Asymptotic Structure for the Clique Density Theorem

Jaehoon Kim*, Hong Liu† Oleg Pikhurko‡ Maryam Sharifzadeh§

Received 17 August 2019; Published 30 December 2020

Abstract: The famous Erdős-Rademacher problem asks for the smallest number of $r$-cliques in a graph with the given number of vertices and edges. Despite decades of active attempts, the asymptotic value of this extremal function for all $r$ was determined only recently, by Reiher [Annals of Mathematics, 184 (2016) 683–707]. Here we describe the asymptotic structure of all almost extremal graphs. This task for $r = 3$ was previously accomplished by Pikhurko and Razborov [Combinatorics, Probability and Computing, 26 (2017) 138–160].

Key words and phrases: graph limits, graphon, clique density theorem, stability, etc.

1 Introduction

Let $K_r$ denote the complete graph on $r$ vertices and let the Turán graph $T_r(n)$ be the complete $r$-partite graph with $n$ vertices and balanced part sizes (that is, every two parts differ in size by at most 1).

It is fair to say that extremal graph theory was born with the following fundamental theorem of Turán [31]: among all graphs on $n$ vertices without $K_r$, the Turán graph $T_{r-1}(n)$ is the unique (up to isomorphism) graph with the maximum number of edges. The case $r = 3$ of this theorem was proved earlier by Mantel [17].

*Supported by the Leverhulme Trust Early Career Fellowship ECF-2018-538, by the POSCO Science Fellowship of POSCO TJ Park Foundation, and by the KAIX Challenge program of KAIST Advanced Institute for Science-X.
†Supported by the UK Research and Innovation Future Leaders Fellowship MR/S016325/1 and the Leverhulme Trust Early Career Fellowship ECF-2016-523.
‡Supported by the Leverhulme Research Project Grant RPG-2018-424.
§Supported by the European Unions Horizon 2020 research and innovation programme under the Marie Curie Individual Fellowship agreement No 752426.
Rademacher (unpublished, see e.g. [6, 7]) showed that, for even \( n \), any \( n \)-vertex graph with \( t_2(n) + 1 \) edges contains at least \( n/2 \) triangles, where \( t_r(n) := e(T_r(n)) \) is the number of edges in the \( r \)-partite Turán graph \( T_r(n) \). In 1955, Erdős [6] asked the more general question to determine \( G_r(n,m) \), the minimum number of copies of \( K_r \) in an \((n,m)\)-graph, that is, a graph with \( n \) vertices and \( m \) edges. This question is now known as the Erdős-Rademacher problem. Early papers on this problem ([6, 7, 18, 20, 22], etc) dealt mainly with the case when \( m \) is slightly larger than \( t_{r-1}(n) \), the threshold when copies of \( K_r \) start to appear. But even this special case turned out to be quite difficult. For example, the conjecture of Erdős [6] that \( G_3(n,t_2(n)+q) \geq q \lfloor n/2 \rfloor \) for \( q < \lfloor n/2 \rfloor \) when \( n \) is large was proved only two decades later by Lovász and Simonovits [13], with a proof of the conjecture also announced by Nikiforov and Khadzhiivanov [21].

Lovász and Simonovits [13, Conjecture 1] made the following bold conjecture. Let \( \mathcal{H} \) consist of all graphs that can be obtained from a complete multipartite graph by adding a triangle-free graph into one of the parts. Let \( H_r(n,m) \) be the minimum number of \( r \)-cliques in an \((n,m)\)-graph from \( \mathcal{H} \). Clearly, \( G_r(n,m) \leq H_r(n,m) \). In this notation, the conjecture of Lovász and Simonovits states that we have equality here, provided \( n \) is sufficiently large:

**Conjecture 1.1 (Lovász and Simonovits [13]).** For every integer \( r \geq 3 \), there is \( n_0 \) such that for all \( n \geq n_0 \) and \( 0 \leq m \leq \binom{n}{2} \), we have that \( G_r(n,m) \geq H_r(n,m) \).

Of course, minimising the number of \( r \)-cliques over \((n,m)\)-graphs from the restricted class \( \mathcal{H} \) is easier than the unrestricted version. The computation of \( H_3(n,m) \) for all \((n,m)\) appears in [11, Proposition 1.5]. Some large ranges of parameters when the conjecture has been proved are when \( m \) is slightly above \( t_r(n) \) by Lovász and Simonovits [14] and when \( r = 3 \) and \( m/\binom{n}{2} \) is bounded away from 1 by Liu, Pikhurko and Staden [11]. Still, Conjecture 1.1 remains open.

Let us turn to the asymptotic version. Namely, given \( \alpha \in [0,1] \) take any integer-valued function \( 0 \leq e(n) \leq \binom{n}{2} \) with \( e(n)/\binom{n}{2} \to \alpha \) as \( n \to \infty \) and define

\[
G_r(\alpha) := \lim_{n \to \infty} \frac{G_r(n,e(n))}{\binom{n}{2}},
\]

\[
h_r(\alpha) := \lim_{n \to \infty} \frac{H_r(n,e(n))}{\binom{n}{2}}.
\]

It is not hard to see from basic principles (see e.g. [24, Lemma 2.2]) that the limits exist and do not depend on the choice of the function \( e(n) \). Thus, the determination of \( g_r(\alpha) \) amounts to estimating the Erdős-Rademacher function within additive error \( o(n^r) \). Clearly, \( h_r(\alpha) \) is an upper bound on \( g_r(\alpha) \).

For each \( \alpha \in [0,1] \), it is not hard to find some sequence of graphs \( (H_{\alpha,n})_{n \in \mathbb{N}} \), that give the value of \( h_r(\alpha) \). If \( \alpha = 1 \), then we can let \( H_{1,n} := K_n \) be the complete graph. Suppose that \( 0 \leq \alpha < 1 \). Let integer \( k \geq 1 \) satisfy \( \alpha \in \left[ 1 - \frac{1}{k}, 1 - \frac{1}{k+1} \right) \). Then fix the (unique) \( c > \frac{1}{k+1} \) so that the complete \((k+1)\)-partite graph \( H_{\alpha,n} := K(V_1,\ldots,V_{k+1}) \) with parts \( V_1,\ldots,V_{k+1} \), where \( |V_i| = \cdots = |V_k| = \lfloor cn \rfloor \), has edge density \( (\alpha + o(1)) \binom{n}{2} \) as \( n \to \infty \). It is easy to show, see (2.8), that \( c \leq \frac{1}{k} \). Thus \( |V_{k+1}| = n - k \lfloor cn \rfloor \geq 0 \). (In fact, our choice to round \( cn \) down was rather arbitrary, just to make the graph \( H_{\alpha,n} \) well-defined for each \( \alpha \) and \( n \).) It is routine to show (see e.g. [19, Theorem 1.3] for a derivation) that these ratios give the value of \( h_r(\alpha) \), that is,

\[
h_r(\alpha) = r! \left( \binom{k}{r} c^r + \binom{k}{r-1} c^{r-1}(1-kc) \right). \tag{1.1}
\]
The function $h_r$ stays 0 when $\alpha \leq 1 - \frac{1}{r-1}$ (when $k \leq r-2$). Also, as Lemma 2.7 implies, $h_r$ consists of countably many concave “scallops” with cusps at $\alpha = 1 - \frac{1}{m}$ for integer $m \geq r - 1$.

For a while, the best known lower bound on the limit function $h_r$, by Bollobás [1], was the piecewise linear function which coincides with $h_r$ on the cusp points. Fisher [9] showed that $g_3(\lambda) = h_3(\lambda)$ for all $1/2 \leq \lambda \leq 2/3$, that is, he determined $g_3$ in the first scallop. Razborov used his newly developed theory of flag algebras first to give a different proof of Fisher’s result in [26] and then to determine the whole function $g_3$ in [27]. The function $g_4$ was determined by Nikiforov [19] and the function $g_r$ for any $r \geq 5$ was determined by Reiher [28] (with these two papers also giving new proofs for the previously solved cases of $r$).

1.1 Main result

We are interested in the asymptotic structure of (almost) extremal graphs for the Erdős-Rademacher $K_r$-minimisation problem, that is, a description up to $o(n^2)$ edges of every $(n,m)$-graph with $G_r(n,m) + o(n^2)$ copies of $K_r$ as $n \to \infty$. Of course, such a result tells us more about the problem than just the value of $g_r$.

Asymptotic structure results are often very useful for proving enumerative and probabilistic versions of the corresponding extremal problem. For example, the more general problem of understanding the structure of graphons with any given edge and $r$-clique densities appears in the study of exponential random graphs (see Chatterjee and Diaconis [4] and its follow-up papers), phases in large graphs (see the survey by Radin [25]), and large deviation inequalities for the clique density (see the survey by Chatterjee [3]). Last but not least, asymptotic structure results often greatly help, as a first step, in obtaining the exact structure of extremal graphs via the so-called stability approach pioneered by Simonovits [30]. Here, knowing extremal $n$-vertex graphs within $o(n^2)$ edges greatly helps in the ultimate aim of ruling out even a single “wrong” adjacency. In fact, almost all cases when the Erdős-Rademacher problem was solved exactly were established via the stability approach.

In order to state the main result of this paper, we have to define further graph families. For $\alpha = 1$, we let $\mathcal{H}_{1,n} := \{K_n\}$. For $\alpha \in [0,1)$, by using the notation defined before (1.1), let $\mathcal{H}_{\alpha,n}$ consist of all graphs that are obtained from the complete $k$-partite graph on parts $V_1, \ldots, V_{k-1}$ and $U := V_k \cup V_{k+1}$ by adding a triangle-free graph on $U$ with $|V_k| \cdot |V_{k+1}|$ edges. (In particular, $\mathcal{H}_{0,n} := \{K_n\}$ consists of the empty graph only.) Clearly, for any $r \geq 3$, the number of $r$-cliques in the obtained graph does not depend on the choice of the graph added on $U$. Also, $\mathcal{H}_{\alpha,n} \supseteq H_{\alpha,n}$ is always non-empty (but typically has many non-isomorphic graphs). Finally, for $r \geq 3$, let $\mathcal{H}_{r,n}$ be the union of $\mathcal{H}_{\alpha,n}$ over all $\alpha \in [0,1]$ together with the family of all $K_r$-free $n$-vertex graphs.

Pikhurko and Razborov [23] proved that every almost extremal $(n,m)$-graph $G$ is $o(n^2)$ close in edit distance to some graph in $\mathcal{H}_{3,n}$. Our main result is to extend this structural result to all $r \geq 4$:

**Theorem 1.2.** For every real $\varepsilon > 0$ and integer $r \geq 4$, there are $\delta > 0$ and $n_0$ such that every graph $G$ with $n \geq n_0$ vertices and at most $(g_r(\varepsilon) + \delta)(\binom{n}{r})$ $r$-cliques, where $\alpha := e(G)/\binom{n}{2}$, can be made isomorphic to some graph in $\mathcal{H}_{r,n}$ by changing at most $\varepsilon n^2$ adjacencies.

Note that $\mathcal{H}_{3,n} \subseteq \mathcal{H}_{r,n}$. Also, all graphs in $\mathcal{H}_{r,n} \setminus \mathcal{H}_{3,n}$ are $K_r$-free but may contain triangles; these are “trivial” minimisers for the $g_r$-problem that need not be minimisers for the $g_3$-problem. Thus, apart from these graphs, the $g_3$ and $g_r$ extremal problems have the same set of approximate minimisers and we

**DISCRETE ANALYSIS, 2020:19, 26pp.**
exploit this in our proof as follows. In brief, we take any almost $G_r(n,m)$-extremal graph $G$. Suppose that $G$ has strictly more than $G_3(n,m) + o(n^3)$ triangles for otherwise $G$ is $o(n^2)$-close in edit distance to $\mathcal{H}_{3,n}$ by the result in [23] and we are done since $\mathcal{H}_{3,n} \subseteq \mathcal{H}_{r,n}$. If $\alpha := m/(\binom{n}{2})$ is $1 - \frac{1}{r} + o(1)$ for some integer $k \geq r$ (that is, the edge density of $G$ is close to that of some Turán graph $T_k(n)$), then we use the result of Lovász and Simonovits [14] that $G$ has to be $o(n^2)$-close in edit distance to $T_k(n)$, giving the desired conclusion. Thus we can assume that the edge density is strictly inside one of the scallops. Lemma 2.7 shows that the function $h_r$ is differentiable for such $\alpha$. This allows us to derive various properties of $G$ via variational principles. The property that we will need is that, for a typical vertex $x$ of $G$, there is an asymptotic linear relation between the degree of $x$ and the number of $r$-cliques containing $x$. This relation comes from the Lagrange multiplier method. Since we know the extremal function $g_r$, we can determine all Lagrange multipliers and write an explicit relation. Since the graph $G$ is “heavy” on triangles, we can find a typical vertex $x$ that is “heavy” in terms of triangles containing it. When we restrict ourselves to the graph $G'$ induced by the set of neighbours of $x$, then the counts of triangles and $r$-cliques in $G$ containing $x$ correspond to the counts of respectively edges and $(r-1)$-cliques in $G'$. Some calculations show that $G'$ is too “heavy” on $K_2$ when compared to the number of $(r-1)$-cliques, contradicting the asymptotic result for $r = 1$ and finishing the proof. Thus, the main results on which our proof of Theorem 1.2 for a given $r \geq 4$ crucially relies are the values of $g_r$ and $g_{r-1}$ as well as the asymptotic structure for $r = 3$.

We found it more convenient to present our proof in terms of graphons that are analytic objects representing subgraph densities in large dense graphs. This reduces the number of parameters in various statements. For example, the statement that some “natural” property fails for $o(n)$ vertices corresponds in the limit to the statement that the set of failures has measure 0. Also, the variational principles are easier to state and derive using the graphon language, in particular it is much cleaner to define the limit versions of the families $\mathcal{H}_{r,n}$, namely the families $\mathcal{H}_r$ defined in Section 1.2. Some downside of this is that we have to use various non-trivial (but standard) facts of measure theory. We rectify this by giving discrete analogues of some analytic constructions and properties that we use. Also, we believe that graphons, as a tool in extremal graph theory, are by now standard and widely known.

**Organisation.** The rest of the paper is organised as follows. We will rephrase our main result in terms of the structure of extremal graphons in the next subsection, Section 1.2. Further properties of graphons and of the family of extremal graphons are discussed in Sections 2.2 and 2.3. This is preceded by Section 2.1 that contains some notions and results of measure theory that we will need. In Section 2.4, we derive Theorem 1.2 from our result on graphons. Then in Section 3, we present the proof of the graphon version of our main result.

### 1.2 Graphons with minimum clique density

For an introduction to graphons, we refer the reader to the excellent book by Lovász [12].

For the purposes of this paper, it is convenient to define a graphon as a pair $(W, \mu)$, where $W : [0,1] \times [0,1] \to [0,1]$ is a symmetric Borel function and $\mu$ is a non-atomic probability measure on Borel subsets of $[0,1]$. By small abuse of notation, we may call just the function $W$ a graphon (if the measure $\mu$ is understood). Each graph $G = (V, E)$ with $V = \{v_1, \ldots, v_n\}$ corresponds naturally to a graphon $(W_G, \mu)$ with $\mu$ being the Lebesgue measure on $[0,1]$ and $W_G$ being the adjacency function of $G$ which assumes value 1 on $\left[\frac{s-1}{n}, \frac{s}{n}\right] \times \left[\frac{t-1}{n}, \frac{t}{n}\right]$ for each $\{v_s, v_t\} \in E(G)$ and 0 otherwise.
For a graph $F$ on $[r]$, define its homomorphism density in a graphon $W$ by
\[ t(F, W) := \int_{[0,1]^r} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{i=1}^r d\mu(x_i). \]

In particular, $t(K_2, W)$ is called the (edge)-density of the graphon $W$. If $W = W_G$, then we get the homomorphism density $t(F, G) := t(F, W_G)$ of $F$ in $G$, which is the probability that a uniformly at random chosen function $f : V(F) \to V(G)$ maps every edge of $F$ to an edge of $G$.

A sequence of graphons $W_n$ converges to $W$ if for every graph $F$ we have
\[ \lim_{n \to \infty} t(F, W_n) = t(F, W). \]

In the special case when $W_n = W_{G_n}$, we get the convergence of graphs $G_n$ to $W$. Let us call two graphons $U$ and $W$ weakly isomorphic if $t(F, U) = t(F, W)$ for every graph $F$. Theorem 13.10 in [12] gives several equivalent definitions of weak isomorphism. Let $\mathcal{W}$ denote the equivalence class of a graphon $W$ up to weak isomorphism and let
\[ \mathcal{W} := \{ [W] : \text{graphon } W \}. \]

If we fix some enumeration $\mathcal{F} = \{ F_1, F_2, \ldots \}$ of all graphs up to isomorphism, then one can identify each $[W] \in \mathcal{W}$ with the sequence $(t(F, W))_{F \in \mathcal{F}} \in [0,1]^\mathcal{F}$ and the above convergence is the one corresponding to the product topology on $[0,1]^\mathcal{F}$. Since the product of compact sets is compact, the closed subspace $\mathcal{W} \subset [0,1]^\mathcal{F}$ is compact. As $\mathcal{F}$ is countable, every infinite sequence of graphons/graphs has a convergence subsequence. Also, for each $F$, the function $W \mapsto t(F, W)$ is continuous (as it is just the projection on the $F$-th coordinate). Another key property of graphons is that finite graphs are dense in $\mathcal{W}$.

Let $B \subseteq [0,1]$ be a Borel subset of $[0,1]$ with $\mu(B) > 0$. We define the graphon $W|B$ induced by $B$ to be the graphon $(W, \mu')$ with $\mu'(A) := \mu(A \cap B) / \mu(B)$ for Borel $A \subseteq [0,1]$. Since the new measure $\mu'$ is 0 on $[0,1] \setminus B$, this effectively restricts everything to $B$ (while the scaling ensures that $\mu'$ is a probability measure).

**Definition 1.3** (Graphon family $H_r$). Let $r \geq 3$. Define $\mathcal{F} := \{ [W] \in \mathcal{W} : t(K_r, W) = 0 \}$ and, for each integer $k \geq r - 1$, let $\mathcal{G}_k$ be the set of weak isomorphism classes of graphons $(W, \mu)$ satisfying the following: there exist a real $c \in (\frac{1}{k+1}, \frac{1}{k})$ and a Borel partition $[0,1] = \Omega_1 \cup \cdots \cup \Omega_k$ with $\mu(\Omega_i) = c$ for each $i \in [k-1]$ such that

1. $t(K_2, W[\Omega_k]) = 2c(b-c)/b^2$, where $b := 1 - (k-1)c$,
2. $t(K_3, W[\Omega_k]) = 0$,
3. For all $x_1 \in \Omega_{k_1}$ and $x_2 \in \Omega_{k_2}$, we have $W(x_1, x_2) = \begin{cases} 0, & \text{if } k_1 = k_2 \leq k - 1, \\ 1, & \text{if } k_1, k_2 \in [k] \text{ satisfy } k_1 \neq k_2. \end{cases}$

Further let
\[ H_r := \{ [1_{[0,1]^2}] \} \cup \mathcal{F} \cup \bigcup_{k=r-1}^\infty \mathcal{G}_k. \]

**Discrete Analysis**, 2020:19, 26pp. 5
This is a direct analogue of the graph family $\mathcal{H}_{r,n}$: for example, to construct a graphon in $\mathcal{G}_k$ we take a “complete partite” graphon with parts $\Omega_1, \ldots, \Omega_k$ and add a triangle-free graphon into the last part. In fact, by Lemma 2.12, $\mathcal{H}_r$ is precisely the set of possible limits of increasing graph sequences $(G_n)_{n=1}^{\infty}$ with each $G_n \in \mathcal{H}_{r,n}$. This (or an easy direct calculation) gives that $t(K_r, W) = h_r(t(K_2, W))$ for each $W \in \mathcal{H}_r$, where $h_r$ is defined in (1.1). Also, note that, by definition, $\mathcal{H}_3 \subseteq \mathcal{H}_4 \subseteq \ldots \subseteq \mathcal{W}$.

The result of Reiher [28] that determines the function $g_r$ can be equivalently rephrased in the language of graphons as follows.

**Theorem 1.4 ([28]).** For each integer $r \geq 3$ and a graphon $W$, it holds that $t(K_r, W) \geq h_r(t(K_2, W))$.

We say that a graphon $W$ is $K_r$-extremal if $t(K_r, W) = g_r(t(K_2, W))$. In other words, it is $K_r$-extremal if it has the minimum $K_r$-density among all graphons with the same edge density. In fact, the asymptotic structure result for triangles by Pikhurko and Razborov was first derived via a statement about graph limits (see [23, Theorem 2.1]):

**Theorem 1.5 ([23]).** A graphon $W$ is $K_3$-extremal if and only if $[W] \in \mathcal{H}_3$.

Our main result in terms of graphons is as follows.

**Theorem 1.6.** For each $r \geq 3$, a graphon $W$ is $K_r$-extremal if and only if $[W] \in \mathcal{H}_r$.

This completely characterises all graphons achieving the equality in Theorem 1.4 (and implies Theorem 1.2, see Section 2.4).

## 2 Preliminaries

Let $\mathbb{N} := \{1, 2, \ldots\}$ be the set of natural numbers. For $t \in \mathbb{N}$, let $I_t := [1 - \frac{1}{t}, 1 - \frac{1}{t+1}]$; these intervals partition $[0, 1)$. Also, we denote $[t] := \{1, \ldots, t\}$ and, for a set $X$, let $\binom{X}{t}$ consist of all $t$-subsets of $X$. For $k, t \in \mathbb{N}$, the $t$-th falling power of $k$ is $k^{(t)} := k(k-1)\ldots(k-t+1)$; note that $k^{(t)} = 0$ if $t > k$. The indicator function $1_Y$ of a set $Y$ assumes value 1 on $Y$ and is 0 otherwise. An unordered pair $\{x, y\}$ may be abbreviated to $xy$.

A graph is always a finite graph with non-empty vertex set. For a graph $G = (V, E)$ and bijection $\phi$ from $V$ to some set $X$, we denote by $\phi(G)$ the graph on $X$ with edge set $\phi(E) := \{\{\phi(x), \phi(y)\} : x, y \in E(G)\}$. The edit distance $|G \triangle H|$ between two graphs $G$ and $H$ of the same order is the minimum of $|E(G) \triangle E(\phi(H))|$ over all bijections $\phi : V(H) \to V(G)$; in other words, this is the minimum number of edge edits needed to make $G$ and $H$ isomorphic.

For $a, b, c \in \mathbb{R}$ we write $a = b \pm c$ if $b - c \leq a \leq b + c$. The constants in the hierarchies used to state our results have to be chosen from right to left. More precisely, if we claim that a result holds whenever e.g. $c \ll b \ll a_1, \ldots, a_k$, then this means that there are coordinate-wise non-decreasing functions $f : [0, 1] \to [0, 1]$ and $g : (0, 1]^k \to (0, 1]$ such that the claimed result holds whenever $0 < c < f(b)$ and $0 < b < g(a_1, \ldots, a_k)$.
2.1 Some notions and results of measure theory

Let us recall some basic notions that apply when \( \mathcal{A} \) is a \( \sigma \)-algebra on a set \( X \) and \( \nu \) is a measure on \( (X, \mathcal{A}) \). A function \( f : X \to \mathbb{R} \) is called \( \mathcal{A} \)-measurable if the preimage of any open (equivalently, Borel) subset of \( \mathbb{R} \) is in \( \mathcal{A} \). This class of functions is closed under arithmetic operations, pointwise limits, etc., see e.g. [5, Section 2.1]. A set \( Y \subseteq X \) is called \( (\nu-) \) null if there is \( Z \in \mathcal{A} \) with \( Z \supseteq Y \) and \( \nu(Z) = 0 \). We say that a property holds \( (\nu-) \) a.e. if the set of \( x \) where it fails is \( \nu \)-null. The \( \nu \)-completion \( \mathcal{A}_\nu \) of \( \mathcal{A} \) consists of those \( A \subseteq X \) for which there exist \( B, C \in \mathcal{A} \) with \( B \subseteq A \subseteq C \) and \( \nu(B) = \nu(C) \); equivalently, \( \mathcal{A}_\nu \) is the \( \sigma \)-algebra generated by the union of \( \mathcal{A} \) and all \( \nu \)-null sets.

For \( k \in \mathbb{N} \), let \( \mathcal{B}([0,1]^k) \) consist of all Borel subsets of \([0,1]^k\) (i.e., it is the \( \sigma \)-algebra generated by open subsets of \([0,1]^k\)). It is easy to show (see e.g. [5, Example 5.1.1]) that \( \mathcal{B}([0,1]^k) \) is equal to the product of \( k \) copies of the Borel \( \sigma \)-algebra on \([0,1]\). When \( k \) is clear, we abbreviate \( \mathcal{B}([0,1]^k) \) to \( \mathcal{B} \).

Let \( \mu \) be a probability measure on \(([0,1], \mathcal{B}) \). By \( \mu^k \), we denote the measure on \(([0,1]^k, \mathcal{B}) \) which is the product of \( k \) copies of \( \mu \). We call the sets in the \( \mu^k \)-completion of \( \mathcal{B}([0,1]^k) \) measurable and, when \( k \) is understood, denote this \( \sigma \)-algebra by \( \mathcal{B}_\mu \). We call a function \( f : [0,1]^k \to \mathbb{R} \) Borel (resp. measurable) if it is \( \mathcal{B} \)-measurable (resp. \( \mathcal{B}_\mu \)-measurable).

Let us state some results that will be useful for us. The first one is an easy consequence of the countable additivity of a measure, see e.g. [5, Proposition 1.2.5].

**Lemma 2.1** (Continuity of measure). For every measure space \((X, \mathcal{A}, \nu)\) and every nested sequence \( X_0 \subseteq X_1 \subseteq X_2 \subseteq \ldots \) of sets in \( \mathcal{A} \), the measure of their union \( \bigcup_{n \in \mathbb{N}} X_n \) is equal to \( \lim_{n \to \infty} \nu(X_n) \).

The following result will allow us to work with just Borel sets and functions.

**Lemma 2.2.** For every measure space \((X, \mathcal{A}, \nu)\) and an \( \mathcal{A}_\nu \)-measurable function \( f : X \to [0,1] \), there is an \( \mathcal{A} \)-measurable function \( f' : X \to [0,1] \) such that \( f' = f \) \( \nu \)-a.e.

**Proof.** For \( x \in X \), write \( f(x) \in [0,1] \) in binary, \( f(x) = \sum_{i=0}^{\infty} b_i(x) 2^{-i} \) where each \( b_i(x) \in \{0,1\} \) and, for definiteness, we require that infinitely many of \( b_i(x) \) are 0 (thus we do not allow expansions where all digits are eventually 1). Note that each \( B_i := \{ x \in X : b_i(x) = 1 \} \) is \( \mathcal{A}_\nu \)-measurable because it is the preimage under the measurable function \( f \) of the set consisting of \( r \in [0,1] \) such that \( i \)-th binary digit of \( r \) is 1, which is Borel as a union of some intervals.

By the definition of \( \mathcal{A}_\nu \), for each \( i \in \mathbb{N} \) there are \( A_i, N_i \in \mathcal{A} \) such that \( \nu(N_i) = 0 \) and \( A_i \cap B_i \subseteq N_i \). Let \( g := \sum_{i=0}^{\infty} 2^{-i} 1_{A_i} \). Then \( g \) is \( \mathcal{A} \)-measurable as the countable convergent sum of \( \mathcal{A} \)-measurable functions. Also, the set where \( f \) and \( g \) differ is a subset of \( N := \bigcup_{i=0}^{\infty} N_i \), which has measure 0. Some points in \( A_0 \cup N_0 \) may have \( g \)-value greater than 1, so let \( f'(x) := \min\{g(x), 1\} \). The set where the \( \mathcal{A} \)-measurable function \( f' \) differs from \( f \) is still a subset of the null set \( N \), as required.

The following result will be frequently used (allowing us, in particular, to change the order of integration), so we state it fully. For a proof, see e.g. [5, Proposition 5.2.1].

**Theorem 2.3** (Tonelli’s theorem). Let \((X, \mathcal{A}, \mu)\) and \((Y, \mathcal{C}, \nu)\) be \( \sigma \)-finite measure spaces and let \( f : X \times Y \to [0, \infty] \) be \( A \times C \)-measurable. Then

1. for every \( x \in X \) the function \( f(x, \cdot) : Y \to [0, \infty], y \mapsto f(x,y) \), is \( \mathcal{C} \)-measurable;
2. the function $x \mapsto \int_{Y} f(x, y) \, dv(y)$ is $\mathcal{A}$-measurable;

3. we have $\int_{X} (\int_{Y} f(x, y) \, dv(y)) \, d\mu(x) = \int_{X \times Y} f(x, y) \, d(\mu \times v)(x, y)$. 

Also, we will need the following result.

**Theorem 2.4** (Sierpiński’s theorem). If $(X, \mathcal{A}, v)$ is a non-atomic measure space with $v(X) < \infty$, then for every $\rho \in [0, v(X)]$ there is $Y \in \mathcal{A}$ with $v(Y) = \rho$.

**Proof.** We will use only the special case $(X, \mathcal{A}) = ([0, 1], \mathcal{B})$ of the theorem, whose proof is very simple. Namely, the function $x \mapsto \mu([0, x))$ for $x \in [0, 1]$ is continuous by Lemma 2.1 (and by the non-atomicity of $\mu$); now, the Intermediate Value Theorem gives the required. 

### 2.2 Further results on graphons

Usually, a graphon $W$ is defined as a symmetric measurable function $[0, 1]^2 \to [0, 1]$ with respect to the Lebesgue measure $\mu$ on $[0, 1]$. Given such $W$, we can, by Lemma 2.2, modify it on a $\mu^2$-null set to obtain a Borel function $U' : [0, 1]^2 \to [0, 1]$. Then letting $U(x, y) := \frac{1}{2} (U'(x, y) + U'(y, x))$ for $(x, y) \in [0, 1]^2$, we obtain a Borel symmetric function which is still a.e. equal to $W$. Thus $[U] = [W]$ and our requirement that graphons are represented by Borel functions is not a restriction. Also, on the other hand, we did not enlarge $W$ by allowing non-uniform measures on $[0, 1]$ as our definition is a special case of the general form of graphons, as defined in [12, Section 13.1]. The motivation for our definition is that our proof requires changing the measure a few times while the assumption that the function $W$ is Borel ensures that all sets and functions that we will encounter are everywhere defined and Borel.

We could have also applied the so-called **purification** of the graphon introduced by Lovász and Szegedy [16] (see also [12, Section 13.3]) which would eliminate a few (simple) applications of the continuity of measure in our proof. However, we decided against using this (non-trivial) result as this could obscure the simplicity of this step.

One consequence of Tonelli’s theorem (Theorem 2.3) and the identity $\mathcal{B}([0, 1]^2) = \mathcal{B}([0, 1]) \times \mathcal{B}([0, 1])$ (and our requirement that the function $W$ is Borel) is that for every $x \in [0, 1]$ the section $W(x, \cdot) : [0, 1] \to [0, 1]$, $y \mapsto W(x, y)$, is a Borel function.

For a graph $F$ on $[r]$ with $1, \ldots, k$ designated as roots, define its **rooted homomorphism density** in $W$ by

$$t_{x_1, \ldots, x_k}(F, W) := \int_{[0, 1]^{r-k}} \prod_{i,j \in E(F)} W(x_i, x_j) \prod_{i=k+1}^{r} d\mu(x_i), \quad \text{for } (x_1, \ldots, x_k) \in [0, 1]^k.$$ 

By Tonelli’s theorem, this is an (everywhere defined) Borel function $[0, 1]^k \to [0, 1]$.

Note that if $F'$ is obtained from a graph $F$ by rooting it on a fixed vertex, then

$$\int_{[0, 1]} t_{x}(F', W) \, d\mu(x) = t(F, W).$$

If $F$ has two roots 1 and 2 that are adjacent and $F^-$ is obtained from $F$ by removing the edge $\{1, 2\}$, then $t_{x_1, x_2}(F, W) = W(x_1, x_2) t_{x_1, x_2}(F^-, W).$ 

**DISCRETE ANALYSIS, 2020:19, 26pp.**
When the graphon is understood, we may abbreviate \( B \) for every \( x \in \{ 0, 1 \} \) in \( W \) by

\[
d_W(x) := t_x(K_2, W) = \int_{[0,1]} W(x,y) \, d\mu(y).
\]

When the graphon is understood, we may abbreviate \( d_W(x) \) to \( d(x) \). By Tonelli’s theorem, \( d(x) \) is defined for every \( x \in [0,1] \) and

\[
\int_{[0,1]} d_W(x) \, d\mu(x) = t(K_2, W). \tag{2.1}
\]

**Definition 2.5.** For a graphon \((W, \mu)\) and \( x \in [0,1] \) with \( d(x) \neq 0 \), define the neighbourhood \( N_W(x) = N(x) \) of \( x \) in \( W \) as the graphon \((W, \mu')\), where

\[
d\mu'(y) := \frac{W(x,y)}{d(x)} \, d\mu(y),
\]

that is, \( \mu'(A) := \int_A \frac{W(x,y)}{d(x)} \, d\mu(y) \) for Borel \( A \subseteq [0,1] \).

Note that, in the above definition, the function \( W \) remains the same and only the measure is changed. With this definition, we have that for every \( r \geq 2 \)

\[
t(K_r, N_W(x)) = \int_{[0,1]^r} \prod_{1 \leq i < j \leq r} W(x_i, x_j) \prod_{i=1}^r d\mu'(x_i)
\]

\[
= \int_{[0,1]^r} \prod_{1 \leq i < j \leq r} W(x_i, x_j) \prod_{i=1}^r W(x_i) \prod_{i=1}^r d\mu(x_i)
\]

\[
= t_x(K_{r+1}, W) \frac{d_W(x)}{(d_W(x))^r}. \tag{2.2}
\]

For \( r, \ell, m \in \mathbb{N}, B \subseteq [0,1]^m \) and \( y = (y_1, y_2) \in [0,1]^2 \), define

\[
B^{r,\ell} := \left\{ x \in [0,1]^r : \left| \left\{ S \in \binom{[r]}{m} : (x_i)_{i \in S} \in B \right\} \right| = \ell \right\},
\]

\[
B^{r,\ell}_+ := \bigcup_{i \geq \ell} B^{r,i},
\]

\[
B^{r,\ell}(y) := \left\{ (y_3, \ldots, y_r) \in [0,1]^{r-2} : (y_1, \ldots, y_r) \in B^{r,\ell} \right\}, \text{ and}
\]

\[
B^{r,\ell}_+(y) := \bigcup_{i \geq \ell} B^{r,i}(y). \tag{2.3}
\]

For example, \( B^{r,\ell} \) consists of those \( r \)-tuples from \([0,1] \) such that the number of \( m \)-subtuples that belong to \( B \subseteq [0,1]^m \) is exactly \( \ell \). We will later need the property that \( B^{r,2}_+ \) is much smaller in measure than

**Discrete Analysis**, 2020:19, 26pp.
This is not true in general: for instance, consider $U \subseteq [0,1]$ with $\mu(U) = \eta \ll 1/r$ and $B = \{(x,y) : x,y \in U \neq \emptyset\}$ when we have $\mu'(B^{r^2}) = \Omega(\mu^2(B))$. However, the following lemma gives the desired property provided that we can pass to a subset of $B$ first. We call a set $B \subseteq [0,1]^2$ symmetric if $(x,y) \in B$ implies that $(y,x) \in B$.

**Lemma 2.6.** Let $\mu$ be a non-atomic measure on $([0,1],\mathcal{B})$, $\eta^{-1} \in \mathbb{N}$ and $B \subseteq [0,1]^2$ be a symmetric Borel set. Then there exists a symmetric Borel subset $C \subseteq B$ satisfying the following for all $r \geq 3$:

(C1) $\eta^2 \mu^2(B) \leq \mu^2(C) \leq \mu^2(B)$;

(C2) For each $x \in [0,1]^2$, we have $\mu'^{-2}(C^r_+(x)) \leq 2r\eta$;

(C3) $\mu'(C^r_+) \leq r^3 \eta \mu^2(C)$.

**Proof.** Let $t := \eta^{-1}$. Using Theorem 2.4, we can partition $[0,1]$ into Borel sets $I_1, \ldots, I_t$ with $\mu(I_i) = \eta$ for each $i \in [t]$. For all $i, j \in [t]$, let $I_{i,j} := I_i \times I_j$. As $\mu^2(B) = \sum_{i,j \in [t]} \mu^2(B \cap I_{i,j})$, there exists $(i_0, j_0) \in [t]^2$ such that $\mu^2(B \cap I_{i_0,j_0}) \geq \eta^2 \mu^2(B)$. Let

$$C := (B \cap I_{i_0,j_0}) \cup (B \cap I_{j_0,i_0}).$$

Clearly, $C$ is a symmetric Borel set that satisfies (C1).

Given $x = (x_1, x_2) \in [0,1]^2$, we consider the following random experiment. We choose $x_3, \ldots, x_r \in [0,1]$ independently at random with respect to the probability measure $\mu$. Let $E$ be the event that $|\{ij \in \binom{[2]}{2} : (x_i, x_j) \in C\}| \geq 2$. Note that by the definition of $C$, if the event $E$ happens then at least one of $x_3, \ldots, x_r$ lies inside $I_{i_0} \cup I_{j_0}$. Thus the probability of $E$ satisfies

$$\mathbb{P}[E] \leq \sum_{i=3}^r \mathbb{P}[x_i \in I_{i_0} \cup I_{j_0}] \leq 2r\eta.$$

This implies (C2).

Finally, by the symmetry between the variables $x_i$, (C2) implies that

$$\mu'(C^r_+) \leq \binom{r}{2} \int_{x \in C} \mu'^{-2}(C^r_+(x)) \prod_{i \in [2]} \mathrm{d}\mu(x_i) \leq \binom{r}{2} \mu^2(C) \cdot 2r\eta \leq r^3 \eta \mu^2(C).$$

Thus we have (C3). \hfill \Box

**2.3 Properties of $\mathcal{H}_r$ and $h_r$**

Everywhere in this section, $r \geq 3$ is fixed. In order to deal with the graphon family $\mathcal{H}_r$, it is convenient to define some related parameters.

For $t, \ell \in \mathbb{N}$ with $\ell \geq 2$ and $\gamma \in \mathbb{R}$, define

$$\kappa_{t,\ell}(\gamma) := \ell! \left( \binom{t}{\ell} \gamma^\ell + \binom{t}{\ell-1} \gamma^{\ell-1} (1-t\gamma) \right).$$

**Discrete Analysis, 2020:19, 26pp.**
Asymptotic Structure for the Clique Density Theorem

\[ = t^{(\ell-1)}\gamma^{-1}(\ell - (\ell - 1)(t+1)\gamma). \quad (2.4) \]

Note that if \( 0 \leq \gamma \leq \frac{1}{t} \) then \( \kappa_{\ell,t}(\gamma) \) is the asymptotic density of \( \ell \)-cliques in a complete partite graph with \( t \) parts of size \( \gamma n \) and one part of size \( (1-t\gamma)n \) as \( n \to \infty \). Next, for \( x \leq 1 - \frac{1}{\ell+t} \), let

\[ \gamma(x) := \frac{1}{t+1} + \frac{\sqrt{t(t-(t+1)x)}}{t(t+1)}. \quad (2.5) \]

This formula comes from taking \( \gamma(x) \) to be the larger root \( \gamma \) of the quadratic equation \( x = \kappa_{\ell,t}(\gamma) \). Further, again for \( x \leq 1 - \frac{1}{\ell+t} \), define

\[ p_{r,t}(x) := \kappa_{r,t}(\gamma(x)) = \frac{t^{(r-1)}}{r'((t+1)^{r-1}} \left( t + \sqrt{t(t-(t+1)x)} \right)^{r-1} \left( t - (r-1)\sqrt{t(t-(t+1)x)} \right). \quad (2.6) \]

With this preparation we are ready to define the two main parameters, \( k \) and \( c \), associated with the edge density \( \alpha \in [0,1] \), namely

\[ k = k(\alpha) \in \mathbb{N} \quad \text{such that} \quad \alpha \in I_k \] (that is, \( 1 - \frac{1}{k} \leq \alpha < 1 - \frac{1}{k+1} \)),

\[ c = c(\alpha) := \gamma(\alpha). \quad (2.7) \]

Note that \( c = c(\alpha) \) in (2.7) is the same as in (1.1) and Definition 1.3; also, (2.5) gives an explicit formula for \( c \).

In other words, the function \( c : [0,1) \to [0,1) \) is obtained by taking \( \gamma \) on the interval \( I_t \) for \( t \in \mathbb{N} \). (Recall that these intervals partition \( [0,1) \) is the same as in (1.1) and Definition 1.3; also, (2.5) gives an explicit formula for \( c \).

Also, easy calculations based on e.g. (2.5) show that for all \( \alpha \in [0,1) \) we have, with \( c = c(\alpha) \) and \( k = k(\alpha) \), that

\[ \frac{1}{k+1} < c \leq \frac{1}{k}. \quad (2.8) \]

Furthermore, \( h_r(\alpha) = \kappa_{rk}(c) = p_{rk}(\alpha) \), where the function \( h_r(\alpha) \) was defined in (1.1). In fact, as we will show in Lemma 2.11, \( h_r(\alpha) = \max \{ p_{rk}(\alpha) : t \geq k \} \), that is, \( h_r \) is the maximum of those functions \( p_{rk} \) that are defined at a given point, with \( p_{rk} \) being a largest one on \( I_t \).

The following lemma computes the first two derivatives of \( p_{r,t} \) (and thus of \( h_r \) in all interior points of each \( I_t \), where we write these derivatives in terms of \( \gamma \) for convenience. Note that \( h_r \) is not differentiable at points \( 1 - \frac{1}{t} \) for integers \( t \geq r-1 \): the left and right derivatives of \( h_r \) exist at these points but are different.

**Lemma 2.7.** For each \( t \in \mathbb{N} \) and \( x \leq 1 - \frac{1}{\ell+t} \), we have that

\[ p'_{r,t}(x) = \left( \begin{array}{c} r \\ 2 \end{array} \right) (t-1)^{r-2} \gamma(x)^{r-2} \quad \text{and} \quad p''_{r,t}(x) = \frac{3}{2} (t-1)^{(r-2)} \gamma(x)^{r-3} \]

\[ 2r(t-1)(t+1)\gamma(x). \]
Proof. It is easy to calculate that $\kappa'_{t,t}(\gamma) = 2t(1-(t+1)\gamma)$. Denote $\gamma = \gamma_t(x)$. Since $x = \kappa_{t,t}(\gamma)$, we derive from (2.4) by the Implicit Function Theorem that

$$p'_{t,t}(x) = \frac{\kappa'_{t,t}(\gamma)}{\kappa_{t,t}(\gamma)} = \frac{t(r-1)r(r-1)\gamma^{r-2}(1-t+1)\gamma}{2t(1-(t+1)\gamma)} = \left(\frac{r}{2}\right)(t-1)(r-2)\gamma^{r-2}.$$ 

Likewise, $p''_{t,t}(x) = \frac{d}{dx}\left((\frac{r}{2})(t-1)(r-2)\gamma^{r-2}\right) / \kappa_{t,t}(\gamma)$, giving the stated formula. \qed

An informal explanation of the above formula for $h'_r(\alpha)$ is that this derivative measures the increment in the number of $r$-cliques in $H_{\alpha,n}$, normalised by $\binom{n}{r}^2$, when we increase $\alpha$ as $n \to \infty$. When we add $\lambda = o(n^2)$ new edges between the last two parts, we create around $\lambda \binom{k-1}{r-2} (cn)^{r-2}$ copies of $K_r$ while the change in the ratio $c$ has negligible effect because the (optimal) vector of part ratios is critical. Now note that $\binom{k-1}{r-2} (cn)^{r-2} \cdot \frac{n}{r} = \binom{r}{2} (k-1)(r-2)c^{r-2} + o(1)$.

The following lemma directly follows from the previous lemma and Taylor’s approximation.

Lemma 2.8. Let $\alpha \in [0,1]$. Let $k \in \mathbb{N}$ and $c$ be as in (2.7). If $\alpha \neq 1-\frac{1}{r}$ (that is, $\alpha$ is in the interior of $I_k$), then there is $\varepsilon > 0$ such that for each $\alpha' = \alpha \pm \varepsilon$, we have $\alpha' \in I_k$ and

$$h_r(\alpha') = h_r(\alpha) + \left(\frac{r}{2}\right)(k-1)(r-2)c^{r-2} (\alpha' - \alpha) \pm |\alpha' - \alpha|^{3/2}. \quad \Box$$

The following lemma proves Theorem 1.6 for the special edge densities where the function $h_r$ is not differentiable.

Lemma 2.9. Let $t \geq r-1$ be integer, $\alpha = 1-\frac{1}{r}$, and let $W$ be a graphon with $t(K_r,W) = \alpha$. If $t(K_r,W) = h_r(\alpha)$, then $W \in [W_{K_r}]$.

Proof. The quickest way to prove the lemma is to use the weaker version of a result of Lovász and Simonovits [14, Theorem 2] that every graph of order $n \to \infty$ with $(\alpha + o(1))\binom{n}{r}$ edges and $(h_r(\alpha) + o(1))\binom{n}{r}$ copies of $K_r$ is $o(n^2)$-close in edit distance to the Turán graph $T_r(n)$. Applying this result to a sequence of graphs $(G_n)_{n=1}^\infty$, where $G_n$ has $n$ vertices, that converges to the graphon $W$, we can transform each $G_n$ into $T_r(n)$ by changing $o(n^2)$ adjacencies. This change does not affect the convergence to $W$. Now, the limit of the $r$-partite Turán graphs is clearly $[W_{K_r}]$, giving the required. \qed

Remark 2.10. Alternatively, one can prove Lemma 2.9 operating with graphons only. Namely, the proof of Lovász and Simonovits [14, Theorems 1–2] for graphons would be to write $t(K_r,W)/t(K_r,W)$ as a telescopic product over $3 \leq s \leq r$ of $t(K_s,W)/t(K_{s-1},W)$ and bound each ratio separately, using the Cauchy-Schwartz Inequality (with double counting replaced by Tonelli’s theorem). In particular, since $t(K_r,W)$ is smallest possible, the graphon $W$ also minimises the triangle density for the given edge density $\alpha = 1-\frac{1}{r}$. By unfolding the corresponding argument from [14], one can show that the induced density of 3-sets spanning exactly one edge is 0. It follows with a bit of work that, similarly to graphs, $W$ is a complete partite graphon a.e. Now, a routine optimisation (see e.g. [19, Theorem 1.3]) shows that, apart a null-set, there are exactly $t$ parts of equal measure.

We will also need the following result, which is essentially a consequence of the piecewise concavity of the function $h_r$.
Lemma 2.11. For every $t \in \mathbb{N}$ and $\alpha \in [0, 1 - \frac{1}{t}]$, we have that $h_r(\alpha) \geq p_{r,t}(\alpha)$.

Proof. Let $x_0 := 1 - \frac{1}{t}$ and define

$$L_{r,t}(x) := p_{r,t}(x_0) + p_{r,t}'(x_0)(x - x_0), \quad \text{for } x \in \mathbb{R}.$$ 

In other words, $y = L_{r,t}(x)$ is the line tangent to the curve $y = p_{r,t}(x)$ at $x = x_0$. Since $y(x) \geq \frac{1}{t} > \frac{1}{1+t}$ for $x \leq x_0$ by (2.5), Lemma 2.7 gives that the function $p_{r,t}$ has the negative second derivative and is thus concave. We conclude that $p_{r,t}(x) \leq L_{r,t}(x)$ for all $x \leq x_0$.

Thus we are done if we show that $h_r(x) \geq L_{r,t}(x)$ for all $x \in [0, x_0]$. Note that $h_r(x_0) = L_{r,t}(x_0)$.

Since $h_r$ is a continuous function which is differentiable for every $x \in [0, x_0]$ apart finitely many points, it is enough to show by the Mean Value Theorem that $h_r'(x) \leq p_{r,t}'(x_0)$ for each $x \in [0, x_0]$ where $h_r$ is differentiable. So, let $x \in I_t$ with $0 \leq x < t$. Since $h_r = p_{r,t}$ on $I_t$ and, by Lemma 2.7, the derivative $p_{r,t}'$ is a decreasing function, it is enough to check that $p_{r,t}'(1 - \frac{1}{x^2}) \leq p_{r,t}'(1 - \frac{1}{x})$. Note that $y_m(1 - \frac{1}{m}) = \frac{1}{m}$ for each $m \in \mathbb{N}$. If $s \geq r - 2$, then by Lemma 2.7 we have that

$$\frac{p_{r,t}'(1 - \frac{1}{x})}{p_{r,t}'(1 - \frac{1}{x^2})} = \frac{(s - 1)(r - 2)}{(t - 1)(r - 2)} = \prod_{i=1}^{t-2} \frac{(s-i)}{s(t-i)} \leq 1$$

because $t(s - i) - s(t - i) = i(s - t) \leq 0$. If $s \leq r - 2$, then $p_{r,t}'(1 - \frac{1}{x}) = 0$ while $p_{r,t}'(x_0) \geq 0$, also giving the desired inequality. \hfill \Box

Lemma 2.12. Every graphon $(W, \mu)$ in $K_r$ is the limit of some sequence $(H_n)_{n=1}^\infty$ where $H_n \in K_{r,n}$ for each integer $n \geq 1$. Also, for all integers $n_1 < n_2 < \ldots$ and graphs $H_i \in K_{r,n_i}$ such that the sequence $(H_i)_{i=1}^\infty$ converges, its limit is in $K_r$.

Proof. Assume that $\alpha := t(K_2, W) < 1$ (as otherwise we can take $H_n$ to be the complete graph) and that $t(K_r, W) > 0$ (as $K_{r,n}$ contains all $K_r$-free graphs of order $n$). Let $\Omega_1 \cup \cdots \cup \Omega_k$ be the partition of the underlying space $[0, 1]$ for the graphon $W$, as in Definition 1.3.

For each $n \geq 1$, let $G_n \sim G(n, W)$ be a graph on $[n]$ which is an $n$-vertex sample of $W$, that is, we pick $n$ points $x_1, \ldots, x_n \in [0, 1]$ using the probability measure $\mu$ and make $i, j \in [n]$ adjacent with probability $W(x_i, x_j)$, with all choices being independent. Then the sequence $G_n$ converges to $W$ with probability 1, see Lovász and Szegedy [15, Corollary 2.6]. Each graph $G_n$ comes with a vertex partition $V_{n,1}, \ldots, V_{n,k}$, where we put $i \in V(G_n)$ into $V_{n,j}$ if $x_{n,i} \in \Omega_j$. By the Chernoff Bound, we have that $|V_{n,j}|/n$ converges to $\mu(\Omega_j) = c$ for every $j \in [k]$ as $n \to \infty$, with probability 1. Since $W$ is an (explicit) $\{0, 1\}$-valued function, we know all edges of $G_n$ apart from the ones inside $V_{n,k}$. Using that $\lim_{n \to \infty} t(K_3, G_n) = t(K_3, W)$ in the special cases $s = 2, 3$, we derive that $V_{n,k}$ induces $o(n^2)$ triangles in $G_n$ as well as the asymptotically correct number of edges. Fix a sequence $(G_n)_{n=1}^\infty$ that satisfies all above properties.

Now we are ready to show that the edit distance between $G_n$ and some graph in $K_{\alpha,n}$ is $o(n^2)$, which will be enough to prove the first part of the lemma. For each $n$, move $o(n)$ vertices between the parts of $G_n$ so that $|V_{n,i}| = \lfloor cn \rfloor$ for each $i \in [k-1]$. (The new adjacencies of a moved vertex are determined by its new part, except if we move a vertex into $V_{n,k}$ we make it adjacent to e.g. every other vertex for definiteness.) The new graphs $G_n$ still satisfy the above properties and have the correct part sizes. Using the Triangle Removal Lemma [29, 8] (see e.g. [10, Theorem 2.9]), we make $G_n[V_{n,k}]$ triangle-free by...
changing \( o(n^2) \) adjacencies. The definition of \( \mathcal{H}_{\alpha,n} \) requires to have exactly \( \lfloor cn \rfloor \cdot (|V_{n,k}|-\lfloor cn \rfloor) \) edges in \( V_{n,k} \). This can be achieved by [23, Lemma 2.2] which states that if \( G \) is triangle-free graph with \( m \rightarrow \infty \) vertices and \( s = e(G) + o(m^2) \) is at most \( t_2(m) \), then \( G \) is \( o(m^2) \)-close in edit distance to a triangle-free graph with exactly \( s \) edges, as desired.

Let us now show the second part of the lemma. Assume that a sequence \( (H_i)_{i=1}^{\infty} \) contradicts the statement. As \( \mathcal{H}_c \) contains the constant-1 graphon, the limiting density \( \alpha := \lim_{i \to \infty} t(K_2,H_i) \) must be in \( [0,1) \). Also, \( \lim_{i \to \infty} t(K_r,H_i) > 0 \) since \( \mathcal{H}_c \) contains all graphons with zero \( K_r \)-density.

Let \( V(H_i) = V_{i,1} \cup \cdots \cup V_{i,k-1} \cup U_i \) be the partition from the definition of \( H_i \). Let \( F_i := H_i[U_i] \). By the compactness of \( \mathcal{W} \), some subsequence of \( (F_i)_{i=1}^{\infty} \) converges to some graphon \( W' \). The limiting graphon \( W' \) has zero triangle density. Since we know all edges of \( H_i \) except inside \( U_i \), the graphon \( W' \) has the correct edge density. Now, define \( W \in \mathcal{H}_r \) as in Definition 1.3 with \( c = c(\alpha) \), \( k = k(\alpha) \), and \( W[\Omega_k] \) being weakly isomorphic to \( W' \). Since we know all adjacencies except inside \( \Omega_k \), a routine calculation shows that \( H_i \) converges to \( W \), as required.

\[ \square \]

### 2.4 Asymptotic structure from extremal graphons

We are ready to show that Theorem 1.6 implies Theorem 1.2 by adopting the analogous step from [23, Section 2.2].

**Proof of Theorem 1.2.** Suppose for the sake of contradiction that Theorem 1.2 is false, which is witnessed by some \( r \geq 4 \) and \( \varepsilon > 0 \). Thus we can find a sequence \( (G_n)_{n \in \mathbb{N}} \) of graphs of increasing orders \( v_n := v(G_n) \) such that \( t(K_r,G_n) = g_r(t(K_2,G_n)) + o(1) \) and each \( G_n \) is \( v_n^2 \)-far in edit distance from \( \mathcal{H}_{r,v_n} \). By using the compactness of \( \mathcal{W} \) and passing to a subsequence, we can additionally assume that the sequence \( (G_n)_{n \in \mathbb{N}} \) is convergent to some graphon \( W \). Let \( \alpha := t(K_2,W) \). Clearly, \( t(K_r,W) = g_r(\alpha) \). By Theorem 1.6, \( [W] \in \mathcal{H}_r \). By Lemma 2.12 pick \( H_n \in \mathcal{H}_{r,v_n} \) such that the sequence \( (H_n)_{n \in \mathbb{N}} \) converges to \( W \).

For two graphs \( G \) and \( H \) of the same order \( n \), define the cut distance \( \hat{\delta}_\square(G,H) \) to be the minimum over all bijections \( \phi : V(H) \rightarrow V(G) \) of \( \hat{d}(G,\phi(H)) \), where for graphs \( G \) and \( F \) with \( V(G) = V(F) \) we define

\[
\hat{d}(G,F) := \max_{S \subseteq V(G)} \frac{|e_G(S,T) - e_F(S,T)|}{v(G)^2},
\]

with \( e_G(S,T) := |\{(x,y) \in S \times T : \{x,y\} \in E(G)\}| \). Informally speaking, \( \hat{d}(G,F) \) is small if the two graphs on the same vertex set have similar edge distributions over all vertex cuts, while \( \hat{\delta}_\square \) is the version of \( \hat{d} \) where we look only at the isomorphism types of the graphs.

Theorems 2.3 and 2.7 in Borgs et al [2] give that \( \hat{\delta}_\square(G_n,H_n) \rightarrow 0 \). Namely, [2, Theorem 2.7] states that if two graphs have similar subgraph densities, then they are close in the fractional cut-distance \( \hat{\delta}_\square \) (which is defined the same way as \( \hat{\delta}_\square \) except, informally speaking, \( \phi \) distributes each vertex of \( H \) fractionally among \( V(G) \)), while [2, Theorem 2.3] provides an upper bound of \( \hat{\delta}_\square \) in terms of \( \hat{\delta}_\square \).

Up to relabelling of each \( H_n \), assume that \( \hat{d}(G_n,H_n) \rightarrow 0 \). Take any \( n \in \mathbb{N} \) and let \( v := v_n \). Fix the partition \( V(H_n) = V_1 \cup \cdots \cup V_{k-1} \cup U \) that was used to define \( H_n \in \mathcal{H}_{r,v} \). For \( i \in [k-1] \), if we use \( S = V_i \) and \( T = V_j \) (resp. \( T = V(H_n) \setminus V_i \)) in \( (2.9) \), then we conclude that \( V_i \) spans \( o(v^2) \) edges (resp. \( V_i \) is almost complete to the rest). 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 subgraphs induced by \( U \). Suppose that \( |U| = \Omega(v) \) for otherwise we get a
contradiction to Lemma 2.9. 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 \( K_r \)-density changes by \( o(1) \). Also, each edge of \( G_n[U] \) (and of \( H_n[U] \)) is in the same number of \( r \)-cliques whose remaining \( r-2 \) vertices are in \( V(G_n) \setminus U \). Since \( H_n[U] \) is triangle-free and \( G_n \) is asymptotically extremal, we conclude that \( G_n[U] \) spans \( o(v^3) \) triangles. We can change \( o(v^3) \) adjacencies and make \( G_n[U] \) to be triangle-free by the Triangle Removal Lemma and have the “correct” number of edges by [23, Lemma 2.2]. The obtained graph (which is \( o(v^2) \)-close in edit distance to \( G_n \)) is in \( \mathcal{H}_{r,v} \), contradicting our assumption.

3 Proof of the main result

Suppose that Theorem 1.6 is not true. Let \( r \geq 3 \) be the minimum integer such that there exists a \( K_r \)-extremal graphon \( W = (W, \mu) \) which does not belong to \( \mathcal{H}_r \). Let
\[
\alpha := t(K_2, W), \quad k := k(\alpha), \quad \text{and} \quad c := c(\alpha) \quad \text{as in (2.7)}.
\]
As \( [W] \notin \mathcal{H}_r \), we have \( 0 < \alpha < 1 \) and Theorem 1.5 implies that \( r \geq 4 \). We may further assume the following properties.

(W1) \( t(K_r, W) = h_r(\alpha) > 0 \) and \( t(K_3, W) > h_3(\alpha) \).

(W2) \( \alpha \in I_k \setminus \{1 - \frac{1}{k}\} \) and \( c \in (\frac{1}{k+1}, \frac{1}{k}) \).

Indeed, we may assume that \( W \) is not \( K_3 \)-extremal as otherwise \( W \in \mathcal{H}_3 \subseteq \mathcal{H}_r \) by Theorem 1.5. This, together with Theorem 1.4, implies (W1). As \( [W_k] \in \mathcal{H}_r \) for all \( t \in \mathbb{N} \), Lemma 2.9 implies (W2).

Our strategy is as follows. Using (W1), we will find a point \( x \in [0,1] \) such that \( t_x(K_r, W) \) is small while \( t_x(K_3, W) \) is large. (Recall that these are the densities of respectively \( K_r \) and \( K_3 \) rooted at \( x \).) Note that (2.2) translates these two values together with \( d_W(x) \) into \( t(K_r-1, N_W(x)) \) and \( t(K_2, N_W(x)) \). Hence, this will eventually enable us to translate the assumption \( [W] \notin \mathcal{H}_r \) into some conclusion about \( N_W(x) \), which will violate Theorem 1.4 for \( r-1 \).

In order to work in \( N_W(x) \), we need to relate \( d_W(x) \) and \( t_x(K_r, W) \). For this purpose, we will make use of the following auxiliary functions. For integer \( t \geq 3 \) and real \( x \in [0,1] \), define
\[
q_t(x) := (t-1)(d_W(x) - (k-1)c)(k-1)^{(t-2)}e^{t-2} + (k-1)^{(t-1)}e^{t-1}, \quad \text{and} \quad f_t(x) := q_t(x) - t_x(K_r, W).
\]

By Tonelli’s theorem (Theorem 2.3), \( d_W(x) \) and \( t_x(K_r, W) \) are (everywhere defined) Borel functions of \( x \in [0,1] \), so \( q_t \) and \( f_t \) are also Borel. Later, in Claim 1, we will show that \( f_t(x) = 0 \) for almost all \( x \), which provides the desired relation between \( d_W(x) \) and \( t_x(K_r, W) \).

We first prove the following lemma, which partly motivates the definition of the function \( f_t \).

Lemma 3.1. For each integer \( t \geq 3 \), we have
\[
\int_{[0,1]} f_t(x) \, d\mu(x) = h_t(\alpha) - t(K_r, W).
\]
We first show that we increase or decrease the density of positive. Indeed, if say \( \mu \) is of negligible measure. Note that both sets are Borel as while \( \alpha = t(K_3, W) \) is the interior of \( I_k \), where \( h_i \) is differentiable. Since we cannot push \( h_i(t(K_2, W)) - t(K_3, W) = 0 \) into positive values by Theorem 1.4, it follows that \( f_i(x) = 0 \) for almost every \( x \in [0, 1] \).

For the reader’s convenience, we present a direct proof of the claim. For each \( \gamma \geq 0 \), let

\[ U_{1, \gamma} := \{ x \in [0, 1] : f_i(x) > \gamma \} \quad \text{and} \quad U_{2, \gamma} := \{ x \in [0, 1] : f_i(x) < -\gamma \}. \]

Note that \( U_{1, \gamma}, U_{2, \gamma} \) are both Borel sets for all \( \gamma \geq 0 \) as the function \( f_i \) is Borel. Suppose that \( \mu(M_0) = \mu(U_{1, 0} \cup U_{2, 0}) > 0 \).

By (W1) and Lemma 3.1, we have \( \int_{[0, 1]} f_i(x) \, d\mu(x) = 0 \), implying that both \( \mu(U_{1, 0}) \) and \( \mu(U_{2, 0}) \) are positive. Indeed, if say \( \mu(U_{1, 0}) = 0 \) while \( \mu(U_{2, 0}) > 0 \), then \( \int_{[0, 1]} f_i(x) \, d\mu(x) < 0 \), a contradiction. For
where the final equality follows from Tonelli’s theorem (Theorem 2.3). In particular, as \( \gamma > 0 \) such that \( \mu(U_{1, \gamma}) > \gamma \) by Lemma 2.1 (applied to a countable sequence of \( \gamma \to 0 \)).

In brief, we derive a contradiction to the minimality of \( W \) by replacing a small subset of vertices with negative \( f_r \) by those with positive \( f_r \), and showing that this strictly decreases \( t(K_r, W) - h_r(t(K_2, W)) \).

Formally, we choose \( \eta, \epsilon \) such that

\[
0 < \eta \ll \epsilon \ll \gamma, 1/k, 1/r, c \leq 1.
\]

By Sierpinski’s theorem (Theorem 2.4) and since the measure \( \mu \) is non-atomic by the definition of a graphon, there exist sets \( U_1 \subseteq U_{1, \gamma} \) and \( U_2 \subseteq U_{2, \gamma} \) such that \( \mu(U_1) = \mu(U_2) = \eta/2 \). Let \( U := U_1 \cup U_2 \).

Consider the density function

\[
u(z) := \begin{cases} 1, & z \in [0, 1] \setminus U, \\ 1 + \epsilon, & z \in U_1, \\ 1 - \epsilon, & z \in U_2. \end{cases}
\]

Then \( \int_{[0,1]} \nu(z) \, d\mu(z) = 1 \), so \( d\mu'(z) := \nu(z) \, d\mu(z) \) is also a Borel probability measure. Let \( W' = (W, \mu') \) be the graphon with the same function \( W \) but with the new probability measure \( \mu' \). Recall the definitions in (2.3). As \( \mu(U) = \eta \ll 1/r \), the Union Bound gives that

\[
\mu'(U_{1, r_1}) \leq r\eta, \quad (3.2)
\]

and

\[
\mu'(U_{1, r_2}) \leq \sum_{\ell \geq 2} \binom{r}{\ell} \eta^\ell \leq r^2 \eta^2. \quad (3.3)
\]

Note that \( U^{2,0} = ([0,1] \setminus U) \times ([0,1] \setminus U) \). For each \( j \in [2] \), let

\[
V_j := (U_j \times ([0,1] \setminus U)) \cup (([0,1] \setminus U) \times U_j).
\]

Then \( U^{2,1} = V_1 \cup V_2 \). As \( U^{2,0} \) does not contribute to \( t(K_2, W') - t(K_2, W) \), we obtain

\[
t(K_2, W') - t(K_2, W) = \sum_{i \in [2]} \int_{U^{2,i}} \left( W'(x_1, x_2) \prod_{i \in [2]} d\mu'(x_i) - W(x_1, x_2) \prod_{i \in [2]} d\mu(x_i) \right)
\]

\[
= \left( \int_{V_1} - \int_{V_2} \right) \epsilon W(x_1, x_2) \prod_{i \in [2]} d\mu(x_i) \pm r^2 \eta^2
\]

\[
= 2 \left( \int_{U_1 \times [0,1]} - \int_{U_2 \times [0,1]} \right) \epsilon W(x_1, x_2) \prod_{i \in [2]} d\mu(x_i) \pm 2 \mu^2(U^2) \pm r^2 \eta^2
\]

\[
= 2 \epsilon \left( \int_{U_1} - \int_{U_2} \right) d\mu(x) \, d\mu(x) \pm 2 r^2 \eta^2, \quad (3.4)
\]

where the final equality follows from Tonelli’s theorem (Theorem 2.3). In particular, as \( \mu(U_1) = \mu(U_2) = \eta/2 \) and \( d\mu(x) \leq 1 \) for all \( x \in [0,1] \), (3.4) with the fact \( \eta \ll \epsilon \) implies

\[
|t(K_2, W') - t(K_2, W)| \leq 2 \epsilon \mu(U) + 2r^2 \eta^2 \leq 3\epsilon \eta. \quad (3.5)
\]
We now consider the increment in $K_r$-density. Again, $U^{r_0}$ does not contribute to $t(K_r, W') - t(K_r, W)$. Hence

$$
t(K_r, W') - t(K_r, W) = \int_{x \in U^{r_0}} \prod_{i \in [r]} W(x_i, x_j) \left( \prod_{i \in [r]} d\mu'(x_i) - \prod_{i \in [r]} d\mu(x_i) \right) \pm r^2 \eta^2
$$

where the second equality holds by symmetry between the variables $x_i$. By the definitions of $U_1$ and $U_2$, and by (3.1)–(3.3), we further have

$$
(3.6) \quad \leq \quad r \varepsilon (r - 1) (k - 1) (r - 2) e^{-2} \left( \int_{U_1} - \int_{U_2} \right) dW(x) d\mu(x) - r \gamma \varepsilon \mu(U) \pm 2r^2 \eta^2
$$

Let $\alpha' := t(K_2, W')$. Then, the above inequality, together with (3.4), Lemma 2.7 and (W2), implies that

$$
t(K_r, W') \quad \leq \quad t(K_r, W) + \binom{r}{2} (k - 1) (r - 2) e^{-2} (\alpha' - \alpha \pm 2r^2 \eta^2) - r \gamma \varepsilon \eta / 2
$$

On the other hand, by Lemma 2.8 and (3.5), we see that

$$
h_r(\alpha') \geq h_r(\alpha) + h'_r(\alpha')(\alpha' - \alpha) - |\alpha' - \alpha|^3 / 2 > h_r(\alpha) + h'_r(\alpha')(\alpha' - \alpha) - \eta^{3/2}.
$$

Hence, as $\eta \ll \varepsilon$, we have

$$
t(K_r, W') < h_r(\alpha) + h'_r(\alpha)(\alpha' - \alpha) - r \gamma \varepsilon \eta / 3 < h_r(\alpha) + h'_r(\alpha)(\alpha' - \alpha) - \eta^{3/2} < h_r(\alpha').
$$

This contradicts Theorem 1.4, proving the claim.

We next show that the set $N_0$, which consists of “$K_3$-heavy” points, has positive measure.

Claim 2. We have $\mu(N_0) > 0$. 

Discrete Analysis, 2020:19, 26pp.
We shall show in this subsection that almost every  

For each  

Proof. For each  

Let  

By Lemma 3.1, we have  

Thus we have  

implying that  

Consequently, as  

Claim 3. We have  

Proof. This claim also follows from Razborov’s differential calculus [26, Corollary 4.6]. In terms of graphs, the argument roughly says that if, on the contrary,  

then by removing a carefully selected subset of such edges we can destroy so many  

This claim also follows from Razborov’s differential calculus [26, Corollary 4.6]. In terms of graphs, the argument roughly says that if, on the contrary,  

implying that  

Consequently, as  

Again, let us give a direct proof of the claim. Suppose  

Note that  

forms a collection of nested sets and  

As  

forms a collection of nested sets and  

Thus  

We fix such  

By lowering the value of  

if necessary, and choosing a constant  

Assuming that  

for all }
By Sierpinski’s Theorem (Theorem 2.4), take a subset \( B \subseteq B_\varepsilon \) with \( \mu^2(B) = \eta \). By Lemma 2.6, there exists a symmetric Borel set \( C \subseteq B \) satisfying (C1)–(C3).

Define
\[
W'(x, y) := \begin{cases} 
(1 - \varepsilon)W(x, y), & \text{if } (x, y) \in C, \\
W(x, y), & \text{if } (x, y) \in [0, 1]^2 \setminus C.
\end{cases}
\]

As \( C \) and \( W \) are Borel, the function \( W' \) is also Borel. Let \( \alpha' := t(K_2, W') \). As \( \eta \ll \alpha \), we have
\[
\alpha' - \alpha = t(K_2, W') - t(K_2, W) = -\varepsilon \int_C W(x_1, x_2) \prod_{i=2} \mu(x_i).
\]

Since \( \varepsilon < W(x, y) \leq 1 \) for all \( (x, y) \in C \subseteq B_\varepsilon \), we also have that
\[
-\varepsilon \mu^2(C) \leq \alpha' - \alpha \leq -\varepsilon^2 \mu^2(C).
\]

For each \( x = (x_1, x_2) \in C \), (C2) implies that
\[
\int_{(x_3, \ldots, x_\ell) \in C^\ell(x)} \prod_{i \in \{\frac{\ell}{2}\} \setminus \{1, 2\}} W(x_i, x_j) \prod_{i=3}^{r} d\mu(x_i)
= \left( \int_{[0, 1]^\ell - 2} \int_{C^\ell(x)} \prod_{i \in \{\frac{\ell}{2}\} \setminus \{1, 2\}} W(x_i, x_j) \prod_{i=3}^{r} d\mu(x_i) \right)
= \left( \int_{(x_1, x_2) \in C} W(x_1, x_2) (t_{x_1, x_2}(K_2^-, W) \pm 2\varepsilon) \prod_{i \in [2]} d\mu(x_i) \right).
\]

Thus, by the symmetry between the variables \( x_i \) and Tonelli’s theorem, we have
\[
\int_{x \in C^\ell} \prod_{i \in \{\frac{\ell}{2}\}} W(x_i, x_j) \ d\mu'(x)
= \left( \int_{(x_1, x_2) \in C} W(x_1, x_2) \int_{(x_3, \ldots, x_\ell) \in C^\ell(x_1, x_2)} \prod_{i \in \{\frac{\ell}{2}\} \setminus \{1, 2\}} W(x_i, x_j) \prod_{i=3}^{r} d\mu(x_i) \right)
= \left( \int_{(x_1, x_2) \in C} W(x_1, x_2) t_{x_1, x_2}(K_2^-, W) \prod_{i \in [2]} d\mu(x_i) \pm 2\varepsilon \mu^2(C) \right).
\]

We shall bound \( K_\varepsilon \)-density in \( W' \) in two ways to derive a contradiction. First, note that for all \( 2 \leq \ell \leq \left( \frac{\ell}{2} \right) \) and \( (x_1, \ldots, x_\ell) \in C^{\ell,1} \), we have that
\[
\prod_{ij \in \{\frac{\ell}{2}\}} W'(x_i, x_j) = (1 - \varepsilon)^\ell \prod_{ij \in \{\frac{\ell}{2}\}} W(x_i, x_j).
\]

As \( C^{\ell,0} \) does not contribute to the change in \( K_\varepsilon \)-density, we have
\[
t(K_\varepsilon, W') - t(K_\varepsilon, W)
\]

**Discrete Analysis, 2020:19, 26pp.**

20
\[ = -\varepsilon \int_{x \in C^{r,1}} \prod_{ij \in \binom{[n]}{2}^r} W(x_i, x_j) \, d\mu^r(x) + \sum_{2 \leq l \leq \binom{n}{2}^r} (1 - (1 - \varepsilon)^{l-1}) \int_{x \in C^{r,l}} \prod_{ij \in \binom{[n]}{2}^r} W(x_i, x_j) \, d\mu^r(x) \]

\[ \overset{(3.11)}{=} -\varepsilon \binom{r}{2} \left( \int_{(x_1, x_2) \in C} W(x_1, x_2) t_{x_1, x_2}(K^r, W) \prod_{i \in [2]} d\mu(x_i) \pm 2r\eta \mu^2(C) \right) \pm r^4 \mu^r(C^2)\]

\[ \overset{(3.3)}{=} -\varepsilon \binom{r}{2} \left( \int_{(x_1, x_2) \in C} W(x_1, x_2)((k - 1)^{r-2} c^{r-2} + \varepsilon) \prod_{i \in [2]} d\mu(x_i) \right) \pm \eta \mu^2(C) \]

\[ \overset{(3.8)}{=} \binom{r}{2} ((k - 1)^{r-2} c^{r-2} + \varepsilon)(\alpha' - \alpha) \pm \eta \mu^2(C) \]

\[ \overset{\text{Lem.} \ 2.7}{=} h'_r(\alpha)(\alpha' - \alpha) + \binom{r}{2} \varepsilon(\alpha' - \alpha) \pm \eta \mu^2(C) \]

\[ \overset{(3.9)}{\leq} h'_r(\alpha)(\alpha' - \alpha) - \varepsilon^3 \mu^2(C)/2. \quad (3.12) \]

On the other hand, as \( \eta \ll \varepsilon \ll 1/k, 1/c, 1/r, \alpha \), by Theorem 1.4, Lemma 2.8 and (W2), we have

\[ t(K_r, W') \geq h_r(\alpha') \geq h_r(\alpha) + h'_r(\alpha)(\alpha' - \alpha) - |\alpha' - \alpha|^3/2 \]

\[ \overset{(3.9)}{\geq} t(K_r, W) + h'_r(\alpha)(\alpha' - \alpha) - \varepsilon^3/2(\mu^2(C))^{3/2} \]

\[ \overset{(C1)}{>} t(K_r, W) + h'_r(\alpha)(\alpha' - \alpha) - \varepsilon^3 \mu^2(C)/2, \]

a contradiction to (3.12). This proves the claim. \( \square \)

We can now show that \( D \), the set of “large degree” points, is negligible, thus imposing an additional “maximum degree” condition on our graphon.

**Claim 4.** We have \( \mu(D) = 0. \)

*Proof.* In graph theory language, the argument is informally as follows. Claim 3 bounds the number of \( r \)-cliques per typical edge of an almost extremal graph \( G \). This, by double counting, bounds the number of \( r \)-cliques per typical vertex \( x \) in terms of its degree. On the other hand, the last two parameters are linearly related by Claim 1. Putting all together, we derive the claim.

Let us provide details. Recall the definition of \( B_\gamma \) in (3.7). For each \( \gamma \geq 0 \), let

\[ B(\gamma) := \{ x \in [0, 1] : \mu \left( \{ y \in [0, 1] : (x, y) \in B_\gamma \} \right) > \gamma \}. \]

By Tonelli’s Theorem and Claim 3, we have

\[ 0 = \mu^2(B_\gamma) = \int_{x \in [0, 1]} \int_{y \in [0, 1]} d\mu(x) d\mu(y) \geq \int_{x \in B(\gamma)} \gamma \, d\mu(x) \geq \gamma \mu(B(\gamma)). \]

Thus we have \( \mu(B(\gamma)) = 0 \) for all \( \gamma > 0 \) and, by Lemma 2.1, \( \mu(B(0)) = \mu(\cup_{\gamma > 0} B(\gamma)) = 0 \). Hence, by Claim 1 it suffices to prove \( D \subseteq M_0 \cup B(0) \). By the definition of \( M_0 \), for each \( x \notin M_0 \cup B(0) \), we have \( f_r(x) = q_r(x) - t_\alpha(K_r, W) = 0 \). Tonelli’s theorem then implies that

\[ q_r(x) = (r - 1)(dW(x) - (k - 1)c)(k - 1)^{(r-2)c^{r-2} + (k - 1)(r-1)c^{r-1}} \]
We are now ready to derive the final contradiction. By Claims 1, 2 and 4, we have
\[ \mu \leq (\{ y \in \Omega : (x, y) \in B_s \}) + (k - 1)^{(r-2)}c^{r-2}d_W(x) \]
where the final inequality follows from the assumption \( x \notin B(0) \). Rearranging this, we obtain
\[ d_W(x) \leq kc, \]
showing that \( x \notin D \). Hence, \( D \subseteq M_0 \cup B(0) \) as claimed. \( \square \)

3.3 Putting everything together

We are now ready to derive the final contradiction. By Claims 1, 2 and 4, we have \( \mu(N_0 \setminus (D \cup M_0)) > 0 \).

Fix a point \( x \in N_0 \setminus (D \cup M_0) \).

By Tonelli’s Theorem, the function \( W(x, \cdot) \) is Borel. For brevity, set \( d := d_W(x) \). Note that \( f_\tau(x) = q_\tau(x) - t_\tau(K_r, W) = 0 \) as \( x \notin M_0 \).

Suppose first that \( d = 0 \). Then we have
\[ t_\tau(K_r, W) = q_\tau(x) = -(r - 1)(k - 1) \cdot (k - 1)^{(r-2)}c^{r-1} + (k - 1)^{(r-1)}c^{r-1} \]
a contradiction. Thus, we may assume that \( d > 0 \).

Let \( \tau := c/d \). As \( x \notin D \), we have
\[ \tau \geq \frac{1}{k}. \tag{3.13} \]

Consider \( W' := N_W(x) \), the neighbourhood of \( x \) in \( W \) as in Definition 2.5. As \( x \in N_0 \), we have \( f_3(x) = q_3(x) - t_3(K_3, W) < 0 \). We then derive from (2.2) and (W1) that
\[ \alpha' := r(K_2, W') \cdot \frac{t_3(K_3, W)}{d^2} > \frac{q_3(x)}{d^2} = \frac{2(d - (k - 1)c)(k - 1)c + (k - 1)(k - 2)c^2}{d^2} = 2(k - 1)\tau - k(k - 1)\tau^2. \tag{3.14} \]

Further define
\[ \rho := 2(k - 1)\tau - k(k - 1)\tau^2 = 1 - \frac{1}{k} - k(k - 1)\left( \tau - \frac{1}{k} \right)^2 \leq 1 - \frac{1}{k}. \tag{3.15} \]

Let us briefly overview where we stand in the proof now. Here, \( \alpha' \) is the edge density of \( N_W(x) \). Also, \( \rho \) is the edge density of the neighbourhood of a vertex of degree \( d = d_W(x) \) in an \( r \)-extremal graphon.
of the overall edge density $\alpha$, provided that this degree $d$ is realisable. The relation $f_r(x) = 0$ implies that the density of $K_{r-1}$ in $N_r(x)$ is as “expected”, see (3.16) below. In order to derive a contradiction to $\rho < \alpha$ we also need to exclude the case that $\rho$ is in the interior of the region when $h_{r-1}$ is zero.

**Claim 5.** $\rho \geq 1 - \frac{1}{r^2}$.

**Proof.** Since $f_r(x) = 0$, we have that $q_r(x) = t_s(K_r, W)$ is non-negative. This implies via an easy calculation that $d = d(x)$ is at least $\frac{r-2}{r-1} c k$. In turn, this and (3.13) give that $\tau = c/d$ lies between $\frac{1}{k}$ and $\frac{r-1}{(r-2)k}$. Since $\rho$ is a concave quadratic function of $\tau$, it is enough to verify that $\rho - (1 - \frac{1}{r^2})$ is non-negative for these end-points. Routine calculations give respectively $\frac{1}{r-2} - \frac{1}{k}$ and $\frac{k+1-r(r-3)}{k(r-2)^2}$, both of which are non-negative as $k \geq r-2$ (by $h_r(\alpha) > 0$).

Recall that $t_s(K_r, W) = q_r(x)$ and so (2.2) implies that

$$t(K_{r-1}, W') = \frac{t_s(K_r, W)}{d^{r-1}} = (r-1)\frac{(d-(k-1)c)(k-1)^{(r-2)}c^{r-2} + (k-1)^{(r-1)}c^{r-1}}{d^{r-1}}$$

$$= (r-1)(k-1)^{(r-2)}\tau^{r-2} - (r-2)k^{(r-1)}\tau^{r-1}. \quad (3.16)$$

We can also deduce from the definition of $\rho$ that

$$(k-1)(k-1-k\rho) \stackrel{(3.15)}{=} \left((k-1)(k\tau - 1)\right)^2.$$  

Here, the left-hand side is exactly the expression that appears under the square root when we define $p_{r-1,k-1}(\rho)$ in (2.6). Thus we have by (3.13) that

$$p_{r-1,k-1}(\rho) = \frac{(k-2)^{(r-3)}}{(k-1)^{(r-2)}k^{r-2}} \frac{(k-1+(k-1)(k\tau - 1))^{r-2} (k-1-(r-2)(k-1)(k\tau - 1))}{(r-1)(k-1)^{(r-2)}\tau^{r-2} - (r-2)k^{(r-1)}\tau^{r-1}}$$

$$\quad \stackrel{(3.16)}{=} t(K_{r-1}, W').$$

On the other hand, by (2.7) and (3.15), we have $k(\rho) \leq k - 1$. Recall that $\alpha' > \rho \geq 1 - \frac{1}{r^2}$. Thus, by Lemma 2.11 and the fact that $h_{r-1}$ is a strictly increasing function on $[1 - \frac{1}{r^2}, 1]$, we have

$$h_{r-1}(\alpha') > h_{r-1}(\rho) = p_{r-1,k(\rho)}(\rho) \geq p_{r-1,k-1}(\rho) = t(K_{r-1}, W').$$

Hence, we have $t(K_{r-1}, W') < h_{r-1}(\alpha')$ while $t(K_2, W) = \alpha'$, a contradiction to Theorem 1.4.

This completes the proof of Theorem 1.6.

**Acknowledgements**

The authors are grateful to the anonymous reviewer for carefully reading this paper.
References

[1] B. Bollobás. On complete subgraphs of different orders. *Math. Proc. Camb. Phil. Soc.*, 79:19–24, 1976. 3

[2] 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. 14

[3] S. Chatterjee. An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc. (N.S.)*, 53:617–642, 2016. 3

[4] S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41:2428–2461, 2013. 3

[5] D. L. Cohn. *Measure theory*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser/Springer, New York, second edition, 2013. 7

[6] P. Erdős. Some theorems on graphs. *Riveon Lematomatika*, 9:13–17, 1955. 2

[7] P. Erdős. On a theorem of Rademacher-Turán. *Illinois J. Math.*, 6:122–127, 1962. 2

[8] 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. 13

[9] D. C. Fisher. Lower bounds on the number of triangles in a graph. *J. Graph Theory*, 13:505–512, 1989. 3

[10] J. Komlós and M. Simonovits. Szemerédi’s regularity lemma and its applications to graph theory. In D. Miklós, V. T. Sós, and T. Szőnyi, editors, *Combinatorics, Paul Erdős is Eighty*, volume 2, pages 295–352. Bolyai Math. Soc., 1996. 13

[11] H. Liu, O. Pikhurko, and K. Staden. The exact minimum number of triangles in graphs of given order and size. *Forum of Math., Pi*, 8:144pp., 2020. 2

[12] L. Lovász. *Large Networks and Graph Limits*. Colloquium Publications. Amer. Math. Soc., 2012. 4, 5, 8

[13] L. Lovász and M. Simonovits. On the number of complete subgraphs of a graph. In *Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975)*, pages 431–441. Congressus Numerantium, No. XV, Winnipeg, Man., 1976. Utilitas Math. 2

[14] L. Lovász and M. Simonovits. On the number of complete subgraphs of a graph. II. In *Studies in pure mathematics*, pages 459–495. Birkhäuser, Basel, 1983. 2, 4, 12

[15] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory* (B), 96:933–957, 2006. 13
### Asymptotic Structure for the Clique Density Theorem

[16] L. Lovász and B. Szegedy. Regularity partitions and the topology of graphons. In I. Bárány and J. Solymosi, editors, *An Irregular Mind (Szemerédi is 70)*, volume 21 of *Bolyai Soc. Math. Studies*, pages 415–446. Springer, 2010.

[17] W. Mantel. Problem 28. *Winkundige Opgaven*, 10:60–61, 1907.

[18] J. W. Moon and L. Moser. On a problem of Turán. *Publ. Math. Inst. Hungar. Acad. Sci.*, 7:283–287, 1962.

[19] V. Nikiforov. The number of cliques in graphs of given order and size. *Trans. Amer. Math. Soc.*, 363:1599–1618, 2011.

[20] V. S. Nikiforov. On a problem of P. Erdős. *Annaire Univ. Sofia Fac. Math. Méc.*, 71:157–160, 1976/77.

[21] V. S. Nikiforov and N. G. Khadzhivanov. Solution of the problem of P. Erdős on the number of triangles in graphs with $n$ vertices and $[n^2/4]+l$ edges. *C. R. Acad. Bulgare Sci.*, 34:969–970, 1981.

[22] E. A. Nordhaus and B. M. Stewart. Triangles in an ordinary graph. *Can. J. Math.*, 15:33–41, 1963.

[23] O. Pikhurko and A. Razborov. Asymptotic structure of graphs with the minimum number of triangles. *Combin. Probab. Computing*, 26:138–160, 2017.

[24] O. Pikhurko, J. Sliacan, and K. Tyros. Strong forms of stability from flag algebra calculations. *J. Combin. Theory* (B), 135:129–178, 2019.

[25] C. Radin. Phases in large combinatorial systems. *Ann. Inst. Henri Poincaré D*, 5:287–308, 2018.

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

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

[28] C. Reiher. The clique density theorem. *Annals of Math.*, 184:683–707, 2016.

[29] 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.

[30] 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.

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

Jaehoon Kim
Korea Advanced Institute of Sciences and Technology
Daejeon, Republic of Korea
jaehoon.kim@kaist.ac.kr
https://sites.google.com/view/jaehoon-kim/

Hong Liu
Mathematics Institute and DIMAP
University of Warwick
Coventry, UK
h.liu.9@warwick.ac.uk
http://homepages.warwick.ac.uk/staff/H.Liu.9/

Oleg Pikhurko
Mathematics Institute and DIMAP
University of Warwick
Coventry, UK
o.pikhurko@warwick.ac.uk
https://homepages.warwick.ac.uk/~maskat/

Maryam Sharifzadeh
Department of Mathematics and Mathematical Statistics
Umeå University
Umeå, Sweden
maryam.sharifzadeh@umu.se