Asymptotic Structure of Graphs with
the Minimum Number of Triangles

Oleg Pikhurko∗
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213
and
Mathematics Institute
University of Warwick
Coventry CV4 7AL, UK

Alexander Razborov†
Department of Computer Science
University of Chicago
Chicago, IL 60637

April 12, 2012

Abstract

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.

1 Introduction

The famous theorem of Turán [Tur41] determines $\text{ex}(n, K_r)$, the maximum number of edges in a graph with $n$ vertices that does not contain the $r$-clique $K_r$ (see also Mantel [Man07]). 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 $\text{ex}(n, K_r) = (1 - \frac{1}{r-1} + o(1)) \binom{n}{2}$.

Rademacher (unpublished, 1941) proved that a graph with $\text{ex}(n, K_3) + 1$ edges has at least $\lfloor n/2 \rfloor$ triangles.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 [Erd55] 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 Goodman [Goo59], Erdős [Erd62, Erd69], Moon and Moser [MM62], Bollobás [Bo176], Fisher [Fis89], Lovász and Simonovits [LS76, LS83], Razborov [Raz07, Raz08], Nikiforov [Nik11], Reiher [Rei12], and others.

∗Partially supported by the National Science Foundation, Grant DMS-1100215.
†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

\[ g_r(a) \overset{\text{def}}{=} \lim_{n \to \infty} \frac{g_r\left(\binom{n}{2}, n\right)}{\binom{n}{r}} \]

for any given \( a \) 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 recently (for \( r = 3 \) by Razborov [Raz08], for \( r = 4 \) by Nikiforov [Nik11], and for \( r \geq 5 \) by Reiher [Rei12]).

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

\[ a \in \left[1 - \frac{1}{t}, 1 - \frac{1}{t+1}\right] \tag{1} \]

and let real

\[ c = \frac{t + \sqrt{t(t-a(t+1))}}{t(t+1)} \tag{2} \]

be the (unique) root of the quadratic equation

\[ 2\left(\binom{t}{2}c^2 + tc(1-tc)\right) = a \tag{3} \]

with \( c \geq \frac{1}{t+1} \). Since \( a \geq 1 - \frac{1}{t} \), it follows from (3) that \( c \leq \frac{1}{t} \). Partition the vertex set \([n] = \{1, \ldots, n\}\) into \( t + 1 \) parts \( V_1, \ldots, V_{t+1} \) with \(|V_1| = \cdots = |V_t| = \lfloor cn \rfloor \) for \( i \in [t] \). Let \( G \) be obtained from the complete \( t \)-partite graph \( K(V_1, \ldots, V_{t-1}, U) \), where \( U = V_t \cup V_{t+1} \), by adding an arbitrary triangle-free graph on \( U \) with \(|V_1| |V_{t+1}| \) edges. One possible choice is to take \( G[U] = K(V_t, V_{t+1}) \), resulting in \( G = K(V_1, \ldots, V_{t+1}) \). Since each edge of \( G[U] \) belongs to exactly \(|V_1| + \cdots + |V_{t-1}| \) triangles, the choice of \( G[U] \) has no effect on the triangle density. Clearly, the edge density of \( G \) is \( a + o(1) \). Thus \( g_3(a) \leq h(a) \), where

\[ h(a) \overset{\text{def}}{=} 6\left(\binom{t}{3}c^3 + \binom{t}{2}c^2(1-tc)\right). \tag{4} \]

If \( a = 1 \), we let \( G \) be the complete graph \( K_n \) and define \( h(1) = 1 \). For \( a \in [0, 1] \), let \( \mathcal{H}_{a,n} \) be the set of all possible graphs \( G \) on \([n] \) that arise this way, \( \mathcal{H}_a \overset{\text{def}}{=} \bigcup_{n \in \mathbb{N}} \mathcal{H}_{a,n} \), and \( \mathcal{H} \overset{\text{def}}{=} \bigcup_{a \in [0,1]} \mathcal{H}_a \).

If \( a = 1 - \frac{1}{s} \) for integer \( s \geq 1 \), then we have two choices for \( t \) but \( c = \frac{1}{s} \) and \( h(a) = 6\left(\binom{s}{3}/s^3\right) \) are still uniquely defined. 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 [Raz08, Nik11, Rei12] 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.
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.

This theorem is obtained by building upon the flag algebra approach from [Raz08]. 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 [Raz08] 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. [DHM+12, HHK+11, Pik11, Pik12].)

Our initial motivation was the following conjecture of Lovász and Simonovits [LS76, Conjecture 1] for \( r = 3 \).

**Conjecture 1.2** 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 this conjecture is proved, then one may consider the problem of determining \( g_r(m,n) \) as combinatorially solved: the number of \( K_r \)-subgraphs in a such 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. [Nik11, Section 3]) that the optimal part ratios are approximately as 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.)

We hope that Theorem 1.1 may help in proving Conjecture 1.2 in the same way as the so-called stability approach is useful in obtaining exact results. One example where this approach succeeded is the clique minimization problem in the special case when \( a = 1 - \frac{1}{t} \) for some integer \( t \geq 2 \). First, the results of Goodman [Goo59] (for \( r = 3 \)) and Moon and Moser [MM62] (for \( r \geq 4 \)) imply that for any \( m, n \) we have

\[
g_r(m, n) \geq \frac{t(t-1)\ldots(t-r+1)}{r!} \left( \frac{n}{t} \right)^r, \quad \text{if } t \geq r - 1,
\]

where the real \( t \) is defined by \( m = (1 - 1/t)n^2/2 \). (Note that the Turán graph \( T_t(n) \) shows that (5) is asymptotically best possible when \( t \) is an integer; thus \( g_r(1 - \frac{1}{t}) = r!\binom{t}{r}/t^r \) in this case.) While Moon and Moser [MM62] only indicate how to prove (5), a complete proof can be found in Lovász and Simonovits [LS83, Theorem 1] who also deduced that all almost extremal graphs are close to \( T_t(n) \) in the edit distance:

**Theorem 1.3** For every \( r \) and \( \varepsilon > 0 \), there are \( \delta > 0 \) and \( n_0 \) such that, for any integer \( t \geq r - 1 \), every graph \( G \) with \( n \geq n_0 \) vertices, \( (1 - \frac{1}{t} \pm \delta)\binom{n}{2} \) edges, and at most \( (g_r(1 - \frac{1}{t}) + \delta)\binom{n}{r} \) copies of \( K_r \) can be made isomorphic to \( T_t(n) \) by changing at most \( \varepsilon \binom{n}{2} \) edges.
In fact, a sharper form of this result (with an explicit $\delta = \delta(r, t, \varepsilon, n)$) was proved by Lovász and Simonovits [ES83] who used it to establish Conjecture 1.2 when $\text{ex}(n, K_s) \leq m \leq \text{ex}(n, K_s) + \varepsilon n^2$ for some $\varepsilon = \varepsilon(r, s) > 0$. Unfortunately, we have not succeeded in proving the case $r = 3$ of Conjecture 1.2 so far.

This paper is organized as follows. We outline the main ideas behind flag algebras and state some of the key inequalities from [Raz08] 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 [Raz08] 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 [Raz07]. We do not see any reasonable way of making this paper self-contained, without making it quite long and repeating large passages from [Raz07]. Therefore, we restrict ourselves to sketching the proofs in [Raz07, Raz08], during which we informally illustrate the main ideas by providing some analogs from the discrete world. This serves two purposes: to state the key inequalities from [Raz07, Raz08] that we need here and to provide some guiding intuition for the reader who is about to start reading [Raz07]. We stress that some flag algebra concepts do not have direct combinatorial analogs or require a plethora of constants to state them in terms of graphs. Here we just try to distill and present some motivational ideas.

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 that 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 unlabeled and pairwise non-isomorphic. Let $\mathcal{F}^0$ consist of all (unlabeled non-isomorphic) graphs and let $\mathbb{R}\mathcal{F}^0$ be the vector space that has $\mathcal{F}^0$ for 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:

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

(6)

where $\mathcal{F}_\ell^0 \subset \mathcal{F}_\ell^0$ 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

$$\mathcal{A}_0 \overset{\text{def}}{=} \mathbb{R}\mathcal{F}_0/\mathcal{K}_0.$$ 

By (6), any element of $\mathcal{A}_0$ can still be identified with an evaluation on (sufficiently large) graphs. A theorem of Whitney [Whi32, Theorem 5a] implies that a linear combination $\sum_{i=1}^m \alpha_i F_i$ that always evaluates to 0 is necessarily in $\mathcal{K}_0$ (see the discussion in [HN11, Page 551]).

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 $U_1, U_2 \subset V(G)$ 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 $|U_1| = \ell_1$. For $F \in \mathcal{F}_{\ell_1 + \ell_2}^0$, let $p(F_1, F_2; F)$ denote the probability that $G[U_1] \cong F_1$ (i.e. the subgraph of $G$ induced by $U_1$ is isomorphic to $F_1$) for both $i = 1, 2$ when conditioned on $G[U] \cong F$. This probability depends on $F_1$, $F_2$, and $F$ only. Thus

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

(7)

Since we are interested in the case when $n \to \infty$, 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 $\mathcal{A}_0$ into a commutative associative algebra with the empty graph being the multiplicative identity, see [Raz07, Lemma 2.4].

Unfortunately, we do not have the property that graph evaluations preserve multiplication exactly. This can be fixed 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 $|V(G_1)| < |V(G_2)| < \ldots$ (we call such sequences increasing) and for every graph $F$ the limit

$$\phi(F) \overset{\text{def}}{=} \lim_{n \to \infty} p(F, G_n)$$

(8)

exists. We extend $\phi$ by linearity to $\mathbb{R}\mathcal{F}_0$. It is routine to check that $\phi$ 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 still denote by $\phi$), see [Raz07, Theorem 3.3]. We say that $\phi$ is the limit of $\{G_n\}$ and, following the notation in [Raz07, Section 3.1], denote this as $\phi = \lim_{n \to \infty} p^{G_n}$. 

5
Clearly, \( \phi \) is non-negative, that is, \( \phi(F) \geq 0 \) for every graph \( F \). Let \( \text{Hom}^+(A^0, \mathbb{R}) \) be the set of all non-negative homomorphisms.

It turns out that every non-negative homomorphism \( \phi : A^0 \to \mathbb{R} \) is the limit of some sequence of graphs. It is instructive to sketch a proof of this, see Lovász and Szegedy [LS06, Lemma 2.4] (or [Raz07, Theorem 3.3]) for details. Take some integer \( n \). Since the identity \( \sum_{F \in \mathcal{F}} F = 1 \) holds in \( A^0 \), we have that \( \sum_{F \in \mathcal{F}_n} \phi(F) = 1 \), that is, \( \phi \) defines some probability distribution on \( \mathcal{F}_n \). Let \( G_{n,\phi} \in \mathcal{F}_n \) be drawn according to this distribution with the choices for different values of \( n \) being independent. Fix some \( F \) and \( \varepsilon > 0 \). Let \( n \geq |V(F)| \). Easy calculations show 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 \( O(1/n) \). Chebychev’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, \( \varepsilon \in \{1, \frac{1}{2}, \frac{1}{3}, \ldots \} \), 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. [HHK+11, Section 2.2].

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 show that

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

then we can conclude that indeed \( g_3(a) = h(a) \), as it was done by [Raz08].

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

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

$$\phi(F) = \frac{1}{\text{aut}(F)} \sum_{(V_1, \ldots, V_k)} \left( \frac{|V(F)|}{|V_1|, \ldots, |V_k|} \right)^k \prod_{i=1}^k \left( \alpha_i^{|V_i|, \phi_i(F[V_i])} \right),$$

where the summation runs over all possible ways (up to isomorphism) to partition $V(F) = V_1 \cup \cdots \cup V_k$ into $k$ labeled 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 $K^0$) 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

$$\bigvee(0, \ldots, 0, \psi; c_1, \ldots, c_t - 1 - (t - 1)c),$$

where $0$ denote the (unique) non-negative homomorphism of edge-density $0$, $1/(t+1) \leq c \leq 1/t$, and $\psi$ is an arbitrary non-negative homomorphism with $\psi(K_3) = 0$ and $\psi(K_2) = 2c(1-tc)/(1-(t-1)c)^2$.

Our next 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 [FK99]. We omit the definition of $\delta_{\square}$ but refer the reader to [BCL+08, Definition 2.2].

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 $a = \phi_0(K_2)$. By (9), we know that $\phi_0(K_3) = h(a)$. 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 $G_n$ and $H_n$ have asymptotically same statistics of fixed subgraphs. This does not necessarily implies 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 et al [BCL+08] 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 [BCL+08, Theorem 2.3], we can relabel $V(H_n)$ so that for every disjoint $S, T \subseteq V(G_n)$ we have

$$|e(G_n[S,T]) - e(H_n[S,T])| = o(\nu^2),$$

(11)
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 
\[
V(H_n) = V_1 \cup \cdots \cup V_{t-1} \cup U
\]
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 (11), then we conclude that the number of \( S - T \) edges that are missing from 
\( G_n \) is \( o(v^2) \). Also, the size of \( G[V_i] \) is \( o(v^2) \) for otherwise a random partition \( V_i = S \cup T \) would contradict (11). 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 easily 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 Removal Lemma [RS78, EFR86] (see e.g. [KS96, 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 \( e(H_n[U]) \) are left, in which case the obtained graph \( G_n \) belongs to \( H_{a,n} \) 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 
\[
e(G) < s \leq \min(e(G) + \varepsilon n^2, \lfloor n^2/4 \rfloor)
\]
(12)
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 \( c \gg \delta \). Let \( n \) be large and let \( s \) satisfy (12). 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 by the Stability Theorem of Erdős [Erd67] and Simonovits [Sim68] 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 $m' \geq m + cn\left(\frac{4m}{n} - \delta n\right) + (cn)^2$ 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
$$\frac{(m + cn\left(\frac{4m}{n} - \delta n\right) + (cn)^2)}{\binom{n}{2}} \geq m + \frac{c^2(n^2 - 4m) - 2c\delta n^2}{(1 + 2c)^2}. $$
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 $2cn^2 \leq \varepsilon n^2$ adjacencies, as required.

3 Sketch of Proof of $\phi(K_3) \geq h(\phi(K_2))$

Let us sketch the proof of (9) from [Raz07, Raz08], being consistent with the notation defined
there. Consider the “defect” functional $f(\phi) = \phi(K_3) - h(\phi(K_2))$, where $h$ is defined by (4).
We can identify each homomorphism $\phi \in \text{Hom}(\mathcal{A}^0, \mathbb{R})$ with the sequence
$$p_\phi \overset{\text{def}}{=} (\phi(F))_{F \in \mathcal{F}^0} \subseteq \mathbb{R}^{\mathcal{F}^0}$$
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 con-
tinuous and assumes 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 $a = \phi_0(K_2)$. 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 $a = \phi(K_2) \leq 1/2$, then $h(a) = 0$ and there is nothing to do.

Let us write an explicit formula for the function $h(t)$ defined in (4) when $1 - \frac{1}{t} \leq x \leq 1 - \frac{1}{t + 1}$:
$$h_t(x) \overset{\text{def}}{=} \frac{(t - 1)(t - 2\sqrt{t(t - x(t + 1))})(t + \sqrt{t(t - x(t + 1)})^2}{t^4(t + 1)^2}. \quad (13)$$

If $a = 1 - \frac{1}{t}$ or $a = 1 - \frac{1}{t + 1}$, then we are done by Goodman’s bound (that is, the case $r = 3$
of (5)). So let us assume that $a$ lies in the open interval $(1 - \frac{1}{t}, 1 - \frac{1}{t + 1})$. 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
$h_t'(a)\binom{n}{2}^{-1} \approx (t - 1)c\binom{n}{3}^{-1}$. 

9
Let us see which properties $\phi_0$ has. Let $\{G_n\}$ converge to $\phi_0$ with $|V(G_n)| = n$. Let $\varepsilon > 0$ be fixed.

It is impossible that at least $\varepsilon n^2$ edges of $G_n$ are each in more than $((t - 1)c + \varepsilon)n$ triangles. If we remove 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 [Raz08, (3.3)] reads

$$\phi_0^E(K_2^E) \leq \frac{1}{3} h_1^i(a) \quad \text{a.e. (almost everywhere).} \quad (14)$$

Let us informally explain (14). It involves counting triangles that contain a specified edge. Let $F^E$ consist of $E$-flags, by which we mean graphs with some two adjacent vertices being labeled as 1 and 2. Any isomorphism has to preserve the labels. We may represent elements of $F^E$ as $(G; x_1, x_2)$, where $G \in F^0$ 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 $(F; y_1, y_2) \in F^E$ as an evaluation on $F^E$ 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 $\ell \geq |F|$ vertices, then we can determine the density of $(F; y_1, y_2)$ by the analogy of (6). So we can define the corresponding linear subspace $K^E$ and let $A^E \overset{\text{def}}{=} \mathbb{R} F^E / K^E$. The obvious analog of (7) holds, and the corresponding coefficients define a multiplication on $\mathbb{R} F^E$ that turns $A^E$ into a commutative algebra. The multiplicative identity is $E \in F^E$, the unique $E$-flag on $K_2$. As in the unlabeled case, the limits of convergent sequences of $E$-flags are precisely non-negative algebra homomorphisms from $A^E$ to the reals ([Raz07, 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 labeling its endpoints by 1 and 2. Thus for each $n$ we have a probability distribution on $E$-flags which weakly converges to the (unique) distribution on $\text{Hom}^+(A^E, \mathbb{R})$, see [Raz08, Section 3.2]. In (14), $\phi_0^E$ denotes the extension of $\phi_0$ (that is, a random homomorphism 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 $\binom{n}{2}/(n-1) = 1 + \frac{2}{n} + \mathcal{O}(n^{-2})$. Thus the edge density changes by $-d(x)/\binom{n}{2} + 2a/n + \mathcal{O}(n^{-2})$. Likewise, the triangle density changes by $-3K_3^E(x)/\binom{n}{3} + 3b/n + \mathcal{O}(n^{-2})$, where $K_3^E(x)$ is the number of triangles per $x$. Thus for all but at most $\varepsilon n$ vertices $x$ we have $(-2d(x)/n + 2a)h_1^i(a) < -3K_3^E(x)/\binom{n}{3} + 3b + \varepsilon$, for otherwise by removing such vertices (and taking the limit as $n \to \infty$) we can strictly decrease the defect.
functional $f$. In the flag algebra language this reads

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

where $\mathcal{F}^1$ consists of all graphs with one vertex labeled 1, $K_2^1, K_3^1 \in \mathcal{F}^1$ “evaluate” the edge and triangle density at the labeled vertex, and $\phi_0^1 \in \text{Hom}^+(A^1, \mathbb{R})$ is the random extension of $\phi_0$.

Note that if we take the expectation of each side of (15) with respect to the random $\phi_0^1 \in \text{Hom}^+(A^1, \mathbb{R})$, 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 (15) is in fact equality a.e. ([Raz08, (3.2)]).

How can (14) and (15) be converted into statements about $\phi_0^1$? If, for example, one applies the averaging operator $[\ldots]_1$ ([Raz07, Section 2.2]) to (15), that is, taking the expected value of (15) over $\phi_0^1$, then one obtains the identity $0 = 0$, as we have just mentioned. However, one can multiply both sides of (15) by some 1-flag $F$ and then average. (In terms of graphs this corresponds to weighting vertices of $G$ proportionally to the density of $F$-subgraphs rooted at them.) What sufficed in [Raz07, Raz08] was to take $F = K_2^1$. Denoting $e = K_2^1$ for convenience and rearranging terms, we get ([Raz08, (3.4)]):

$$\phi_0^1(3\left[\begin{array}{c}eK_3^1\end{array}\right]_1 - 2h'_t(a)\left[\begin{array}{c}e^2\end{array}\right]_1) = a(3b - 2ah'_t(a)).$$

Applying the operator $[\ldots]_E$ (averaging over $\phi_0^E$) directly to (14) is not useful. Namely, if we take a graph $G \in \mathcal{H}_a$, then the graph analog of (14) may have slack for edges that connect two larger parts; thus the obtained inequality will not be best possible. The trick in [Raz07] was first to multiply (14) 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 ([Raz08, (3.5)]):

$$\phi_0^E([\bar{P}_3^E K_3^1]_E) \leq \frac{1}{3} h'_t(a)\phi_0^E([\bar{P}_3^E]_E) = \frac{1}{9} h'_t(a)\phi_0^E(\bar{P}_3).$$

We will also need the following identity which may be routinely checked (compare with [Raz08, Lemma 3.2]):

$$3\left[\begin{array}{c}eK_3^1\end{array}\right]_1 + 3\left[\begin{array}{c}\bar{P}_3^E K_3^1\end{array}\right]_E = 2K_3 + K_4 + \frac{1}{4} \bar{K}_{1,3},$$

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

$$\frac{1}{3} \bar{P}_3 + 2\left[\begin{array}{c}e^2\end{array}\right]_1 = K_2 + K_3.$$  

Now, if we apply $\phi_0$ to (18) and (19) and combine with (16) and (17), then we obtain the following inequality (cf. [Raz08, (3.6)]):

$$b \geq \frac{a(2a - 1)h'_t(a) + \phi_0(K_4) + \frac{1}{4} \phi_0(\bar{K}_{1,3})}{h'_t(a) + 3a - 2}.$$
If $\phi_0(K_{1,3}) = 0$ and $\phi_0(K_4)$ is equal to the limiting $K_4$-density in $\mathcal{H}_a$, then the right-hand side of (20) 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 [Raz07]. Of course, the result of Nikiforov [Nik11] who determined $g_4(a)$ for all $a$ would suffice here but in order to prove our new Theorem 2.1 we need to analyze the argument of [Raz08] further.

Following [Raz08, Page 612] define
\begin{align*}
A & \overset{\text{def}}{=} \frac{2}{3}h'(a) = 2(t - 1)c, \\
B & \overset{\text{def}}{=} Aa - b = \frac{2}{3}ah'(a) - b.
\end{align*}

Then, for example, (15), which is an equality a.e., can be rewritten as
\begin{equation}
\phi_0^1(K^3_3) = A\phi_0^1(e) - B \quad \text{a.e.} \tag{21}
\end{equation}

Also, let us apply the averaging operator $\llbracket \ldots \rrbracket_{E,1}$ to (14). Informally speaking, given the labeled vertex $x_1 \in V(G_n)$, we pick the second labeled vertex $x_2$ uniformly at random and take the expectation of (14) multiplied by the indicator function of $x_1$ and $x_2$ being adjacent. Since $[K^E_3]_{E,1} = K^1_3$ and $[E]_{E,1} = e$, we get ([Raz08, (3.8)])
\begin{equation}
\phi_0^1(K^1_3) \leq \frac{1}{3}h'(a)\phi_0^1(e) = \frac{A}{2}\phi_0^1(e) \quad \text{a.e.} \tag{22}
\end{equation}

From (21) and (22) we obtain
\begin{equation}
0 < \frac{B}{A} \leq \phi_0^1(e) \leq \frac{2B}{A} \quad \text{a.e.} \tag{23}
\end{equation}

Now let us take any individual $\phi^1 \in \text{Hom}^+(\mathcal{A}^1, \mathbb{R})$ for which (21)–(23) hold. Let
\begin{equation}
\psi \overset{\text{def}}{=} \phi^1\pi^e \in \text{Hom}^+(\mathcal{A}^0, \mathbb{R}), \tag{24}
\end{equation}

see [Raz08, Page 612]. Informally, we take an arbitrary vertex $x$ for $G_n$ and assume that the density of edges/triangles containing $x$ satisfies (21)–(23). Then $\psi$ correspond to taking the subgraph $H_n$ of $G_n$ induced by the neighborhood of $x$. For example, the edge density of $H_n$ can be calculated by taking the triangle density at $x$ and multiplying it by $\left(\frac{n-1}{2}\right)/(\frac{d(x)}{2}) \approx \left(\frac{n-1}{d(x)}\right)^2$.

In the flag algebra formalism this reads ([Raz08, (3.13)])
\begin{equation}
\psi(K_2) = \frac{\phi^1(K^1_3)}{\phi^1(e)^2} = \frac{A\phi^1(e) - B}{\phi^1(e)^2} = \frac{z - \mu}{z^2}, \tag{25}
\end{equation}

where following [Raz08, Page 612] we define
\begin{equation}
z \overset{\text{def}}{=} \frac{\phi^1(e)}{A} \quad \text{and} \quad \mu \overset{\text{def}}{=} \frac{B}{A^2}. \tag{26}
\end{equation}

Some calculations based on Goodman’s bound (5) show that ([Raz08, (3.15)])
\begin{equation}
\psi(K_2) \leq 1 - \frac{1}{t}, \tag{27}
\end{equation}

12
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 neighborhood $N(x)$ of $x$. Moreover, this density is below $1 - \frac{1}{t} + o(1)$. 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. Another complication is that $t(\psi(K_2))$ may assume different values for different choices of $\phi$. Both of these difficulties are overcome by proving the following lower bound on $\phi(K_4) = \psi(K_3)(\phi(e))^3$ which is a linear function of $\phi(e)$ that does not depend on $t(\psi(K_2))$ ([Raz08, (3.24)]):

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

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

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

that lies in the interval $[\mu, 2\mu]$, see [Raz08, (3.17)]. Thus the random extension $\phi_0$ satisfies (28) a.e. and we can average it, obtaining a lower bound on $\phi_0(K_4)$, which is Inequality (3.25) in [Raz08]. (Note that the expectation of $\phi_0(K_4)$ is $\phi_0(K_4)$.) It turns out that this lower bound, when substituted into (20) suffices for proving the desired conclusion $b \geq h(a)$. The derivations (also those of (28)) are rather messy and do not involve any generic flag algebras calculations. So we omit them and refer the reader to [Raz08] for all details.

4 Proof of Theorem 2.1

All notation here is compatible with that of [Raz07, Raz08]. We exclusively work with the theory of graphs. As before, let 0, 1, and $E$ denote the (unique) types with respectively 0, 1 and 2 (adjacent) vertices. Also, $\rho \overset{def}{=} K_2 \in F_2^0$ and $e \overset{def}{=} K_2^1 \in F_2^1$ are the (unique) 0- and 1-flags having two adjacent vertices.

We offer a Mathematica code that verifies some laborious flag algebra (in)equalities from this paper at this URL: http://homepages.warwick.ac.uk/~maskat/Papers/K3.nb

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

We prove Theorem 2.1 (that is, the claim that $\phi_0(\Phi)$ by induction on $t = t(a)$ (cf. (1)). If $t = 1$, then $a \leq 1/2$ (e.g. by Goodman’s bound (5)), $b = 0$, 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 $a = 1 - \frac{1}{s}$ for some integer $s$. Apply Theorem 1.3 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 $(1 - \frac{1}{t}, 1 - \frac{1}{t+1})$. Let $c$ be defined by (3). We assume that the reader is familiar with the proof in [Raz08] (that we sketched in Section 3) and we use the facts established there. In particular, we know that $b = h(a)$. This gives some noticeable simplifications:

\[
\begin{align*}
B &= t(t-1)c^2, \\
\mu &= \frac{B}{A^2} = \frac{t}{4(t-1)}, \\
\eta_{t-1} &= 1/2.
\end{align*}
\]

(30)

Since any type $\sigma$ is simply a labeled graph [Raz07, Section 2], after unlabeled its vertices, $\sigma$ can be viewed as an element of $F^0$. If $\phi_0(\sigma) > 0$, then the random extension $\phi^\sigma_0 \in \text{Hom}^+(A^\sigma, \mathbb{R})$ is well-defined ([Raz08, Section 3.2]). Let $S^\sigma(\phi_0)$ denote its support, that is, the smallest closed subset of $\text{Hom}^+(A^\sigma, \mathbb{R})$ of measure 1. 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 (21)–(23) hold a.e., thus $\phi^1$ satisfies them. In particular, (23) simplifies to

\[0 < \frac{tc}{2} \leq \phi^1(e) \leq tc < 1.\]

(31)

So, we can define $\psi$ by (24).

Let us prove that $\psi$ is extremal (that is, has the smallest possible triangle density given its edge density).

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

We know that [Raz08, (3.25)] is equality for every $\phi^1 \in S^1(\phi_0)$. Since it is obtained by averaging [Raz08, (3.24)] (our formula (28)), the latter is also equality. The analysis of the calculations in [Raz08] shows that [Raz08, (3.16)] (which is equivalent to $\psi(K_3) \geq h_s(\psi(\rho))$) is also equality. Thus the homomorphism $\psi \in \text{Hom}^+(A^0, \mathbb{R})$ is extremal. By (27) 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 (25). In view of (30), the quadratic equation $\frac{z - \mu}{z^2} = 1 - \frac{1}{t-1}$ has two roots: $z = \frac{1}{2}$ and $z = \frac{t}{2(t-2)}$. By (31), it is impossible that $z \geq \frac{t}{2(t-2)}$ (which is equivalent to $\phi^1(e) \geq \frac{t(t-1)}{2t}c$). Thus, if we assume that $s \leq t - 2$, then $\psi(\rho) \leq 1 - \frac{1}{t-1}$ and $z \leq 1/2 = \eta_{t-1}$.

Thus, when we apply the proof of [Raz08, Claim 3.3], the case $z \leq \eta_{t-1}$ takes place. This implies that [Raz08, (3.21)] is tight. Then [Raz08, (3.23)] is also tight. Its proof on Page 615 of [Raz08] shows that this is possible only if $\mu = \frac{1}{4} \phi^1$ is the largest element of $[\frac{1}{2}, \frac{s+1}{4s}]$, the admissible interval for $\mu$. By (30) we have that $s = t - 1$, as required.\[\blacksquare\]
The intuition behind the following claim is as follows. Identity (21) gives a linear relation between triangle and edge densities via a vertex. By Claim 4.1 we know that (21) also holds for the subgraph induced by the neighborhood 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 4.2 For every \( r \geq 3 \), we have

\[
\phi_0(K_r) = 2(t - r + 2)c\phi_0(K_{r-1}) - (t - r + 3)(t - r + 2)c^2\phi_0(K_{r-2}).
\] (32)

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 (32) can be routinely checked.

Suppose that (32) 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 \( \psi(\rho) \in [1 - \frac{1}{r-1}, 1 - \frac{1}{t}] \). Let \( \gamma = c(\psi(\rho)) \), where \( c(x) \) is defined by (3), that is, \( \gamma \) is the unique root of

\[
2 \left( \binom{t-1}{2} \gamma^2 + (t - 1)\gamma(1 - (t - 1)\gamma) \right) = \psi(\rho)
\] (33)

with \( \gamma \geq 1/t \). We have that \( \gamma = c/\phi^1(e) \). Indeed, this value satisfies (33) by (25) and is at least 1/t by (31). (An informal reason is that all derived inequalities are sharp for \( \Phi \) and, if we pass to a neighborhood 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 \( \gamma = c/\phi^1(e) \) and \( \psi(K_s) = \phi^1(K^1_{s+1})/(\phi^1(e))^s \), cancel all occurrences of \( (\phi^1(e))^{-r} \), and average the result, we obtain exactly what we needed.

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

\[
h^{(r)}(x) \overset{\text{def}}{=} r! \left( \binom{t}{r} e^x + \binom{t}{r-1} e^{x-1}(1 - tc) \right),
\]

where \( c = c(x) \) is defined by (3). In other words, \( h^{(r)}(x) \) is the limiting density of \( K_r \) in the graphs from \( \mathcal{H}_{x,n} \) as \( n \to \infty \). It is an upper bound on \( g_r(x) \) and, as it was recently shown by Reiher [Rei12], we have in fact equality.

Claim 4.2 has the following useful corollary.

Claim 4.3 Let \( r \geq 3 \). Then \( \phi_0(K_r) = h^{(r)}(a) \), that is, the density of each clique is determined. 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 4.2 by induction on \( r \).

The following lemma verifies Theorem 2.1 if \( \phi_0 \) is the limit of complete partite graphs. Note the special role of \( c_0 \): it corresponds to the proportion of vertices that are in parts of relative size \( o(1) \).

**Lemma 4.4** Let \( 0 \leq a < 1 \). Let reals \( c_1 \geq \cdots \geq c_s > 0 \) and \( c_0 \geq 0 \) satisfy \( \sum_{i=0}^{s} c_i = 1 \) and \( c_0^2/2 + \sum_{0 \leq i < j \leq s} c_i c_j = a/2 \). Then

\[
\frac{c_0^3}{3!} + \frac{c_0^2}{2!} (1 - c_0) + \sum_{0 \leq i < j < h \leq s} c_i c_j c_h \geq \frac{h(a)}{3!}
\]

with equality if and only if \( c_0 = 0 \), \( s = t + 1 \), and \( c_1 = \cdots = c_t = c \), where \( c \) and \( t \) are defined by (1) and (3).

**Proof.** The claim is routine to establish, see the calculations in [Nik11, Section 3].

Suppose first that \( \phi_0(\bar{P}_3) = 0 \), where \( \bar{P}_3 \) denotes the complement of the 3-vertex path. Take a sequence \( \{G_n\} \) convergent to \( \phi_0 \). By the Induced Removal Lemma of Alon, Fischer, Krivelevich, and Szegedy [AFKS00], we can change \( o(|V(G_n)|^2) \) adjacencies in \( G_n \) and ensure that \( G_n \) has no \( P_3 \)-subgraph. As it is well-known, a graph without an induced copy of \( P_3 \) is complete multipartite. Here, Theorem 2.1 follows from Lemma 4.4.

So suppose that \( \phi_0(\bar{P}_2) > 0 \). We need a few auxiliary results in this case. For a graph \( F \in \mathcal{F}_\ell^0 \), let \( F^{(1)} \in \mathcal{F}_{\ell+1}^1 \) be the 1-flag obtained by adding a new vertex \( x \) that is connected to all vertices of \( F \) (i.e., taking the join \( F \lor K_1 \)) and labeling \( x \) as 1.

**Claim 4.5** \( \phi_0([\bar{P}_3^{(1)}]_1) > 0 \).

**Proof.** By Claim 4.3 we have that \( \phi_0(K_4) = h^{(4)}(a) \). When we substitute this value into (20) we obtain the tight inequality except for the extra term involving \( K_{1,3} \) (a triangle plus an isolated vertex). We conclude that

\[
\phi_0(\bar{K}_{1,3}) = 0.
\]

(34)

Inequality (17) 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 \( \phi_0([\bar{P}_3 E K_3^E]_E) > 0 \), where \( \bar{P}_3 E \) is the unique \( E \)-flag on \( \bar{P}_3 \). But

\[
[\bar{P}_3 K_3^E]_E = \frac{1}{4} \bar{K}_{1,3} + \frac{1}{3} [\bar{P}_3^{(1)}]_1,
\]

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 (34). Namely, we make up an analog of (20) that is tight on extremal homomorphisms and such that the “overall slackness” involved will cover $G_1$ and $G_2$.

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

$$f^E \overset{\text{def}}{=} \frac{1}{2} P^{E,c}_4 - \frac{1}{2} P^{E,b}_4 - F^E,$$

where $P^{E,c}_4, P^{E,b}_4, F^E \in \mathcal{F}_4$ are shown on Figure 2. Since (17) is tight,

$$\phi_0^E(K_3^E) < \frac{1}{3} h'_t(a) \implies \phi_0^E(\mathcal{P}^E_3) = 0 \text{ a.e.}$$

and, since both $P^{E,b}_4$ and $F^E$ contain $P^E_3$, this implies that

$$\phi_0^E(K_3^E) < \frac{1}{3} h'_t(a) \implies \phi_0^E(f^E) \geq 0 \text{ a.e.} \quad (35)$$

(Recall that $h_t$ is just the restriction of $h$ to the interval $[1 - \frac{1}{t}, 1 - \frac{1}{t+1}]$ as defined by (13).) Thus, by (14), we can multiply the left-hand side of (35) by $f^E$, obtaining a true inequality. If we apply the averaging operator $\lfloor \ldots \rfloor_E$ to this new inequality, we get that

$$\phi_0(\lfloor f^E K_3^E \rfloor_E) \leq \frac{1}{3} h'_t(a) \phi_0(\lfloor f^E \rfloor_E). \quad (36)$$

Next, similarly to [Raz08, (3.4)] but multiplying [Raz08, (3.2)] (i.e. our formula (15) which is equality) by $K^1_3$ rather than by $e$, we obtain

$$\phi_0(3[(K^1_3)^2]_1 - 2h'_t(a)[eK^1_3]_1) = b(3b - 2ah'_t(a)). \quad (37)$$

Subtracting (37) from (36) multiplied by 3, and re-grouping terms,

$$3\phi_0(\lfloor f^E K^E_3 \rfloor_E - \lfloor (K^1_3)^2 \rfloor_1 + h'_t(a)\phi_0(2[eK^1_3]_1 - \lfloor f^E \rfloor_E) \leq b(2ah'_t(a) - 3b). \quad (38)$$
But we have
\[2[εK_{3}^{1}]_{1} - [[F^{E}]_{E} = \frac{4}{3}K_{3} + \frac{2}{3}K_{4} - \frac{1}{3}K_{1,3}\]
(39)
and
\[[F^{E}K_{3}^{E}]_{E} - [(K_{3}^{1})^{2}]_{1} \geq \frac{1}{60}(G_{1} + G_{2}) - \left(\frac{1}{2}K_{4} + \frac{1}{3}hK_{3} + \frac{1}{6}K_{5}\right).\]
(40)
Substituting these relations into (38), and using Claim 4.3, we conclude by (34) that
\[\frac{1}{20}\phi_{0}(G_{1} + G_{2}) \leq b(2ah'(a) - 3b) - h'(a)\left(\frac{4}{3}b + \frac{2}{3}h^{(4)}(a)\right) + \left(\frac{3}{2}h^{(4)}(a) + ab + \frac{1}{2}h^{(5)}(a)\right) = 0.\]
Claim 4.6 is proved.

**Lemma 4.7** Let \(G\) be a graph on \(V = \{x_{1}, x_{2}, x_{3}, y, z\}\) as follows. Let the vertices \(x_{1}, x_{2}, x_{3}\) induce \(P_{3}\) with \(x_{1}x_{2} \in E(G)\). Let \(y\) be adjacent to each \(x_{i}\) and let \(z\) be non-adjacent to at least one \(x_{i}\). If \(yz \notin E(G)\), then \(G\) contains \(K_{1,3}\) 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 \(G - y \cong K_{1,3}\). If \(zx_{1}, zx_{2} \notin E(G)\), then \(G - x_{3} \cong K_{1,3}\). 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.

Let \(σ\) be the 3-vertex type whose graph is \(P_{3}\) with labels 1 and 2 being adjacent. Let \(\{G_{n}\}\) converge to \(ϕ_{0}\) with \(|V(G_{n})| = n\). By Claim 4.5, \(G_{n}\) has \(Ω(n^{4})\) copies of \(F_{0} \in \mathcal{F}_{1}^{0}\), which denotes a triangle with a pendant edge. Let \(F_{1} \in \mathcal{F}_{1}^{1}\) be obtained from \(F_{0}\) by putting label 1 on a vertex of degree 2. Let \(F_{3} \in \mathcal{F}_{1}^{3}\) be the (unique) \(σ\)-flag that can be obtained from \(F_{1}\) by adding labels 2 and 3.

Fix small positive \(ε ≫ δ\). Let \(X = \{x \in V(G_{n}) : p(F_{1}, (G_{n}; x_{1})) > ε\}\). By counting copies of \(F_{0}\) in \(G_{n}\), we conclude that
\[2(ϕ(F_{0}) + o(1))\left(\frac{n}{4}\right) \leq |X|\left(\frac{n - 1}{3}\right) + (n - |X|)ε\left(\frac{n - 1}{3}\right),\]
implying that \(|X| \geq 2εn\). Likewise, by (31), the set \(Y = \{y \in V(G_{n}) : p(e, (G_{n}; x_{1})) < 1 - ε\}\) has size at least \((1 - ε)n\). Thus \(|X ∩ Y| \geq εn\). Similar counting shows that for every \(x_{1} \in X ∩ Y\) there are at least \(δn^{2}\) pairs \((x_{2}, x_{3})\) of vertices with \(p(F_{3}, (G_{n}; x_{1}, x_{2}, x_{3})) \geq δ\).

For \(x_{1}, x_{2}, x_{3}\) as above, 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 \(|V_{1}| \geq δ(n - 3)\). Since \(x_{1} \in Y\), we have \(|V_{2}| \geq ε(n - 1)\) (note that all non-neighbors 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 \(K_{1,3}\) by Lemma 4.7. By (34) 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 \(ε(\overline{G}[V_{1}, V_{2}]) = o(n^{2})\).
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 positive.

Let $i = 1$ or $2$. Each $\phi_i$ is an extremal homomorphism: for example, if there is $\phi_1'$ 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 and the Removal Lemma (or just (5)) imply that $\phi_i(\rho) \leq 1 - \frac{1}{t}$. Thus we can apply the (global) induction and conclude that $\phi_i \in \Phi$.

Let $\phi_0'$ be the limit of complete partite graphs that have the same edge and triangle density as $\phi_i$ and let $\phi_0' = \vee(\phi_1', \phi_2', \alpha_1, \alpha_2)$. Then $\phi_0'$ is also extremal and all the claims that we proved for $\phi_0$ apply to $\phi_0'$. In particular, $\phi_0'(K_{t+2}) = 0$. Thus $\phi_0'$ is the limit of complete partite graphs with $s \leq t + 1$ parts. Assume that $s \geq 3$ for otherwise we are done. By Lemma 4.4, the parts have the correct ratios. We conclude that one of $\phi_i'$, say $\phi_1'$, has equal part sizes and thus $\phi_1' = \phi_1$. By induction, we have $\phi_2 \in \Phi$. The required structure of the join $\phi_0$ routinely follows from Lemma 4.4.

This finishes the proof of Theorem 2.1.

References

[AFKS00] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20:451–476, 2000.

[BCL+08] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219:1801–1851, 2008.

[Bol76] B. Bollobás. On complete subgraphs of different orders. *Math. Proc. Camb. Phil. Soc.*, 79:19–24, 1976.

[DHM+12] S. Das, H. Huang, J. Ma, H. Naves, and B. Sudakov. A problem of Erdős on the minimum number of $k$-cliques. E-print: *arxiv.org*:1203.2723, 2012.

EFR86] P. Erdős, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.*, 2:113–121, 1986.

[Erd55] P. Erdős. Some theorems on graphs. *Riveon Lematematika*, 9:13–17, 1955.

[Erd62] P. Erdős. On a theorem of Rademacher-Turán. *Illinois J. Math.*, 6:122–127, 1962.

[Erd67] P. Erdős. Some recent results on extremal problems in graph theory. Results. In *Theory of Graphs (Internat. Sympos., Rome, 1966)*, pages 117–123 (English); pp. 124–130 (French). Gordon and Breach, New York, 1967.
[Erd69] P. Erdős. On the number of complete subgraphs and circuits contained in graphs. \textit{Časopis Pěst. Mat.}, 94:290–296, 1969.

[ES83] P. Erdős and M. Simonovits. Supersaturated graphs and hypergraphs. \textit{Combinatorica}, 3:181–192, 1983.

[Fis89] D. C. Fisher. Lower bounds on the number of triangles in a graph. \textit{J. Graph Theory}, 13:505–512, 1989.

[FK99] A. Frieze and R. Kannan. Quick approximation to matrices and applications. \textit{Combinatorica}, 19:175–220, 1999.

[Goo59] A. W. Goodman. On sets of acquaintances and strangers at any party. \textit{Amer. Math. Monthly}, 66:778–783, 1959.

[HHK+11] H. Hatami, J. Hladký, D. Král, S. Norine, and A. Razborov. On the number of pentagons in triangle-free graphs. E-print \texttt{arXiv:1102.1634}, 2011.

[HN11] H. Hatami and S. Norine. Undecidability of linear inequalities in graph homomorphism densities. \textit{J. Amer. Math. Soc.}, 24:547–565, 2011.

[KS96] J. Komlós and M. Simonovits. Szemerédi’s regularity lemma and its applications to graph theory. In D. Miklós, V. T. Sós, and T. Szönyi, editors, \textit{Combinatorics, Paul Erdős is Eighty}, volume 2, pages 295–352. Bolyai Math. Soc., 1996.

[LS76] L. Lovász and M. Simonovits. On the number of complete subgraphs of a graph. In \textit{Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975)}, pages 431–441. Congressus Numerantium, No. XV, Winnipeg, Man., 1976. Utilitas Math.

[LS83] L. Lovász and M. Simonovits. On the number of complete subgraphs of a graph. II. In \textit{Studies in pure mathematics}, pages 459–495. Birkhäuser, Basel, 1983.

[LS06] L. Lovász and B. Szegedy. Limits of dense graph sequences. \textit{J. Combin. Theory (B)}, 96:933–957, 2006.

[Man07] W. Mantel. Problem 28. \textit{Winkundige Opgaven}, 10:60–61, 1907.

[MM62] J. W. Moon and L. Moser. On a problem of Turán. \textit{Publ. Math. Inst. Hungar. Acad. Sci.}, 7:283–287, 1962.

[Nik11] V. Nikiforov. The number of cliques in graphs of given order and size. \textit{Trans. Amer. Math. Soc.}, 363:1599–1618, 2011.

[Pik11] O. Pikhurko. The minimum size of 3-graphs without four vertices spanning no or exactly three edges. \textit{Europ. J. Combin.}, 23:1142–1155, 2011.
[Pik12] O. Pikhurko. Minimum number of $k$-cliques in graphs with bounded independence number. E-print: arxiv.org:1203.4393, 2012.

[Raz07] A. Razborov. Flag algebras. J. Symb. Logic, 72:1239–1282, 2007.

[Raz08] A. Razborov. On the minimal density of triangles in graphs. Combin. Probab. Computing, 17:603–618, 2008.

[Rei12] C. Reiher. Minimizing the number of cliques in graphs of given order and edge density. Manuscript, 2012.

[RS78] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. In A. Hajnal and V. Sós, editors, Combinatorics II, pages 939–945. North Holland, Amsterdam, 1978.

[Sim68] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. In Theory of Graphs (Proc. Colloq., Tihany, 1966), pages 279–319. Academic Press, 1968.

[Tur41] P. Turán. On an extremal problem in graph theory (in Hungarian). Mat. Fiz. Lapok, 48:436–452, 1941.

[Whi32] H. Whitney. The coloring of graphs. Annals of Math., 33:688–718, 1932.