ASYMPTOTIC STRUCTURE FOR THE CLIQUE DENSITY THEOREM

JAEHOON KIM, HONG LIU, OLEG PIKHURKO, AND MARYAM SHARIFZADEH

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].

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 [Tur41]: 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 [Man07].

Rademacher (unpublished, see e.g. [Erd55, Erd62]) showed that, for even n, any n-vertex graph with $t_2(n) + 1$ edges contains at least $n/2$ triangles, where $t_r(n) := e(T_r(n))$ is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 [Erd55] 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 ([Erd55, Erd62, MM62, Nik76, NS63], 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 [Erd55] that $G_3(n, t_2(n) + q) \geqslant q \lfloor n/2 \rfloor$ for $q < \lfloor n/2 \rfloor$ when n is large was proved only two decades later by Lovász and Simonovits [LS76], with the proof of the conjecture also announced by Nikiforov and Khadzhiivanov [NK81].

Lovász and Simonovits [LS76, Conjecture 1] made the following bold conjecture. Let $\mathcal{H}$ consist of all graphs that can be obtained from a complete partite 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, $G_r(n, m) \leqslant H_r(n, m)$. 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 [LS76]). *For every integer $r \geqslant 3$, there is n_0 such that for all $n \geqslant n_0$ and $0 \leqslant m \leqslant \binom{n}{2}$, we have that $G_r(n, m) \geqslant H_r(n, m)$.*

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 [LPS17, Proposition 1.8]. Some large ranges of parameters when the conjecture has been proved are when m is slightly above $t_r(n)$ by Lovász and Simonovits [LS83] and

when $r = 3$ and $m/\binom{n}{2}$ is bounded away from 1 by Liu, Pikhurko and Staden [LPS17]. Still, Conjecture 1.1 remains open.

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

$$
\begin{aligned}
g_r(\alpha) &:= \lim_{n \to \infty} \frac{G_r(n, e(n))}{\binom{n}{r}}, \\
h_r(\alpha) &:= \lim_{n \to \infty} \frac{H_r(n, e(n))}{\binom{n}{r}}.
\end{aligned}
$$

It is not hard to see from basic principles (see e.g. [PST19, 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 $0 \leqslant \alpha < 1$. Let integer $k \geqslant 1$ satisfy $\alpha \in [1 - \frac{1}{k}, 1 - \frac{1}{k+1})$. 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 $|V_1| = \cdots = |V_k| = \lfloor cn \rfloor$, has edge density $(\alpha + o(1))\binom{n}{2}$ as $n \to \infty$. It is easy to show, see (2.8), that $c \leqslant \frac{1}{k}$. Thus $|V_{k+1}| = n - k\lfloor cn \rfloor \geqslant 0$. (In fact, our choice to round cn down was rather arbitrary, just to make the graph $H_{\alpha,n}$ well-defined for each α and n.) It is routine to show (see e.g. [Nik11, 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 $\alpha \leqslant 1 - \frac{1}{r-1}$ (when $k \leqslant r-2$). Also, as Lemma 2.7 implies, h_r consists of countably many concave "scallops" with cusps at $\alpha = 1 - \frac{1}{m}$ for integer $m \geqslant r-1$.

For a while, the best known lower bound on the limit function h_r, by Bollobás [Bol76], was the piecewise linear function which coincides with h_r on the cusp points. Fisher [Fis89] showed that $g_3(\lambda) = h_3(\lambda)$ for all $1/2 \leqslant \lambda \leqslant 2/3$, 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 [Raz07] and then to determine the whole function g_3 in [Raz08]. The function g_4 was determined by Nikiforov [Nik11] and the function g_r for any $r \geqslant 5$ was determined by Reiher [Rei16] (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 $G_r(n, m) + o(n^r)$ 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 [CD13] and its follow-up papers), phases in large graphs (see the survey by Radin [Rad18]), and large deviation inequalities for the clique density (see the survey by Chatterjee [Cha16]). 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 [Sim68]. 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, $\mathcal{H}_{0,n} := \{\overline{K_n}\}$ consists of the empty graph only.) Clearly, for any $r \geqslant 3$, the number of r-cliques in the obtained graph does not depend on the choice of the graph added on U. Also, $\mathcal{H}_{\alpha,n} \ni H_{\alpha,n}$ is always non-empty (but typically has many non-isomorphic graphs). Finally, for $r \geqslant 3$, let $\mathcal{H}_{r,n}$ be the union of $\mathcal{H}_{\alpha,n}$ over all $\alpha \in [0, 1]$ as well as the family all K_r-free n-vertex graphs.

Pikhurko and Razborov [PR17] proved that every almost extremal (n, m)-graph G is $o(n^2)$ close in the edit distance to some graph in $\mathcal{H}_{3,n}$. Our main result is to extend this structural result to all $r \geqslant 4$:

Theorem 1.2. *For every real $\varepsilon > 0$ and integer $r \geqslant 4$, there are $\delta > 0$ and n_0 such that every graph G with $n \geqslant n_0$ vertices and at most $(g_r(\alpha) + \delta)\binom{n}{r}$ 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 εn^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; 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 explore 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 the edit distance to $\mathcal{H}_{3,n}$ by the result in [PR17] and we are done since $\mathcal{H}_{3,n} \subseteq \mathcal{H}_{r,n}$. If $\alpha := m/\binom{n}{2}$ is $1 - \frac{1}{k} + o(1)$ for some integer $k \geqslant r$ (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 [LS83] that G has to be $o(n^2)$-close in the 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 α. 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 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 $r \geqslant 4$ 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, especially that the limit versions of the families $\mathcal{H}_{r,n}$, namely the families $\mathcal{H}_r$ defined in Section 1.2, are much cleaner to define. 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 preceeded 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 exellent book by Lovász [Lov12].

For the purposes of this paper, it is convenient to define a *graphon* as a pair (W, μ), where $W : [0,1] \times [0,1] \to [0,1]$ is a symmetric Borel function and μ 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 μ is understood). Each graph $G = (V, E)$ with $V = \{v_1, \ldots, v_n\}$ corresponds naturally to a graphon (W_G, μ) with μ being the Lebesgue measure on $[0,1]$ and W_G being the *adjacency function* of G which assumes value 1 on $[\frac{s-1}{n}, \frac{s}{n}) \times [\frac{t-1}{n}, \frac{t}{n})$ 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

$$t(F, W) := \int_{[0,1]^r} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{i=1}^r \mathbf{d}\mu(x_i).$$

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 [Lov12] gives several equivalent definitions of weak isomorphism. Let $[W]$ 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 $\mathcal{W} \subseteq [0,1]^{\mathcal{F}}$ 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, μ') with $\mu'(A) := \mu(A \cap B)/\mu(B)$ for Borel $A \subseteq [0,1]$. Since the new measure μ' is 0 on $[0,1] \setminus B$, this effectively restricts everything to B (while the scaling ensures that μ' is a probability measure).

Definition 1.3 (Graphon family $\mathcal{H}_r$). *Let $r \geqslant 3$. Define $\mathcal{F} := \{ [W] \in \mathcal{W} : t(K_r, W) = 0 \}$ and, for each integer $k \geqslant r - 1$, let $\mathcal{G}_k$ be the set of weak isomorphism classes of graphons (W, μ) 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 $i \in [2]$ and $x_i \in \Omega_{k_i}$, $W(x_1, x_2) = \begin{cases} 0, & \text{if } k_1 = k_2 \leq k - 1, \\ 1, & \text{if } k_1 \neq k_2 \in [k]. \end{cases}$

Further let

$$\mathcal{H}_r := \big\{ \, [\mathbf{1}_{[0,1]^2}] \, \big\} \cup \mathcal{F} \cup \bigcup_{k=r-1}^{\infty} \mathcal{G}_k.$$

This is a direct analogue of the graph family $\mathcal{H}_{r,n}$: for example, to constract 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 graphs 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, $\mathcal{H}_3 \subsetneq \mathcal{H}_4 \subsetneq \ldots \subseteq \mathcal{W}$.

The result of Reiher [Rei16] that determines the function g_r can be equivalently rephrased in the language of graphons as follows.

Theorem 1.4 ([Rei16]). *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 call a graphon W to be 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 [PR17, Theorem 2.1]):

Theorem 1.5 ([PR17]). *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 $\mathbf{I}_t := [1 - \frac{1}{t}, 1 - \frac{1}{t+1})$; 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* $\mathbf{1}_Y$ 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 ϕ 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. $c \ll b \ll a_1, \ldots, a_s$, then this means that there are coordinate-wise non-decreasing functions $f : (0, 1] \to (0, 1]$ and $g : (0, 1]^s \to (0, 1]$ such that the claimed result holds whenever $0 < c < f(b)$ and $0 < b < g(a_1, \ldots, a_s)$.

2.1. **Some notions and results of measure theory.** Let us recall some basic notions that apply when $\mathcal{A}$ is a σ-algebra on a set X and ν 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. [Coh13, Section 2.1]. A set $Y \subseteq X$ is called *(ν-) null* if there is $Z \in \mathcal{A}$ with $Z \supseteq Y$ and $\nu(Z) = 0$. We say that a property holds *(ν-) a.e.* if the set of x where it fails is ν-null. The ν-*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 σ-algebra generated by the union of $\mathcal{A}$ and all ν-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 σ-algebra generated by open subsets of $[0,1]^k$). It is easy to show (see e.g. [Coh13, Example 5.1.1]) that $\mathcal{B}([0,1]^k)$ is equal to the product of k copies of the Borel σ-algebra on $[0,1]$. When k is clear, we abbreviate $\mathcal{B}([0,1]^k)$ to $\mathcal{B}$.

Let μ be a probability measure on $([0,1], \mathcal{B})$. By μ^k, we denote the measure on $([0,1]^k, \mathcal{B})$ which is the product of k copies of μ. We call the sets in the μ^k-completion of $\mathcal{B}([0,1)^k)$ *measurable* and, when k is understood, denote this σ-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 addivitity of a measure, see e.g. [Coh13, 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 $\cup_{n\in\mathbb{N}} X_n$ 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 $f' = f$ ν-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 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 $A_i \triangle B_i \subseteq N_i$. Let $g := \sum_{i=0}^{\infty} 2^{-i} \mathbf{1}_{A_i}$. 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 $N := \cup_{i=0}^{\infty} N_i$, which has measure 0. Some points in $A_0 \subseteq N_0$ 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. $\square$

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. [Coh13, Proposition 5.2.1].

Theorem 2.3 (Tonelli's theorem). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be σ-finite measure spaces and let $f : X \times Y \to [0, \infty]$ be $\mathcal{A} \times \mathcal{C}$-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 $x \mapsto \int_Y f(x, y)\, d\nu(y)$ is $\mathcal{A}$-measurable;
(3) we have $\int_X \left(\int_Y f(x,y)\, d\nu(y) \right) d\mu(x) = \int_{X \times Y} f(x, y)\, d(\mu \times \nu)(x, y)$.

Also, we will need the following result.

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

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 μ); now, the Intermediate Value Theorem gives the required. $\square$

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 μ on $[0,1]$. Given such W, we can, by Lemma 2.2, modify it on μ^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 $\mathcal{W}$ by allowing non-uniform measures on $[0,1]$ as our definition is a special case of the general form of graphon, as defined in [Lov12, 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 [LS10] (see also [Lov12, 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 graph F on $[r]$ with $1, \ldots, k$ designated as roots, define its *rooted homomorphism density* in W by

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

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

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

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

For example, if $W = W_G$ for a graph G with $V(G) = [n]$ and $v_1 \in [n]$ then, for all $x \in [\frac{v_1-1}{n}, \frac{v_1}{n})$, $t_x(F,G)$ is the *rooted density* of F in (G, v_1), namely, the probability for independent uniformly distributed vertices $v_2, \ldots, v_r \in V(G)$ that for every $ij \in E(F)$ we have $v_i v_j \in E(G)$.

Define the *degree* of $x \in [0,1]$ in W by

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

When the graphon is undestood, we abbreviate it to $d(x)$. By Tonelli's theorem, $d(x)$ is defined for every $x \in [0,1]$ and

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

Definition 2.5. *For a graphon (W, μ) 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, μ'), where*

$$\mathbf{d}\mu'(y) := \frac{W(x,y)}{d(x)}\,\mathbf{d}\mu(y),$$

that is, $\mu'(A) := \int_A \frac{W(x,y)}{d(x)}\,\mathbf{d}\mu(y)$ 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 $r \geqslant 2$

$$
\begin{aligned}
t(K_r, N_W(x)) &= \int_{[0,1]^r} \prod_{1 \leqslant i < j \leqslant r} W(x_i, x_j) \prod_{i=1}^{r} \mathbf{d}\mu'(x_i) \\
&= \int_{[0,1]^r} \prod_{1 \leqslant i < j \leqslant r} W(x_i, x_j) \prod_{i=1}^{r} \frac{W(x, x_i)}{d_W(x)} \prod_{i=1}^{r} \mathbf{d}\mu(x_i)
\end{aligned}
$$

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

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

$$
\begin{aligned}
B^{r,\ell} &:= \left\{ \mathbf{x} \in [0,1]^r : \left| \left\{ S \in \binom{[r]}{m} : (x_i)_{i \in S} \in B \right\} \right| = \ell \right\}, \\
B_+^{r,\ell} &:= \bigcup_{i \geq \ell} B^{r,i}, \\
B^{r,\ell}(\mathbf{y}) &:= \left\{ (y_3, \ldots, y_r) \in [0,1]^{r-2} : (y_1, \ldots, y_r) \in B^{r,\ell} \right\}, \text{ and} \\
B_+^{r,\ell}(\mathbf{y}) &:= \bigcup_{i \geq \ell} B^{r,i}(\mathbf{y}).
\end{aligned}
\tag{2.3}
$$

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 ℓ. We will later need the property that $B_+^{r,2}$ is much smaller in measure than $B \subseteq [0,1]^2$. This is not true in general: for instance, consider $U \subseteq [0,1]$ with $\mu(U) = \eta \ll 1/r$ and $B = \{(x,y) : \{x,y\} \cap U \neq \emptyset\}$ when we have $\mu^r(B_+^{r,2}) = \Omega(\mu^2(B))$. 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 μ 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 $\mathbf{x} \in [0,1]^2$, we have $\mu^{r-2}\left(C_+^{r,2}(\mathbf{x}) \right) \leq 2r\eta$;*
(C3) $\mu^r(C_+^{r,2}) \leq r^3 \eta \mu^2(C)$.

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 $\mathbf{x} = (x_1, x_2) \in [0,1]^2$, 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 μ. Let E be the event that $\left| \left\{ ij \in \binom{[r]}{2} : (x_i, x_j) \in C \right\} \right| \geq 2$. 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

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

This implies (C2).

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

$$\mu^r(C_+^{r,2}) \leq \binom{r}{2} \int_{\mathbf{x} \in C} \mu^{r-2}\left(C_+^{r,2}(\mathbf{x}) \right) \prod_{i \in [2]} \mathbf{d}\mu(x_i) \leqslant \binom{r}{2} \mu^2(C) \cdot 2r\eta \leq r^3 \eta \mu^2(C).$$

Thus we have (C3). □

2.3. **Properties of $\mathcal{H}_r$ and h_r.** Everywhere in this section, $r \geqslant 3$ is fixed. In order to deal with the graphon family $\mathcal{H}_r$, it would be convenient to define some related parameters.

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

$$\kappa_{\ell,t}(\gamma) := \ell! \left(\binom{t}{\ell} \gamma^\ell + \binom{t}{\ell-1} \gamma^{\ell-1}(1 - t\gamma) \right)$$
$$= t^{(\ell-1)} \gamma^{\ell-1} (\ell - (\ell-1)(t+1)\gamma). \tag{2.4}$$

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

$$\gamma_t(x) := \frac{1}{t+1} + \frac{\sqrt{t(t - (t+1)x)}}{t(t+1)}. \tag{2.5}$$

This formula comes from taking $\gamma_t(x)$ to be the larger root γ of the quadratic equation $x = \kappa_{2,t}(\gamma)$. Further, again for $x \leqslant 1 - \frac{1}{t+1}$, define

$$p_{r,t}(x) := \kappa_{r,t}(\gamma_t(x))$$
$$= \frac{t^{(r-1)}}{t^r(t+1)^{r-1}} \left(t + \sqrt{t(t - (t+1)x)} \right)^{r-1} \left(t - (r-1)\sqrt{t(t - (t+1)x)} \right). \tag{2.6}$$

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

$$k = k(\alpha) \in \mathbb{N} \text{ such that } \alpha \in \mathbf{I}_k \text{ (that is, } 1 - \frac{1}{k} \leqslant \alpha < 1 - \frac{1}{k+1}),$$
$$c = c(\alpha) := \gamma_k(\alpha). \tag{2.7}$$

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 γ_t on the interval $\mathbf{I}_t$ for $t \in \mathbb{N}$. (Recall that these intervals partition $[0, 1)$.) Since the left and right limits of the function c coincide at any internal boundary point $1 - \frac{1}{t+1}$, with $t \in \mathbb{N}$ (namely, both are $\frac{1}{t+1}$), the explicit formula in (2.5) gives that c is a continuous and strictly monotone decreasing function on $[0, 1)$.

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

$$\frac{1}{k+1} < c \leqslant \frac{1}{k}. \tag{2.8}$$

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

The following lemma computes the first two derivatives of $p_{r,t}$ (and thus of h_r in all interior points of each $\mathbf{I}_t$), where we write these derivatives in terms of γ_t for convenience. Note that h_r is not differentiable at points $1 - \frac{1}{t}$ for integer $t \geqslant r - 1$: 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 $x \leqslant 1 - \frac{1}{t+1}$, we have that*

$$p'_{r,t}(x) = \binom{r}{2}(t-1)^{(r-2)}\gamma_t(x)^{r-2} \quad and \quad p''_{r,t}(x) = \frac{3\binom{r}{3}(t-1)^{(r-2)}\gamma_t(x)^{r-3}}{2t(1 - (t+1)\gamma_t(x))}.$$

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

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

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

An informal explanation of the above formula for $h'_r(\alpha)$ is that this derivative measures the increament in the number of r-cliques in $H_{\alpha,n}$, normalised by $\binom{n}{2}/\binom{n}{r}$, when we increase α 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 $\binom{k-1}{r-2}(cn)^{r-2} \cdot \binom{n}{2}/\binom{n}{r} = \binom{r}{2}(k-1)^{(r-2)}c^{r-2} + o(1)$.

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 $\alpha \neq 1 - \frac{1}{k}$ (that is, α is in the interior of $\mathbf{I}_k$), then there is $\varepsilon > 0$ such that for each $\alpha' = \alpha \pm \varepsilon$, we have $\alpha' \in \mathbf{I}_k$ and*

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

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

Lemma 2.9. *Let $t \geqslant r-1$ be integer, $\alpha = 1 - \frac{1}{t}$, and let W be a graphon with $t(K_2, W) = \alpha$. If $t(K_r, W) = h_r(\alpha)$, then $W \in [W_{K_t}]$.*

Proof. The quickest way to prove the lemma is to use the weaker version of a result of Lovász and Simonovits [LS83, Theorem 2] that every graph of order $n \to \infty$ with $(\alpha + o(1))\binom{n}{2}$ edges and $(h_r(\alpha) + o(1))\binom{n}{r}$ copies of K_r is $o(n^2)$-close in the edit distance to the Turán graph $T_t(n)$. 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 make each G_n into $T_t(n)$ by changing $o(n^2)$ adjacencies. This change does not affect the convergence to W. Now, the limit of the t-partite Turán graphs is clearly $[W_{K_t}]$, giving the required. □

Remark 2.10. Alternatively, one can prove Lemma 2.9 operating with graphons only. Namely, the proof of Lovász and Simonovits [LS83, Theorems 1–2] for graphons would be to write $t(K_r, W)/t(K_2, W)$ as a telescopic product over $3 \leqslant s \leqslant r$ 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 $\alpha = 1 - \frac{1}{t}$. By unfolding the corresponding argument from [LS83], 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. [Nik11, Theorem 1.3]) shows that, apart a null-set, there are exactly t parts of equal measure.

We will also need the following result, 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 $\gamma_t(x) \geqslant \frac{1}{t} > \frac{1}{t+1}$ for $x \leqslant x_0$ 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 $h_r(x) \geqslant L_{r,t}(x)$ for all $x \in [0, x_0]$. Note that $h_r(x_0) = L_{r,t}(x_0)$. Since h_r is a continuos 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 $h'_r(x) \leqslant p'_{r,t}(x_0)$ for each $x \in [0, x_0]$ where h_r is differentiable. So, let $x \in \mathbf{I}_s$ with $0 \leqslant s < t$. Since $h_r = p_{r,s}$ on $\mathbf{I}_s$ and, by Lemma 2.7, the derivative $p'_{r,s}$ is a decreasing function, it is enough to check

that $p'_{r,s}(1 - \frac{1}{s}) \leqslant p'_{r,t}(1 - \frac{1}{t})$. Note that $\gamma_m(1 - \frac{1}{m}) = \frac{1}{m}$ for each $m \in \mathbb{N}$. If $s \geqslant r - 2$, then by Lemma 2.7 we have that

$$\frac{p'_{r,s}(1 - \frac{1}{s})}{p'_{r,t}(1 - \frac{1}{t})} = \frac{(s-1)^{(r-2)}(\frac{1}{s})^{r-2}}{(t-1)^{(r-2)}(\frac{1}{t})^{r-2}} = \prod_{i=1}^{r-2} \frac{t(s-i)}{s(t-i)}$$

is at most 1 because $t(s-i) - s(t-i) = i(s-t) \leqslant 0$. If $s \leqslant r - 2$, then $p'_{r,s}(1 - \frac{1}{s}) = 0$ while $p'_{r,t}(x_0) \geqslant 0$, also giving the desired inequality. □

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

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 $\mathcal{H}_{r,n}$ 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 $n \geqslant 1$, let $G_n \sim \mathbb{G}(n, W)$ be a graph on $[n]$ which is an *n-vertex sample* of W, that is, we pick n points $x_{n,1}, \ldots, x_{n,n} \in [0,1]$ using the probability measure μ 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 [LS06, 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 $\lim_{n \to \infty} t(K_s, G_n) = t(K_s, W)$ in the special cases $s = 2, 3$, we derive that $V_{n,k}$ induces $o(n^3)$ 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 $\mathcal{H}_{\alpha,n}$ 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 [RS78, EFR86] (see e.g. [KS96, 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 [PR17, Lemma 2.2] which states that if G is triangle-free graph with $m \to \infty$ vertices and $s = e(G) + o(m^2)$ is at most $t_2(m)$, then G is $o(m^2)$-close in the edit distance to a triangle-free graph with exactly s edges, as desired.

Let us show the second part of the lemma. Assume that a sequence $(H_i)_{i=1}^{\infty}$ contradicts it. As $\mathcal{H}_r$ 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 $\mathcal{H}_r$ 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 $F_i := H_i[U]$. 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 Ω_k, a routine calculation shows that H_i converges to W, as required. □

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 [PR17, 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 $r \geqslant 4$ 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 εv_n^2-far in the 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 $\phi : V(H) \to V(G)$ of $\hat{d}(G, \phi(H))$, where for graphs G and F with $V(G) = V(F)$ we define

$$\hat{d}(G, F) := \max_{S,T \subseteq V(G)} \frac{|\, e_G(S, T) - e_F(S, T)\,|}{v(G)^2}, \tag{2.9}$$

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 [BCL+08] give that $\hat{\delta}_\square(G_n, H_n) \to 0$. Namely, [BCL+08, Theorem 2.7] states that if two graphs have similar subgraph densities, then they are close in the *fractional cut-distance* $\delta_\square$ (which is defined the same way as $\hat{\delta}_\square$ except, informally speaking, ϕ distributes each vertex of H fractionally among $V(G)$), while [BCL+08, Theorem 2.3] provides an upper bound of $\hat{\delta}_\square$ in terms of $\delta_\square$.

Up to relabelling of each H_n, assume that $\hat{d}(G_n, H_n) \to 0$. 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 $T = V_i$ (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 $o(v^2)$ adjacencies and make $G_n[U]$ to be triangle-free by the Triangle Removal Lemma and have the "correct" number of edges by [PR17, Lemma 2.2]. The obtained graph (which is $o(v^2)$-close in the edit distance to G_n) is in $\mathcal{H}_{r,v}$, contradicting our assumption. $\qquad\square$

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) $\alpha \in \mathbf{I}_k \setminus \{1 - \frac{1}{k}\}$ 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 $[W_{K_t}] \in \mathcal{H}_r$ 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. 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 $t \geqslant 3$ and real $x \in [0, 1]$, define

$$q_t(x) := (t - 1)(d_W(x) - (k - 1)c)(k - 1)^{(t-2)}c^{t-2} + (k - 1)^{(t-1)}c^{t-1}, \quad \text{and}$$
$$f_t(x) := q_t(x) - t_x(K_t, W). \tag{3.1}$$

By Tonelli's theorem (Theorem 2.3), $d_W(x)$ and $t_x(K_t, W)$ 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 $f_r(x) = 0$ 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 $t \geqslant 3$, we have*

$$\int_{[0,1]} f_t(x)\,\mathbf{d}\mu(x) = h_t(\alpha) - t(K_t, W).$$

Proof. By definition, we have

$$\int_{[0,1]} f_t(x)\,\mathbf{d}\mu(x) \quad = \quad (t - 1)\left(\int_{[0,1]} d_W(x)\,\mathbf{d}\mu(x) - (k - 1)c\right)(k - 1)^{(t-2)}c^{t-2}$$
$$+ (k - 1)^{(t-1)}c^{t-1} - \int_{[0,1]} t_x(K_t, W)\,\mathbf{d}\mu(x)$$
$$\overset{(2.1),(2.7)}{=} \quad (t - 1)\big(kc(2 - (k + 1)c) - (k - 1)c\big)(k - 1)^{(t-2)}c^{t-2}$$
$$+ (k - 1)^{(t-1)}c^{t-1} - t(K_t, W).$$

Recalling the definition of $h_t(\alpha)$ from (1.1), one can see that the right hand side above simplifies to $h_t(\alpha) - t(K_t, W)$, as desired. $\qquad\square$

We shall try to locate the desired point $x \in [0, 1]$ as outlined above in the following subsections.

3.1. **Almost all points are "K_r-typical".** We further introduce the following two sets. Let

$$M_0 \quad := \quad \{x \in [0, 1] : f_r(x) \neq 0\},$$

that is, M_0 is the set of "K_r-atypical" points. Let

$$N_0 \quad := \quad \{x \in [0, 1] : f_3(x) < 0\},$$

that is, N_0 is the set of "K_3-heavy" points. Note that both sets are Borel as f_3 and f_r are Borel functions. We first show that M_0 is of negligible measure.

Claim 1. *We have $\mu(M_0) = 0$.*

Proof. The statement that $f_r = 0$ a.e. follows with some calculations from Razborov's differential calculus [Raz07, Corollary 4.6]. Informally speaking, the quantity $f_r(x)$ measures the "contribution" of x to $h_r(t(K_2, W)) - t(K_r, W)$. The terms of f_r that are linear in $d(x)$ and $t_x(K_r, W)$ give the gradient when we increase or decrease the density of μ at x (while the constant term is chosen to make the average of f_r zero). Here $\alpha = t(K_2, W)$ is in the interior of $\mathbf{I}_k$, where h_r is differentiable. Since we cannot push $h_r(t(K_2, W)) - t(K_r, W) = 0$ into positive values by Theorem 1.4, it follows that $f_r(x) = 0$ for almost every $x \in [0, 1]$.

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

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

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

By (W1) and Lemma 3.1, we have $\int_{[0,1]} f_r(x)\,\mathbf{d}\mu(x) = 0$, 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 $\int_{[0,1]} f_r(x)\,\mathbf{d}\mu(x) < 0$, a contradiction. For $i \in [2]$, as $\bigcup_{\gamma>0} U_{i,\gamma} = U_{i,0}$ has positive measure and $\{U_{i,\gamma}\}_{\gamma>0}$ forms a nested collection of sets, there exists $\gamma > 0$ such that $\mu(U_{i,\gamma}) > \gamma$ by Lemma 2.1 (applied to a countable sequence of $\gamma \to 0$).

We choose η, ε such that

$$0 < \eta \ll \varepsilon \ll \gamma, 1/k, 1/r, c \leq 1.$$

By Sierpinski's theorem (Theorem 2.4) and since the measure μ 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

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

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

$$\mu^r(U^{r,1}) \leq r\eta, \tag{3.2}$$

and

$$\mu^r(U_+^{r,2}) \leq \sum_{\ell \geq 2} \binom{r}{\ell} \eta^\ell \leq r^2\eta^2. \tag{3.3}$$

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

$$
\begin{aligned}
t(K_2, W') - t(K_2, W) &= \sum_{\ell \in [2]} \int_{U^{2,\ell}} \left(W'(x_1, x_2) \prod_{i \in [2]} \mathbf{d}\mu'(x_i) - W(x_1, x_2) \prod_{i \in [2]} \mathbf{d}\mu(x_i) \right) \\
&\overset{(3.3)}{=} \left(\int_{V_1} - \int_{V_2} \right) \varepsilon W(x_1, x_2) \prod_{i \in [2]} \mathbf{d}\mu(x_i) \pm r^2\eta^2 \\
&= 2\left(\int_{U_1 \times [0,1]} - \int_{U_2 \times [0,1]} \right) \varepsilon W(x_1, x_2) \prod_{i \in [2]} \mathbf{d}\mu(x_i) \pm 2\mu^2(U^2) \pm r^2\eta^2 \\
&= 2\varepsilon \left(\int_{U_1} - \int_{U_2} \right) d_W(x)\,\mathbf{d}\mu(x) \pm 2r^2\eta^2, \tag{3.4}
\end{aligned}
$$

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

$$|t(K_2, W') - t(K_2, W)| \leq 2\varepsilon\mu(U) + 2r^2\eta^2 \leq 3\varepsilon\eta. \tag{3.5}$$

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

$$
\begin{aligned}
&t(K_r, W') - t(K_r, W) \\
&\quad = \int_{\mathbf{x} \in U^{r,1}} \prod_{ij \in \binom{[r]}{2}} W(x_i, x_j) \left(\prod_{i \in [r]} \mathbf{d}\mu'(x_i) - \prod_{i \in [r]} \mathbf{d}\mu(x_i) \right) \pm \mu^r(U_+^{r,2}) \\
&\quad \overset{(3.3)}{=} r \int_{x_1 \in U} \int_{U^{r-1,0}} \prod_{ij \in \binom{[r]}{2}} W(x_i, x_j) \left(\prod_{i \in [r]} \mathbf{d}\mu'(x_i) - \prod_{i \in [r]} \mathbf{d}\mu(x_i) \right) \pm r^2\eta^2
\end{aligned}
$$

$$= r\varepsilon \left(\int_{x_1 \in U_1} - \int_{x_1 \in U_2} \right) \int_{([0,1]\setminus U)^{r-1}} \prod_{ij \in \binom{[r]}{2}} W(x_i, x_j) \prod_{i \in [r]} \mathbf{d}\mu(x_i) \pm r^2 \eta^2$$

$$= r\varepsilon \left(\int_{x_1 \in U_1} - \int_{x_1 \in U_2} \right) (t_{x_1}(K_r, W) \pm \mu^{r-1}(U_+^{r-1,1})) \, \mathbf{d}\mu(x_1) \pm r^2 \eta^2, \qquad (3.6)$$

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

$$(3.6) \quad \leq \quad r\varepsilon \left(\int_{U_1} (q_r(x) - \gamma) - \int_{U_2} (q_r(x) + \gamma) \right) \mathbf{d}\mu(x) \pm 2r^2 \eta^2$$

$$\overset{\text{def of } q_r}{=} \quad r\varepsilon(r-1)(k-1)^{(r-2)} c^{r-2} \left(\int_{U_1} - \int_{U_2} \right) d_W(x) \, \mathbf{d}\mu(x) - r\gamma\varepsilon\mu(U) \pm 2r^2 \eta^2$$

$$\leq \quad \varepsilon r(r-1)(k-1)^{(r-2)} c^{r-2} \left(\int_{U_1} - \int_{U_2} \right) d_W(x) \, \mathbf{d}\mu(x) - r\gamma\varepsilon\eta/2.$$

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

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

$$< \quad h_r(\alpha) + h_r'(\alpha)(\alpha' - \alpha) - r\gamma\varepsilon\eta/3.$$

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. $\qquad \square$

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

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

Proof. For each $\gamma > 0$, let $N_\gamma := \{x \in [0,1] : f_3(x) < -\gamma\}$. Let

$$\beta := \frac{1}{2}(t(K_3, W) - h_3(\alpha)) \overset{(W1)}{>} 0.$$

By Lemma 3.1, we have $\int_{[0,1]} f_3(x) \, \mathbf{d}\mu(x) = h_3(\alpha) - t(K_3, W) = -2\beta$. On the other hand, we have that, rather roughly, $f_3(x) \geq q_3(x) - 1 \geqslant -k^2$ for all $x \in [0,1]$. Thus we have

$$-2\beta = \int_{[0,1]} f_3(x) \, \mathbf{d}\mu(x) \geq -k^2 \mu(N_\beta) - \beta(1 - \mu(N_\beta)),$$

implying that $\mu(N_\beta) \geq \beta/(k^2 - \beta) > 0$. Consequently, as $N_\beta \subseteq N_0$, we see that $\mu(N_0) \geq \mu(N_\beta) > 0$ as claimed. $\qquad \square$

3.2. **Maximum degree condition.** We shall show in this subsection that almost every $x \in [0,1]$ satisfies $d_W(x) \leq kc$. For this, let

$$D := \{x \in [0,1] : d_W(x) > kc\}$$

be the set of points with "too large degree". We shall see that D has measure zero. To show this we need one more statement about pairs that are "K_r-heavy". Let K_r^- denote the graph obtained from the complete graph on $[r]$ rooted at 1 and 2 by removing the edge $\{1, 2\}$. Define

$$B_* := \left\{ (x, y) \in [0,1]^2 : W(x, y) > 0 \text{ and } t_{x,y}(K_r^-, W) > (k-1)^{(r-2)} c^{r-2} \right\}, \qquad (3.7)$$

which one can think of as the set of edges that are "K_r-heavy". As W is Borel, the set B_* is Borel. The following claim states that most of the pairs of "adjacent" points are not contained in too many copies of K_r.

Claim 3. *We have $\mu^2(B_*) = 0$.*

Proof. This claim also follows from Razborov's differential calculus [Raz07, Corollary 4.6]. In terms of graphs, the argument roughly says that if, on the contrary, $\Omega(n^2)$ edges of an almost extremal (n, m)-graph G are each in too many copies of K_r (namely, in at least $H_r(n, m) - H_r(n, m-1) + \Omega(n^{r-2})$ copies), then by removing a carefully selected subset of such edges we can destroy so many r-cliques so that the asymptotic result (Theorem 1.4) is violated.

Again, let us give a direct proof of the claim. Suppose $\mu^2(B_*) > 0$. For each $\varepsilon > 0$, let

$$B_\varepsilon := \{(x, y) \in [0, 1]^2 : W(x, y) > \varepsilon \text{ and } t_{x,y}(K_r^-, W) \geq (k-1)^{(r-2)} c^{r-2} + \varepsilon\}.$$

Note that B_ε is a Borel set since the function $(x, y) \mapsto t_{x,y}(K_r^-, W)$ is Borel by Tonelli's theorem. As $\{B_\varepsilon\}_{\varepsilon > 0}$ forms a collection of nested sets and $\bigcup_{\varepsilon > 0} B_\varepsilon = B_*$, there is $\varepsilon > 0$ such that the $\mu^2(B_\varepsilon) \geq \varepsilon$. We fix such $\varepsilon > 0$. By lowering the value of ε if necessary, and choosing a constant η, we assume that

$$0 < \eta \ll \varepsilon \ll \mu^2(B), \alpha, 1/r, 1/k, c.$$

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

$$\alpha' - \alpha = t(K_2, W') - t(K_2, W) = -\varepsilon \int_C W(x_1, x_2) \prod_{i \in [2]} \mathbf{d}\mu(x_i). \tag{3.8}$$

Since $\varepsilon < W(x, y) \leq 1$ for all $(x, y) \in C \subseteq B_\varepsilon$, we also have that

$$-\varepsilon \mu^2(C) \leq \alpha' - \alpha \leq -\varepsilon^2 \mu^2(C). \tag{3.9}$$

For each $\mathbf{x} = (x_1, x_2) \in C$, (C2) implies that

$$\int_{(x_3,\ldots,x_r) \in C^{r,1}(\mathbf{x})} \prod_{ij \in \binom{[r]}{2} \setminus \{\{1,2\}\}} W(x_i, x_j) \prod_{i=3}^r \mathbf{d}\mu(x_i)$$

$$= \left(\int_{[0,1]^{r-2}} \pm \int_{C_+^{r,2}(\mathbf{x})} \right) \prod_{ij \in \binom{[r]}{2} \setminus \{\{1,2\}\}} W(x_i, x_j) \prod_{i=3}^r \mathbf{d}\mu(x_i)$$

$$\overset{(C2)}{=} t_{x_1,x_2}(K_r^-, W) \pm 2r\eta. \tag{3.10}$$

Thus, by the symmetry between the variables x_i and Tonelli's theorem, we have

$$\int_{\mathbf{x} \in C^{r,1}} \prod_{ij \in \binom{[r]}{2}} W(x_i, x_j) \, \mathbf{d}\mu^r(\mathbf{x})$$

$$= \binom{r}{2} \int_{(x_1,x_2) \in C} W(x_1, x_2) \int_{(x_3,\ldots,x_r) \in C^{r,1}(x_1,x_2)} \prod_{ij \in \binom{[r]}{2} \setminus \{\{1,2\}\}} W(x_i, x_j) \prod_{i \in [r]} \mathbf{d}\mu(x_i)$$

$$\overset{(3.10)}{=} \binom{r}{2} \int_{(x_1,x_2) \in C} W(x_1, x_2) \left(t_{x_1,x_2}(K_r^-, W) \pm 2r\eta \right) \prod_{i \in [2]} \mathbf{d}\mu(x_i)$$

$$= \binom{r}{2}\left(\int_{(x_1,x_2)\in C} W(x_1,x_2)t_{x_1,x_2}(K_r^-,W)\prod_{i\in[2]}\mathbf{d}\mu(x_i) \pm 2r\eta\mu^2(C)\right). \qquad (3.11)$$

We shall bound K_r-density in W' in two ways to derive a contradiction. First, note that for all $2\leqslant\ell\leqslant\binom{r}{2}$ and $(x_1,\ldots,x_r)\in C^{r,\ell}$, we have that $\prod_{ij\in\binom{[r]}{2}}W'(x_i,x_j)=(1-\varepsilon)^\ell\prod_{ij\in\binom{[r]}{2}}W(x_i,x_j)$. As $C^{r,0}$ does not contribute to the change in K_r-density, we have

$$t(K_r,W')-t(K_r,W)$$

$$= -\varepsilon\int_{\mathbf{x}\in C^{r,1}}\prod_{ij\in\binom{[r]}{2}}W(x_i,x_j)\ \mathbf{d}\mu^r(\mathbf{x}) \pm \sum_{2\leqslant\ell\leqslant\binom{r}{2}}(1-(1-\varepsilon)^\ell)\int_{\mathbf{x}\in C^{r,\ell}}\prod_{ij\in\binom{[r]}{2}}W(x_i,x_j)\ \mathbf{d}\mu^r(\mathbf{x})$$

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

$$\overset{(C3)}{\leq} -\varepsilon\binom{r}{2}\left(\int_{(x_1,x_2)\in C}W(x_1,x_2)((k-1)^{(r-2)}c^{r-2}+\varepsilon)\prod_{i\in[2]}\mathbf{d}\mu(x_i)\right) \pm \eta\mu^2(C)$$

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

$$\overset{\text{Lem } 2.7}{=} h_r'(\alpha)(\alpha'-\alpha)+\binom{r}{2}\varepsilon(\alpha'-\alpha)\pm\eta\mu^2(C)$$

$$\overset{(3.9)}{\leq} h_r'(\alpha)(\alpha'-\alpha)-\varepsilon^3\mu^2(C)/2. \qquad (3.12)$$

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

$$\begin{aligned} t(K_r,W') &\geq h_r(\alpha')\geq h_r(\alpha)+h'(\alpha)(\alpha'-\alpha)-|\alpha'-\alpha|^{3/2} \\ &\overset{(3.9)}{\geq} t(K_r,W)+h_r'(\alpha)(\alpha'-\alpha)-\varepsilon^{3/2}(\mu^2(C))^{3/2} \\ &\overset{(C1)}{>} t(K_r,W)+h_r'(\alpha)(\alpha'-\alpha)-\varepsilon^3\mu^2(C)/2, \end{aligned}$$

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

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 the 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 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_* in (3.7). For each $\gamma\geq 0$, let

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

By Tonelli's Theorem and Claim 3, we have

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

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 $f_r(x) = q_r(x) - t_x(K_r, W) = 0$. Tonelli's theorem then implies that

$$
\begin{aligned}
q_r(x) &= (r-1)(d_W(x) - (k-1)c)(k-1)^{(r-2)}c^{r-2} + (k-1)^{(r-1)}c^{r-1} \\
&= t_x(K_r, W) = \int_{y \in \Omega} t_{x,y}(K_r^-, W)W(x,y)\,\mathbf{d}\mu(y) \\
&= \left(\int_{y:(x,y)\in B_*} + \int_{y:(x,y)\notin B_*} \right) t_{x,y}(K_r^-, W)W(x,y)\,\mathbf{d}\mu(y) \\
&\overset{(3.7)}{\leq} \mu(\{y \in \Omega : (x,y) \in B_*\}) + (k-1)^{(r-2)}c^{r-2}d_W(x) \\
&= (k-1)^{(r-2)}c^{r-2}d_W(x),
\end{aligned}
$$

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. $\qquad\square$

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 $f_r(x) = q_r(x) - t_x(K_r, W) = 0$ as $x \notin M_0$.

Suppose first that $d = 0$. Then we have

$$
\begin{aligned}
t_x(K_r, W) = q_r(x) &= -(r-1)(k-1) \cdot (k-1)^{(r-2)}c^{r-1} + (k-1)^{(r-1)}c^{r-1} \\
&= -(r-2)k(k-1)^{(r-2)}c^{r-1} < 0,
\end{aligned}
$$

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 $f_3(x) = q_3(x) - t_x(K_3, W) < 0$. We then derive from (2.2) and (W1) that

$$
\begin{aligned}
\alpha' := t(K_2, W') = \frac{t_x(K_3, W)}{d^2} &> \frac{q_3(x)}{d^2} = \frac{2(d - (k-1)c)(k-1)c + (k-1)(k-2)c^2}{d^2} \\
&= 2(k-1)\tau - k(k-1)\tau^2. \tag{3.14}
\end{aligned}
$$

Further define

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

Claim 5. $\rho \geqslant 1 - \frac{1}{r-2}$.

Proof. Since $f_r(x) = 0$, we have that $q_r(x) = t_x(K_r, W)$ is non-negative. This implies via an easy calculation that $d = d(x)$ is at least $\frac{r-2}{r-1}ck$. In turn, this and (3.13) give that $\tau = c/d$ lies between $\frac{1}{k}$ and $\frac{r-1}{(r-2)k}$. Since ρ is a concave quadratic function of τ, 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 $k \geqslant r-2$ (by $h_r(\alpha) > 0$). $\qquad\square$

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

$$
\begin{aligned}
t(K_{r-1}, W') &= \frac{t_x(K_r, W)}{d^{r-1}} = \frac{(r-1)(d - (k-1)c)(k-1)^{(r-2)}c^{r-2} + (k-1)^{(r-1)}c^{r-1}}{d^{r-1}} \\
&= (r-1)(k-1)^{(r-2)}\tau^{r-2} - (r-2)k^{(r-1)}\tau^{r-1}. \tag{3.16}
\end{aligned}
$$

We can also deduce from the definition of ρ that

$$(k-1)(k-1-k\rho) \overset{(3.15)}{=} \big((k-1)(k\tau-1)\big)^2.$$

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

$$
\begin{aligned}
p_{r-1,k-1}(\rho) &= \frac{(k-2)^{(r-3)}}{(k-1)^{r-2}k^{r-2}}\left(k-1+(k-1)(k\tau-1)\right)^{r-2}\left(k-1-(r-2)(k-1)(k\tau-1)\right) \\
&= (r-1)(k-1)^{(r-2)}\tau^{r-2}-(r-2)k^{(r-1)}\tau^{r-1} \\
&\overset{(3.16)}{=} t(K_{r-1},W').
\end{aligned}
$$

On the other hand, by (2.7) and (3.15), we have $k(\rho) \le k-1$. Recall that $\alpha' > \rho \geqslant 1-\frac{1}{r-2}$. 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

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

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.

References

[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** (2008), 1801–1851. ↑12

[Bol76] B. Bollobás, *On complete subgraphs of different orders*, Math. Proc. Camb. Phil. Soc. **79** (1976), 19–24. ↑2

[CD13] S. Chatterjee and P. Diaconis, *Estimating and understanding exponential random graph models*, Ann. Statist. **41** (2013), 2428–2461. ↑2

[Cha16] S. Chatterjee, *An introduction to large deviations for random graphs*, Bull. Amer. Math. Soc. (N.S.) **53** (2016), 617–642. ↑2

[Coh13] D. L. Cohn, *Measure theory*, Second, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser/Springer, New York, 2013. ↑5, 6

[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** (1986), 113–121. ↑11

[Erd55] P. Erdős, *Some theorems on graphs*, Riveon Lematematika **9** (1955), 13–17. ↑1

[Erd62] P. Erdős, *On a theorem of Rademacher-Turán*, Illinois J. Math. **6** (1962), 122–127. ↑1

[Fis89] D. C. Fisher, *Lower bounds on the number of triangles in a graph*, J. Graph Theory **13** (1989), 505–512. ↑2

[KS96] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications to graph theory*, Combinatorics, Paul Erdős is eighty, 1996, pp. 295–352. ↑11

[Lov12] L. Lovász, *Large networks and graph limits*, Colloquium Publications, Amer. Math. Soc., 2012. ↑4, 7

[LPS17] H. Liu, O. Pikhurko, and K. Staden, *The exact minimum number of triangles in graphs of given order and size*, 2017. E-print arxiv:1712.00633. ↑1, 2

[LS06] L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory (B) **96** (2006), 933–957. ↑11

[LS10] L. Lovász and B. Szegedy, *Regularity partitions and the topology of graphons*, An irregular mind (Szemerédi is 70), 2010, pp. 415–446. ↑7

[LS76] L. Lovász and M. Simonovits, *On the number of complete subgraphs of a graph*, Proceedings of the fifth british combinatorial conference (univ. aberdeen, aberdeen, 1975), 1976, pp. 431–441. Congressus Numerantium, No. XV. ↑1

[LS83] L. Lovász and M. Simonovits, *On the number of complete subgraphs of a graph. II*, Studies in pure mathematics, 1983, pp. 459–495. ↑1, 3, 10

[Man07] W. Mantel, *Problem 28*, Winkundige Opgaven **10** (1907), 60–61. ↑1

[MM62] J. W. Moon and L. Moser, *On a problem of Turán*, Publ. Math. Inst. Hungar. Acad. Sci. **7** (1962), 283–287. ↑1

[Nik11] V. Nikiforov, *The number of cliques in graphs of given order and size*, Trans. Amer. Math. Soc. **363** (2011), 1599–1618. ↑2, 10

[Nik76] V. S. Nikiforov, *On a problem of P. Erdős*, Annuaire Univ. Sofia Fac. Math. Méc. **71** (1976/77), 157–160. ↑1

[NK81] V. S. Nikiforov and N. G. Khadzhiivanov, *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** (1981), 969–970. ↑1

[NS63] E. A. Nordhaus and B. M. Stewart, *Triangles in an ordinary graph*, Can. J. Math. **15** (1963), 33–41. ↑1

[PR17] O. Pikhurko and A. Razborov, *Asymptotic structure of graphs with the minimum number of triangles*, Combin. Probab. Computing **26** (2017), 138–160. ↑3, 5, 11, 12

[PST19] O. Pikhurko, J. Sliacan, and K. Tyros, *Strong forms of stability from flag algebra calculations*, J. Combin. Theory (B) **135** (2019), 129–178. ↑2

[Rad18] C. Radin, *Phases in large combinatorial systems*, Ann. Inst. Henri Poincaré D **5** (2018), 287–308. ↑2

[Raz07] A. Razborov, *Flag algebras*, J. Symb. Logic **72** (2007), 1239–1282. ↑2, 13, 16

[Raz08] A. Razborov, *On the minimal density of triangles in graphs*, Combin. Probab. Computing **17** (2008), 603–618. ↑2

[Rei16] C. Reiher, *The clique density theorem*, Annals of Math. **184** (2016), 683–707. ↑2, 5

[RS78] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics II, 1978, pp. 939–945. ↑11

[Sim68] M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), 1968, pp. 279–319. ↑2

[Tur41] P. Turán, *On an extremal problem in graph theory (in Hungarian)*, Mat. Fiz. Lapok **48** (1941), 436–452. ↑1

Mathematics Institute and DIMAP, University of Warwick, UK.

E-mail addresses:
{jaehoon.kim.1, h.liu.9, o.pikhurko, m.sharifzadeh}@warwick.ac.uk