Recall that we assume  $\phi_0(\bar{P}_3) > 0$  (as the case  $\phi_0(\bar{P}_3) = 0$  was tackled earlier). We need a few auxiliary results. For a graph  $F \in \mathcal{F}_t^0$ , let  $F^{(1)} \in \mathcal{F}_{t+1}^1$  be the 1-flag obtained by adding a new vertex  $x$  that is connected to all vertices of  $F$  (i.e., taking the join  $F \vee K_1$ ) and labelling  $x$  as 1.

**Claim 4.5.**  $\phi_0(\llbracket \bar{P}_3^{(1)} \rrbracket_1) > 0$ .

**Proof.** By Claim 4.4 we have that  $\phi_0(K_4) = h^{(4)}(a)$ . When we substitute this value into (3.9) we obtain a tight inequality except for the extra term involving  $\bar{K}_{1,3}$  (a triangle plus an isolated vertex). We conclude that

$$\phi_0(\bar{K}_{1,3}) = 0. \tag{4.5}$$

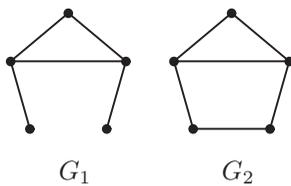


Figure 1. Exceptional graphs.

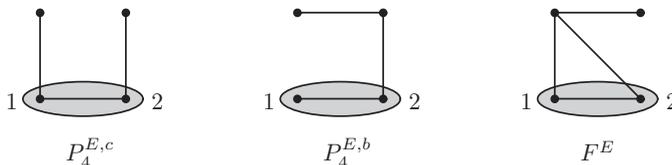


Figure 2. Some E-flags.

Inequality (3.6) is also used in the proof, so it has to be tight. Since we assumed that  $\phi_0(\bar{P}_3) > 0$ , we have that  $\phi_0(\llbracket \bar{P}_3^E K_3^E \rrbracket_E) > 0$ , where  $\bar{P}_3^E$  is the unique E-flag on  $\bar{P}_3$ . But

$$\llbracket \bar{P}_3^E K_3^E \rrbracket_E = \frac{1}{4} \bar{K}_{1,3} + \frac{1}{3} \llbracket \bar{P}_3^{(1)} \rrbracket_1,$$

and the claim follows. □

The two graphs in Figure 1, called  $G_1$  and  $G_2$ , will play a special role.

**Claim 4.6.**  $\phi_0(G_1) = \phi_0(G_2) = 0$ .

**Proof.** We apply the same strategy (although with much more involved calculations) as the one used to prove (4.5). Namely, we make up an analogue of (3.9) that is tight on extremal homomorphisms and such that the ‘overall slackness’ involved will cover  $G_1$  and  $G_2$ .

Form the element  $f^E \in \mathcal{F}_4^E$  as follows:

$$f^E \stackrel{\text{def}}{=} \frac{1}{2} P_4^{E,c} - \frac{1}{2} P_4^{E,b} - F^E,$$

where  $P_4^{E,c}, P_4^{E,b}, F^E \in \mathcal{F}_4^E$  are shown on Figure 2. Since (3.6) is tight,

$$\phi_0^E(K_3^E) < \frac{1}{3} h'_t(a) \implies \phi_0^E(\bar{P}_3^E) = 0 \text{ a.e.}$$

Since both  $P_4^{E,b}$  and  $F^E$  contain  $\bar{P}_3^E$ , this implies that

$$\phi_0^E(K_3^E) < \frac{1}{3} h'_t(a) \implies \phi_0^E(f^E) \geq 0 \text{ a.e.} \tag{4.6}$$

(Recall that  $h_t$  is just the restriction of  $h$  to the interval  $[1 - 1/t, 1 - 1/(t + 1)]$  as defined by (3.1).) Thus, by (3.3), we can multiply the left-hand side of (4.6) by  $f^E$ , obtaining a true inequality. If we apply the averaging operator  $\llbracket \dots \rrbracket_E$  to this new inequality, we get

that

$$\phi_0(\llbracket f^E K_3^E \rrbracket_E) \leq \frac{1}{3} h'_t(a) \phi_0(\llbracket f^E \rrbracket_E). \tag{4.7}$$

Next, similarly to [35, (3.4)] but multiplying [35, (3.2)] (i.e., our formula (3.4) which is equality a.e.) by  $K_3^1$  rather than by  $e$ , we obtain

$$\phi_0(3\llbracket (K_3^1)^2 \rrbracket_1 - 2h'_t(a)\llbracket eK_3^1 \rrbracket_1) = b(3b - 2ah'_t(a)). \tag{4.8}$$

Subtracting (4.8) from (4.7) multiplied by 3, and re-grouping terms, we obtain

$$3\phi_0(\llbracket f^E K_3^E \rrbracket_E - \llbracket (K_3^1)^2 \rrbracket_1) + h'_t(a)\phi_0(2\llbracket eK_3^1 \rrbracket_1 - \llbracket f^E \rrbracket_E) \leq b(2ah'_t(a) - 3b). \tag{4.9}$$

But we also have

$$2\llbracket eK_3^1 \rrbracket_1 - \llbracket f^E \rrbracket_E = \frac{4}{3}K_3 + \frac{2}{3}K_4 - \frac{1}{3}\bar{K}_{1,3} \tag{4.10}$$

and

$$\llbracket f^E K_3^E \rrbracket_E - \llbracket (K_3^1)^2 \rrbracket_1 \geq \frac{1}{60}(G_1 + G_2) - \left( \frac{1}{2}K_4 + \frac{1}{3}\rho K_3 + \frac{1}{6}K_5 \right). \tag{4.11}$$

Substituting these relations into (4.9), and using Claim 4.4, we conclude by (4.5) that

$$\begin{aligned} \frac{1}{20}\phi_0(G_1 + G_2) &\leq b(2ah'_t(a) - 3b) - h'_t(a)\left(\frac{4}{3}b + \frac{2}{3}h^{(4)}(a)\right) \\ &\quad + \left(\frac{3}{2}h^{(4)}(a) + ab + \frac{1}{2}h^{(5)}(a)\right) = 0. \end{aligned}$$

Claim 4.6 is proved. □

**Lemma 4.7.** *Let  $G$  be a graph on  $V = \{x_1, x_2, x_3, y, z\}$  with the following properties. The vertices  $x_1, x_2, x_3$  induce  $\bar{P}_3$  with  $x_1x_2 \in E(G)$ ,  $y$  is adjacent to each  $x_i$  and  $z$  is non-adjacent to at least one  $x_i$ .*

*If  $yz \notin E(G)$ , then  $G$  contains  $\bar{K}_{1,3}$  as an induced subgraph or  $G$  is isomorphic to  $G_1$  or  $G_2$ .*

**Proof.** If  $zx_1, zx_2 \in E(G)$ , then  $zx_3 \notin E(G)$  and  $G - y \cong \bar{K}_{1,3}$ . If  $zx_1, zx_2 \notin E(G)$ , then  $G - x_3 \cong \bar{K}_{1,3}$ . So we can assume without loss of generality that  $zx_1 \in E(G)$  and  $zx_2 \notin E(G)$ . Now, if  $zx_3 \notin E(G)$ , then  $G$  is isomorphic to  $G_1$ ; otherwise  $G \cong G_2$ . □

Now we are ready to put everything together. The next argument would look particularly simple and elegant in genuinely flag-algebraic notation, but it would require introducing some more notions and techniques, notably upward operators ([34, Section 2.3.1]) and relating extensions for different types ([34, Theorem 3.17]). We prefer not to indulge into this endeavour in the concluding part of our paper, so we replace this with (admittedly, crude) translation to the finite world.

Let  $\sigma$  be the 3-vertex type whose graph is  $\bar{P}_3$  with labels 1 and 2 being adjacent. Let  $\{G_n\}$  converge to  $\phi_0$  with  $|V(G_n)| = n$ . By Claim 4.5,  $G_n$  has  $\Omega(n^4)$  copies of  $F_0 \in \mathcal{F}_4^0$ , which denotes a triangle with a pendant edge. Let  $F_1 \in \mathcal{F}_4^1$  be obtained from  $F_0$  by putting label

1 on a vertex of degree 2. Let  $F_3 \in \mathcal{F}_4^\sigma$  be the (unique)  $\sigma$ -flag that can be obtained from  $F_1$  by adding labels 2 and 3.

Fix small positive constants  $\varepsilon \gg \delta$ . Let

$$X = \{x_1 \in V(G_n) : p(F_1, (G_n; x_1)) > \varepsilon\}.$$

By counting copies of  $F_0$  in  $G_n$ , we conclude that

$$2(\phi(F_0) + o(1)) \binom{n}{4} \leq |X| \binom{n-1}{3} + (n - |X|)\varepsilon \binom{n-1}{3},$$

implying that, e.g.,  $|X| \geq \phi(F_0)n/3 \geq 2\varepsilon n$ . An easy counting shows that for every  $x_1 \in X$  there are at least  $\delta n^2$  pairs  $(x_2, x_3)$  of vertices with  $p(F_3, (G_n; x_1, x_2, x_3)) \geq \delta$ . Likewise, by (4.2), the set

$$Y = \{x_1 \in V(G_n) : p(e, (G_n; x_1)) < 1 - \varepsilon\}$$

has size at least  $(1 - \varepsilon)n$ . Thus  $|X \cap Y| \geq \varepsilon n$  and there are at least  $\varepsilon n \cdot \delta n^2$  choices of  $(x_1, x_2, x_3)$  such that  $x_1 \in X \cap Y$  and  $p(F_3, (G_n; x_1, x_2, x_3)) \geq \delta$ . Given such a triple, let  $V_1$  consist of all vertices of  $G_n$  adjacent to all of  $x_1, x_2, x_3$  and let  $V_2 = V(G_n) \setminus V_1$ . We have  $|V_1| \geq \delta(n - 3)$ . Since  $x_1 \in Y$ , we have  $|V_2| \geq \varepsilon(n - 1)$  (note that all non-neighbours of  $x_1$  are in  $V_2$ ). For each non-adjacent  $y \in V_1$  and  $z \in V_2$ , the 5-set  $\{x_1, x_2, x_3, y, z\}$  contains  $G_1, G_2$  or  $\bar{K}_{1,3}$  by Lemma 4.7. By (4.5) and Claim 4.6, each of these graphs has density  $o(1)$  in  $G_n$ . Thus there is a triple  $(x_1, x_2, x_3)$  with  $e(\bar{G}[V_1, V_2]) = o(n^2)$ .

Fix one such choice. By taking a subsequence, we can assume that  $|V_i|/n$  tends to a limit  $\alpha_i$  and that  $G_n[V_i]$  converges to some homomorphism  $\phi_i$ , for  $i = 1, 2$ . Now,  $\phi_0 = \vee(\phi_1, \phi_2, \alpha_1, \alpha_2)$ , where  $\alpha_1 \geq \delta$  and  $\alpha_2 \geq \varepsilon$  are bounded away from 0.

Let  $i = 1$  or  $2$ . Each  $\phi_i$  is an extremal homomorphism: for example, if there exists  $\phi'_1$  with  $\phi'_1(\rho) = \phi_1(\rho)$  and  $\phi'_1(K_3) < \phi_1(K_3)$ , then  $\vee(\phi'_1, \phi_2, \alpha_1, \alpha_2)$  contradicts the extremality of  $\phi_0$ . Since  $\phi_0(K_{t+2}) = 0$  and  $\alpha_{3-i} > 0$ , we have  $\phi_i(K_{t+1}) = 0$  for  $i = 1, 2$ . Turán's theorem implies that  $\phi_i(\rho) \leq 1 - 1/t$ . Thus we can apply the (global) induction and conclude that  $\phi_i \in \Phi$ .

We have proved so far that  $\phi_0$  is a join of two elements from  $\Phi$ ; in particular, it has the form

$$\phi_0 = \vee(\underbrace{0, \dots, 0}_{k \text{ times}}, \psi_1, \psi_2; c_1, \dots, c_k, d_1, d_2), \quad \text{with } c_1, \dots, c_k > 0, \tag{4.12}$$

where  $\psi_1(K_3) = \psi_2(K_3) = 0$ . Let

$$\psi'_i \stackrel{\text{def}}{=} \vee(0, 0; p_i, 1 - p_i),$$

where  $p_i \leq 1/2$  satisfies  $2p_i(1 - p_i) = \psi_i(\rho)$ . Since  $\psi'_i(\rho) = \psi_i(\rho)$  and  $\psi'(K_3) = \psi(K_3) (= 0)$ , after plugging  $\psi'_i$  for  $\psi_i$  into  $\phi_0$ , we will get another extremal homomorphism

$$\phi'_0 \stackrel{\text{def}}{=} \vee(\underbrace{0, \dots, 0}_{k+4 \text{ times}}; c_1, \dots, c_k, d_1 p_1, d_1(1 - p_1), d_2 p_2, d_2(1 - p_2)). \tag{4.13}$$

The equality  $\phi'_0(\bar{P}_3) = 0$ , as we have proved earlier, implies  $\phi'_0 \in \Phi$ , that is, all non-zero weights in (4.12) are equal except for possibly one that is allowed to be smaller than

others. But  $\phi_0(\bar{P}_3) > 0$  which implies that for at least one  $\psi_i$ , say,  $\psi_1$ , we have  $d_1 > 0$  and  $0 < p_1 < 1/2$ . This already creates the exceptional weight  $d_1 p_1$  in (4.13); all other weights must lie in  $\{0, d_1(1 - p_1)\}$ . In particular, either  $d_2 = 0$  or  $p_2 \in \{0, 1/2\}$ ; in the first case  $\psi_2$  can be crossed out from (4.12), and in the second case  $\psi_2 = \psi'_2$  and it can be merged with the first  $k$  terms. Thus,  $\phi_0 \in \Phi$ .

This finishes the proof of Theorem 2.1. □

### References

- [1] Aristoff, D. and Radin, C. (2013) Emergent structures in large networks. *J. Appl. Probab.* **50** 883–888.
- [2] Bollobás, B. (1976) On complete subgraphs of different orders. *Math. Proc. Camb. Phil. Soc.* **79** 19–24.
- [3] Borgs, C., Chayes, J., Lovász, L., Sós, V. T. and Vesztergombi, K. (2008) Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851.
- [4] Chatterjee, S. and Dembo, A. (2014) Nonlinear large deviations. [arXiv:1401.3495](https://arxiv.org/abs/1401.3495)
- [5] Chatterjee, S. and Dey, P. S. (2010) Applications of Stein’s method for concentration inequalities. *Ann. Probab.* **38** 2443–2485.
- [6] Chatterjee, S. and Diaconis, P. (2013) Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461.
- [7] Chatterjee, S. and Varadhan, S. R. S. (2011) The large deviation principle for the Erdős–Rényi random graph. *Europ. J. Combin.* **32** 1000–1017.
- [8] Cummings, J., Král’, D., Pfender, F., Sperfeld, K., Treglown, A. and Young, M. (2013) Monochromatic triangles in three-coloured graphs. *J. Combin. Theory Ser. B* **103** 489–503.
- [9] Das, S., Huang, H., Ma, J., Naves, H. and Sudakov, B. (2013) A problem of Erdős on the minimum number of  $k$ -cliques. *J. Combin. Theory Ser. B* **103** 344–373.
- [10] Erdős, P. (1955) Some theorems on graphs. *Riveon Lematematika* **9** 13–17.
- [11] Erdős, P. (1962) On a theorem of Rademacher–Turán. *Illinois J. Math.* **6** 122–127.
- [12] Erdős, P. (1967) Some recent results on extremal problems in graph theory: Results. In *Theory of Graphs: Rome 1966*, Gordon and Breach, pp. 117–123 (English); pp. 124–130 (French).
- [13] Erdős, P. (1969) On the number of complete subgraphs and circuits contained in graphs. *Časopis Pěst. Mat.* **94** 290–296.
- [14] Erdős, P., Frankl, P. and Rödl, V. (1986) The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.* **2** 113–121.
- [15] Fisher, D. C. (1989) Lower bounds on the number of triangles in a graph. *J. Graph Theory* **13** 505–512.
- [16] Frieze, A. and Kannan, R. (1999) Quick approximation to matrices and applications. *Combinatorica* **19** 175–220.
- [17] Hatami, H., Hladký, J., Král’, D., Norine, S. and Razborov, A. (2013) On the number of pentagons in triangle-free graphs. *J. Combin. Theory Ser. A* **120** 722–732.
- [18] Komlós, J. and Simonovits, M. (1996) Szemerédi’s regularity lemma and its applications to graph theory. In *Combinatorics: Paul Erdős is Eighty* (D. Miklós, V. T. Sós and T. Szőnyi, eds), Vol. 2, Bolyai Mathematical Society, pp. 295–352.
- [19] Lovász, L. (2012) *Large Networks and Graph Limits*, Colloquium Publications, AMS.
- [20] Lovász, L. and Simonovits, M. (1976) On the number of complete subgraphs of a graph. In *Proc. 5th British Combinatorial Conference: Aberdeen 1975. Congress. Numer.* **XV** 431–441.
- [21] Lovász, L. and Simonovits, M. (1983) On the number of complete subgraphs of a graph II. In *Studies in Pure Mathematics*, Birkhäuser, pp. 459–495.
- [22] Lovász, L. and Szegedy, B. (2006) Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957.

- [23] Lubetzky, E. and Zhao, Y. (2014) On the variational problem for upper tails in sparse random graphs. [arXiv:1402.6011](https://arxiv.org/abs/1402.6011)
- [24] Lubetzky, E. and Zhao, Y. (2015) On replica symmetry of large deviations in random graphs. *Random Struct. Alg.* **47** 109–146.
- [25] Mantel, W. (1907) Problem 28. *Wiskundige Opgaven* **10** 60–61.
- [26] Moon, J. W. and Moser, L. (1962) On a problem of Turán. *Publ. Math. Inst. Hungar. Acad. Sci.* **7** 283–287.
- [27] Nikiforov, V. (2011) The number of cliques in graphs of given order and size. *Trans. Amer. Math. Soc.* **363** 1599–1618.
- [28] Nordhaus, E. A. and Stewart, B. M. (1963) Triangles in an ordinary graph. *Canad. J. Math.* **15** 33–41.
- [29] Pikhurko, O. (2011) The minimum size of 3-graphs without four vertices spanning no or exactly three edges. *Europ. J. Combin.* **23** 1142–1155.
- [30] Pikhurko, O. and Vaughan, E. R. (2013) Minimum number of  $k$ -cliques in graphs with bounded independence number. *Combin. Probab. Comput.* **22** 910–934.
- [31] Radin, C. and Sadun, L. (2013) Phase transitions in a complex network. *J. Phys. A* **46** 305002.
- [32] Radin, C. and Yin, M. (2013) Phase transitions in exponential random graphs. *Ann. Appl. Prob.* **23** 2458–2471.
- [33] Radin, C., Ren, K. and Sadun, L. (2014) The asymptotics of large constrained graphs. *J. Phys. A* **47** 175001.
- [34] Razborov, A. (2007) Flag algebras. *J. Symbol. Logic* **72** 1239–1282.
- [35] Razborov, A. (2008) On the minimal density of triangles in graphs. *Combin. Probab. Comput.* **17** 603–618.
- [36] Reiher, C. (2012) The clique density theorem. [arXiv:1212.2454](https://arxiv.org/abs/1212.2454)
- [37] Ruzsa, I. Z. and Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. In *Combinatorics II* (A. Hajnal and V. Sós, eds), North-Holland, pp. 939–945.
- [38] Simonovits, M. (1968) A method for solving extremal problems in graph theory, stability problems. In *Theory of Graphs: Tihany 1966*, Academic, pp. 279–319.
- [39] Turán, P. (1941) On an extremal problem in graph theory (in Hungarian). *Mat. Fiz. Lapok* **48** 436–452.