INSTITUTE OF MATHEMATICS

THE CZECH ACADEMY OF SCIENCES

Cliques in dense inhomogenous random graphs

Martin Doležal
Jan Hladký
András Máthé

Preprint No. 53-2015

PRAHA 2015

CLIQUES IN DENSE INHOMOGENEOUS RANDOM GRAPHS

MARTIN DOLEŽAL, JAN HLADKÝ, AND ANDRÁS MÁTHÉ

Abstract. The theory of dense graph limits comes with a natural sampling process which yields an inhomogeneous variant of the Erdős–Rényi random graph. Here we study the clique number of these random graphs. For a large class of graphons, we establish a formula which gives the almost sure clique number of these random graphs.

In the process of doing so, we make an observation that might be of independent interest: Every graphon avoiding a fixed graph is countably-partite.

1. Introduction

The Erdős–Rényi random graph $\mathbb{G}(n,p)$ is a random graph with vertex set $[n]=\{1,\ldots,n\}$, where each edge is included independently with probability p. Since Gilbert, and independently Erdős and Rényi introduced the model in 1959, this has been arguably the most studied random discrete structure. Here, we recall facts about cliques in $\mathbb{G}(n,p)$; these were among the first properties studied in the model. The key question in the area concerns the order of the largest clique. We write $\omega(G)$ for the order of the largest clique in a graph G. Matula [17], and independently Grimmett and McDiarmid [8] have shown that when $p\in(0,1)$ is fixed, and $\varepsilon>0$ is arbitrary, we have

$$\lim_{n\to\infty}\mathbf{P}\left[\frac{\omega(\mathbb{G}(n,p))}{\log n}=(1\pm\varepsilon)\frac{2}{\log(1/p)}\right]=1\;. \tag{1}$$

Here, as well as in the rest of the paper, we use log for the natural logarithm. The actual result is stronger in two directions: firstly, it extends also to sequences of probabilities $p_n\to 0$, and secondly, Matula, Grimmett and McDiarmid proved an asymptotic concentration of $\omega(\mathbb{G}(n,p))$ on two consecutive values for which they provided an explicit formula.

Our aim however was extending (1) in a different direction. That direction is motivated by the theory of limits of dense graph sequences. Let us introduce the key concepts of the theory; for a thorough treatise we refer to Lovász's book [14]. The key object of that theory are graphons. A *graphon* is a symmetric measurable function $W:\Omega\times\Omega\to[0,1]$, where Ω is a probability space. In their foundational work, Lovász and Szegedy [15] proved that each sequence of finite graphs contains a subsequence that converges — in the so-called *cut metric* — to a graphon. This itself does not justify the graphons as limit objects; it still could be that the space of graphons is unnecessarily big. In other words, one would like to know that every graphon W is attained as a limit of finite graphs. To this end, Lovász and Szegedy introduced the random graph model $\mathbb{G}(n,W)$. The set of vertices of $G\sim\mathbb{G}(n,W)$ is the set $[n]$. To sample G, first generate n independent elements $x_1,\ldots,x_n\in\Omega$ according to the law of Ω. Then, connect i and j by an edge with probability $W(x_i,x_j)$ (independently of other choices). Lovász and Szegedy proved that with probability one, the sequence of samples from $\mathbb{G}(n,W)$ converges to W.

The strength of the theory of graph limits is that convergence in the cut metric implies convergence of many key statistics of graphs (or graphons). These include frequencies of appearances of small subgraphs, and normalized MAX-CUT-like properties. An important direction of research is to understand which other parameters are continuous with respect to the cut metric; those parameters can then be

2010 *Mathematics Subject Classification.* 05C80.

Key words and phrases. random graphs, graph limits, clique number.

JH: This research was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme (Grant Agreement Number 628974, project PaECiDM). Some of the work was done while supported by an EPSRC postdoctoral fellowship while affiliated with DIMAP and Mathematics Institute, University of Warwick.

AM is supported by a Leverhulme Trust Early Career Fellowship.

MD: This work was supported by RVO: 67985840.

defined even on graphons. In Example 3.1 below we show that the clique number can be very discontinuous with respect to the cut metric. Not all is lost even in such a situation. One can still study discontinuous parameters of a graphon W via the sampling procedure $\mathbb{G}(n, W)$. The random sampling will suppress pathological counterexamples such as those in Example 3.1, and thus allow us to associate limit information even to these limit parameters.

Let us explain more concretely what we consider the most important "limit information" in the case of the clique number. Suppose that $W\colon \Omega^2 \to [0,1]$ is such a graphon that $W(x,y) \in [p_1, p_2]$ for every $x, y \in \Omega$, where $0 < p_1 \leqslant p_2 < 1$ are fixed. Then the edges of $\mathbb{G}(n, W)$ are stochastically between $\mathbb{G}(n, p_1)$ and $\mathbb{G}(n, p_2)$. Thus, (1) tells us that the clique number $\omega(\mathbb{G}(n, W))$ asymptotically almost surely satisfies

$$(1 - o(1))\frac{2}{\log(1/p_1)} \leqslant \frac{\omega(\mathbb{G}(n, W))}{\log n} \leqslant (1 + o(1))\frac{2}{\log(1/p_2)} . \tag{2}$$

Thus, it is actually plausible to believe that

$$\frac{\omega(\mathbb{G}(n, W))}{\log n} \tag{3}$$

converges in probability to a positive constant $\kappa(W)$, and the immediate question is what $\kappa(W)$ is. In this paper, we study this and related questions. In Theorem 5.1 we prove that if W is a graphon with strictly positive essential infimum then

$$\kappa(W) = \sup\left\{ \frac{2\|h\|_1^2}{\int_x \int_y h(x)h(y) \log\left(1/W(x,y)\right)} \ : \ h \text{ is a nonnegative } L^1\text{-function on } \Omega \right\}$$

(see also Fact 5.2 and Proposition 6.1 for alternative expressions of $\kappa(W)$).

A particular subclass of the random graph models $\mathbb{G}(n, W)$ are the so-called *stochastic block models* introduced already in 1980's in the field of mathematical sociology, [10]. They are used extensively in many areas of mathematics, computer science, and physics. In our language, (the dense version of) stochastic block models correspond to the case when W is a step-function with finitely many or countably many steps. There exist conceptually simpler proofs to our main results when restricted to stochastic block models. This simplification would occur both on the real-analytic side (i.e., general measurable functions versus step-functions) and on the combinatorial side. See our remark in Section 5.3.

1.1. **Cliques and independent sets.** While we formulate all the problems in terms of cliques, we could have worked with the complementary notion of independent sets instead. Indeed, investigating one of these notions with respect to $\mathbb{G}(n, W)$ is equivalent to investigating the other with respect to $\mathbb{G}(n, 1-W)$.

1.2. **Organization of the paper.** In Section 2 we introduce necessary notation and review basic facts from real analysis and theory of graph limits.

In Section 3 we give some motivating examples. In particular, in Example 3.1 we show that the order of the largest clique in a deterministic sequence of graphs is indeed a very discontinuous parameter. In Example 3.3 we exhibit a family of graphons for which $\omega(\mathbb{G}(n, W))$ for different n's oscillates between typically extremely large values and typically extremely small values. Theorem 3.5 then shows that $\omega(\mathbb{G}(n, W))$ is always concentrated around its mean.

In Section 4 we study a bipartite counterpart to the main problem. It turns out that this problem is easier to attack, and also has a simpler answer.

In Section 5 we present our main result, Theorem 5.1. In Theorem 5.1 we determine the value $\kappa(W)$ coming from (3) for each graphon W whose essential infimum is strictly positive. We provide heuristics for the formula for the $\kappa(W)$. Unfortunately, we were unable to turn these relatively natural heuristics into a rigorous proof. The actual proof of Theorem 5.1 is given in Section 7, building on tools from Section 6.

In Section 8 we treat graphons W for which the expected value of $\omega(\mathbb{G}(n, W))$ is bounded. In particular, we prove that such graphons must be countably-partite. So, while the motivation for the contents of Section 8 stemmed from investigating inhomogeneous random graphs, the main result of this section is structural and deterministic.

Finally, in Section 9 we suggest some further directions of research.

2. Preliminaries

2.1. **Notation.** For $n \in \mathbb{N}$, we write $[n] = \{1, \ldots, n\}$, and $[n]_0 = \{0, 1, \ldots, n\}$. As always, we denote by $\binom{n}{k}$ the binomial coefficient $\frac{n!}{k!(n-k)!}$. For the multinomial coefficients of higher orders, we employ the notation $\binom{n}{k_1|k_2|\ldots|k_l} = \frac{n!}{k_1!k_2!\ldots k_l!(n-\sum k_i)!}$ (here we suppose $k_1 + k_2 + \ldots + k_l \leqslant n$). We omit rounding symbols where it does not affect correctness of the calculations.

We shall always assume that Ω is a standard Borel probability space without atoms. We always write ν for the probability measure associated with Ω. For a function f we write $\int f$ or $\int f \mathrm{d}(\nu)$ for its integral, depending on whether the underlying measure ν is clear from the context.

We shall sometimes make use of tools from real analysis which are available only for $\mathbb{R}$ and $\mathbb{R}^d$. The Lebesgue measure will be denoted by λ. It should be always clear from the context whether we mean the one-dimensional Lebesgue measure on $\mathbb{R}$, two dimensional Lebesgue measure on $\mathbb{R}^2$ or any higher dimensional Lebesgue measure.

We write $\|\cdot\|_1$ to denote the L^1-norm of functions (or vectors in a finite-dimensional space). Non-negative vectors, and non-negative L^1-functions[1] are called *histograms* (see our explanation at the end of Section 5.1). For a histogram f, we write $\mathrm{Box}(f)$ for all histograms g for which $g \leqslant f$ (pointwise). We say that a histogram is *non-trivial* if it is not almost everywhere zero.

We recall the notions of *essential supremum* and *essential infimum.* Suppose that Ω is a space equipped with a measure ν. For a measurable function $f\colon X \to \mathbb{R}$ we define $\operatorname{ess\,sup} f$ as the least number a such that $\nu(\{x \in \Omega : f(x) > a\}) = 0$. The quantity $\operatorname{ess\,inf} f$ is defined analogously.

2.2. **Random graphs $\mathbb{H}(n, W)$.** There is a natural intermediate step when obtaining the random graph $\mathbb{G}(n, W)$ from a graphon W which is often denoted by $\mathbb{H}(n, W)$. To obtain $\mathbb{H}(n, W)$ we sample n random independent points $x_1, \ldots, x_n$ from the probability space underlying W. The random graph has the vertex set $[n]$. The edge-set is an edge-set of a complete graph equipped with edge-weights. The weight of the edge ij is $W(x_i, x_j)$. Self-loops are not included.

2.3. **Graphons.** The above crash course in graph limits almost suffices for the purposes of this paper, and we need only a handful of additional standard definitions. See [14] for further references.

All (non-discrete) probability spaces in this paper are standard Borel probability spaces without atoms. Recall that the Isomorphism Theorem (see e.g. [11, Theorem 17.41]) tells us that there is a measure-preserving isomorphism between each two such spaces (i.e. a bijection between the spaces such that this function and its inverse are measurable and preserve measures). In particular, suppose that $W : \Omega^2 \to [0, 1]$ is a graphon defined on a probability space Ω, and let X be another probability space. Let us fix an isomorphism $\psi : X \to \Omega$ between X and Ω. By a *representation of W on X* we mean the graphon $W' : X^2 \to [0, 1]$, $(x, y) \mapsto W(\psi(x), \psi(y))$. Of course, the representation depends on the actual choice of the isomorphism ψ. Note however that the distribution of $\mathbb{G}(n, W')$ does not depend on the choice of ψ as it is the same as the distribution of $\mathbb{G}(n, W)$.

We shall need to "zoom in" on a certain part of a graphon. The next definition is used to this end.

Definition 2.1. Suppose that $W : \Omega^2 \to [0, 1]$ is a graphon on a probability space Ω with a measure ν. By a *subgraphon of W obtained by restricting to a set $A \subseteq \Omega$* of positive measure we mean a function $U : A \times A \to [0, 1]$ which is simply the restriction $W \restriction_{A\times A}$. When working with this notion, we need to turn A to a probability space. That is, we view U as a graphon on the probability space A endowed with measure $\nu_A(B) := \frac{\nu(B)}{\nu(A)}$ for every measurable set $B \subseteq A$.

Observe that in the above setting for every $B \subseteq A$ of positive measure we have

$$\frac{1}{\nu(B)^2} \int_{B\times B} \log(1/W) \mathrm{d}(\nu^2) = \frac{1}{\nu_A(B)^2} \int_{B\times B} \log(1/W) \mathrm{d}(\nu_A^2) \ . \tag{4}$$

Note that a lower-bound on $\omega(\mathbb{G}(n, W \restriction_{A\times A}))$ provides readily a lower-bound on $\omega(\mathbb{G}(n, W))$. More precisely, suppose that we can show that asymptotically almost surely, $\omega(\mathbb{G}(n, W \restriction_{A\times A})) \geqslant c \log n$. Consider now sampling the random graph $\mathbb{G}(n, W)$. By the Law of Large Numbers, out of the n sampled points $x_1, \ldots, x_n \in \Omega$, there will be $(\nu(A) - o(1))n > \frac{1}{2}\nu(A)n$ many of them contained in A. In other

[1]the vector-/function-space will be clear from the context

words, there is a coupling of $G = \mathbb{G}(n, W)$ and $G' = \mathbb{G}(\frac{1}{2}\nu(A)n, W\restriction_{A\times A})$ such that with high probability, G' is contained in G as a subgraph. We conclude that

$$\omega(\mathbb{G}(n, W)) \overset{\text{a.a.s.}}{\geqslant} \omega\left(\mathbb{G}(\tfrac{1}{2}\nu(A)n, W\restriction_{A\times A})\right) \overset{\text{a.a.s.}}{\geqslant} c\log\left(\tfrac{1}{2}\nu(A)n\right) = (c - o(1))\log n\ . \tag{5}$$

The *homomorphism density* of a graph $H = (\{v_1, \ldots, v_\ell\}, E)$ in a graphon W is defined by

$$t(H, W) = \int_{x_1,\ldots,x_\ell} \prod_{i<j : v_iv_j\in E} W(x_i, x_j)\ .$$

Suppose that Ω is an atomless standard Borel probability space. Let $W_1, W_2 : \Omega^2 \to [0, 1]$ be two graphons. We then define the *cut-norm distance* of W_1 and W_2 by

$$d_\square(W_1, W_2) = \sup_{S,T}\left|\int_{x\in S}\int_{y\in T} W_1(x, y) - W_2(x, y)\right|,$$

where S and T range over all measurable subsets of Ω. Strictly speaking, $d_\square$ is only a pseudometric since two graphons differing on a set of measure zero have zero distance. Based on the cut-norm distance we can define the key notion of *cut distance* by

$$\delta_\square(W_1, W_2) = \inf_\varphi d_\square(W_1^\varphi, W_2)\ , \tag{6}$$

where $\varphi : \Omega \to \Omega$ ranges through all measure preserving automorphisms of Ω, and W_1^φ stands for a graphon defined by $W_1^\varphi(x, y) = W_1(\varphi(x), \varphi(y))$. Then $\delta_\square$ is also a pseudometric.

Suppose that $H = (\{v_1, \ldots, v_\ell\}, E)$ is a graph (which is allowed to have self-loops), and let Ω be an arbitrary atomless standard Borel probability space. By a *graphon representation* W_H *of* H we mean the following construction. We consider an arbitrary partition $\Omega = A_1\dot\cup A_2\dot\cup \ldots \dot\cup A_\ell$ into sets of measure $\frac{1}{\ell}$ each. We then define the graphon W_H as 1 or 0 on each square $A_i \times A_j$, depending on whether v_iv_j forms an edge or not. Note that W_H is not unique since it depends on the choice of the sets $A_1, \ldots, A_\ell$. However, all the possible graphons W_H are at zero distance in the $\delta_\square$-pseudometric. So, when writing W_H we refer to any representative of the above class. With this in mind, we can also define the cut distance of H and any graphon $W : \Omega^2 \to [0, 1]$, denoted by $\delta_\square(H, W)$, as $\delta_\square(W_H, W)$. Also, all of this extends in a straightforward way to weighted graphs with a weight function $w : E \to [0, 1]$.

The key property of the sampling procedures described earlier (both $\mathbb{G}(n, W)$ and $\mathbb{H}(n, W)$) is that the samples are typically very close to the original graphon W in the cut distance. Let us state this fact (for the sampling procedure $\mathbb{H}(n, W)$), proven first in [3], formally.

Lemma 2.2 ([14, Lemma 10.16])**.** *Let W be an arbitrary graphon. Then for each $k \in \mathbb{N}$ with probability at least $1 - \exp(-\frac{k}{2\log k})$, we have $\delta_\square\big(\mathbb{H}(k, W), W\big) \leqslant \frac{20}{\sqrt{\log k}}$.* □

Remark 2.3. In some situations, we shall consider a wider class of kernels, the so-called L^∞-graphons. These are just symmetric L^∞-functions $W : \Omega^2 \to \mathbb{R}_+$. That is, we do not require L^∞-graphons to be bounded by 1 but rather by an arbitrary constant. The random graph $\mathbb{H}(n, W)$ makes sense even for L^∞-graphons.[2] Just by rescaling, we can formulate Lemma 2.2 by saying that $\delta_\square\big(\mathbb{H}(k, W), W\big) \leqslant \frac{20\|W\|_\infty}{\sqrt{\log k}}$ with probability at least $1 - \exp(-\frac{k}{2\log k})$ for an arbitrary L^∞-graphon W. □

Let us note that the proof of Lemma 2.2 is quite involved.

2.4. **Lebesgue points and approximate continuity.** Let $f : (0, 1)^2 \to \mathbb{R}$ be an integrable function. Recall that $(x, y) \in (0, 1)^2$ is a called a *Lebesgue point* of f if we have

$$\lim_{r\to 0+} \frac{1}{\lambda(M_r(x, y)))}\int_{(s,t)\in M_r} |f(s, t) - f(x, y)| = 0, \tag{7}$$

where $M_r(x, y) = [x - r, x + r] \times [y - r, y + r]$. Recall that $(x, y) \in \mathbb{R}^2$ is a *point of density* of a set $A \subseteq \mathbb{R}^2$ if

$$\lim_{r\to 0+} \frac{\lambda(M_r(x, y) \setminus A)}{\lambda(M_r(x, y)))} = 0. \tag{8}$$

[2]But $\mathbb{G}(n, W)$ does not make sense.

Recall also that a function $f\colon (0,1)^2 \to \mathbb{R}$ is said to be *approximately continuous* at $(x,y) \in (0,1)^2$ if for every $\varepsilon > 0$, the point (x,y) is a point of density of the set $\{(s,t) \in (0,1)^2 \colon |f(s,t) - f(x,y)| < \varepsilon\}$.

We will use the following classical result (see e.g. [19, Theorem 7.7]).

Theorem 2.4. *Let $f\colon (0,1)^2 \to \mathbb{R}$ be an integrable function. Then almost every point of $(0,1)^2$ is a Lebesgue point of f. In particular, we have*

$$\lim_{r\to 0+} \frac{1}{\lambda(M_r(x,y)))} \int_{(s,t)\in M_r} f(s,t) = f(x,y)$$

for almost every $(x,y) \in (0,1)^2$. □

An easy consequence of the previous theorem is also the following classical result.

Theorem 2.5. *Let $f\colon (0,1)^2 \to \mathbb{R}$ be a measurable function. Then f is approximately continuous at almost every point of $(0,1)^2$.* □

2.5. **Exhaustion principle.** We recall the principle of exhaustion (see e.g. [6, Lemma 11.12] for a more general statement).

Lemma 2.6. *Let $\mathcal{C}$ be a collection of measurable subsets of $(0,1)$ with positive Lebesgue measure. Suppose that for every $A \subseteq (0,1)$ with positive Lebesgue measure, there is $C \in \mathcal{C}$ such that $C \subseteq A$. Then there is an at most countable subcollection $\mathcal{B}$ of $\mathcal{C}$ of pairwise disjoint sets such that $\sum_{B\in\mathcal{B}} \lambda(B) = 1$.* □

3. Examples, concentration and oscillation

The first set of examples shows that the clique number is not even semicontinuous (when normalized by $\log n$) with respect to the cut-distance.

Example 3.1. Let $f : \mathbb{N} \to \mathbb{N}$ be a function that tends to infinity very slowly.

- Let us consider a sequence of n-vertex graphs consisting of a clique of order $\lceil n/f(n)\rceil$ and $n - \lceil n/f(n)\rceil$ isolated vertices. This sequence converges to the 0-graphon, the smallest graphon in the graphon space. Yet, the clique numbers are unusually high, almost of order $\Theta(n)$.
- Let us consider a sequence of n-vertex $f(n)$-partite Turán graphs. This sequence converges to the 1-graphon, the largest graphon in the graphon space. Yet, the clique numbers tend to infinity very slowly. □

By the previous example, we see that

- there are sequences of finite graphs with clique numbers growing much faster than logarithmic with a limit graphon $W \equiv 0$,
- while there are other sequences of finite graphs with clique numbers growing much slower than logarithmic with a limit graphon $W \equiv 1$.

In the introduction, we saw that for many natural graphons W, $\omega(\mathbb{G}(n,W))$ grows logarithmically. It is easy to construct a graphon for which $\omega(\mathbb{G}(n,W))$ grows for example as $\log\log n$, or another graphon for which $\omega(\mathbb{G}(n,W))$ grows for example as $n^{0.99}$. More surprisingly, our next construction shows that we can have an oscillation between these two regimes even for one graphon. We shall need the following well-known crude bound on the minimum difference between uniformly random points.

Fact 3.2. *Suppose that numbers $x_1,\dots,x_n$ are uniformly sampled from the interval $(0,1)$. Then asymptotically almost surely, $\min_{i\neq j}|x_i - x_j| > n^{-2}$.*

Example 3.3. For an arbitrary function $f : \mathbb{N} \to \mathbb{R}_+$ with $\lim_{n\to\infty} f(n) = +\infty$ there exists a graphon W and a sequence of integers $1 = \ell_0 < k_1 < \ell_1 < k_2 < \ell_2 < \dots$ such that asymptotically almost surely,

$$\omega(\mathbb{G}(k_i,W)) < f(k_i)\ , \text{ and} \tag{9}$$

$$\omega(\mathbb{G}(\ell_i,W)) > \frac{\ell_i}{f(\ell_i)}\ . \tag{10}$$

Let us explain this construction. Consider a sequence of positive numbers $1 = a_1 > a_2 > a_3 > \ldots > 0$, with $\lim_{n\to\infty} a_n = 0$, to be determined later. Define a graphon $W : (0,1)^2 \to [0,1]$ as

$$W(x,y) = \begin{cases} 0 & \text{if } a_{2i-1} \geqslant |x-y| > a_{2i}, \\ 1 & \text{if } a_{2i} \geqslant |x-y| > a_{2i+1}. \end{cases}$$

Let us show how to achieve (9). Suppose that numbers $a_1, \ldots, a_{2i-1}$ were already set. Fix an arbitrary number n large enough such that $n^{-2} < a_{2i-1}$ and $f(n) > 1 + 1/a_{2i-1}$. Then, set $a_{2i} := n^{-2}$. We claim that with high probability, there is no set of $f(n)$ vertices in $\mathbb{G}(n,W)$ forming a clique. Indeed, consider the representation of the vertices of $\mathbb{G}(n,W)$ in the interval $[0,1]$. By Fact 3.2 we can assume that the mutual distances between these points are more than a_{2i}. Consider an arbitrary set $S \subseteq (0,1)$ of these points of size $f(n)$. By the pigeonhole principle there are two points $x, y \in S$ with $|x-y| \leqslant 1/(f(n)-1) < a_{2i-1}$. On the other hand, $|x-y| > a_{2i}$. We conclude that $W(x,y) = 0$, and thus S does not induce a clique.

Next, let us show how to achieve (10). Suppose that numbers $a_1, \ldots, a_{2i}$ were already set. Fix a large number n. In particular, suppose that $n^{-2} < a_{2i}$ and $f(n) > \frac{2}{a_{2i}}$, and let $a_{2i+1} := n^{-2}$. Now, consider the process of generating vertices in $\mathbb{G}(n,W)$. By the Law of Large Numbers, out of n vertices, with high probability, at least $\frac{1}{2}a_{2i}n$ vertices fall in the interval $(\frac{1}{2} - \frac{a_{2i}}{2}, \frac{1}{2} + \frac{a_{2i}}{2})$. By Fact 3.2, with high probability, the differences of pairs of these points are bigger than a_{2i+1}. In particular, the said set of vertices forms a clique of order at least $\frac{1}{2}a_{2i}n > \frac{n}{f(n)}$, as needed for (10). □

Remark 3.4. It may look that replacing the values 0 and 1 by some constants $0 < p_1 < p_2 < 1$ in the above example we get an oscillation between $c_1 \log n$ and $c_2 \log n$. Theorem 5.1 tells us however that this is not the case: the clique number normalized by $\log n$ will converge in probability. □

While Example 3.3 shows that the long-term behavior of $\omega(\mathbb{G}(n,W))$ can be quite wild, for a fixed (but large) n, the distribution of $\omega(\mathbb{G}(n,W))$ is concentrated. The proof of this result was suggested to us by Lutz Warnke.

Theorem 3.5. *For each graphon W and each n, we have that for each $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbf{P}\left[\,\left|\frac{\omega(\mathbb{G}(n,W))}{\mathbf{E}[\omega(\mathbb{G}(n,W))]} - 1\right| > \varepsilon\,\right] = 0\ .$$

Proof. Suppose that W is represented on a probability space Ω. We write $\mu(n,W) := \mathbf{E}[\omega(\mathbb{G}(n,W))]$. Here, we only prove the case when $\mu(n,W) \to \infty$. The complementary case when $\mu(n,W)$ is bounded is covered by Lemma 8.1.

To prove concentration, we shall use Talagrand's inequality. For this, we need to represent $\mathbb{G}(n,W)$ on a suitable product space J. It turns out that the right product space corresponds to "vertex-exposure" technique known in the theory of random graphs. Let $J := \prod_{i=1}^n J_i$, where $J_i = \Omega \times [0,1]^{i-1}$. This indeed is a "vertex-exposure model" of $\mathbb{G}(n,W)$. To see this, consider an arbitrary element $\mathbf{x} \in J$. We can write

$$\mathbf{x} = (x_1, (\,); x_2, (p_{1,2}); x_3, \begin{pmatrix} p_{1,3} \\ p_{2,3} \end{pmatrix}; \ldots; x_n, \begin{pmatrix} p_{1,n} \\ p_{2,n} \\ \cdots \\ p_{n-1,n} \end{pmatrix})\ ,$$

where $x_i \in \Omega$, and $p_{i,j} \in [0,1]$. In the instance of $\mathbb{G}(n,W)$ corresponding to $\mathbf{x}$, vertices i and j are connected if and only if $W(x_i,x_j) \geqslant p_{i,j}$. It is straightforward to check that this gives the right distribution on $\mathbb{G}(n,W)$.

Consider the clique number, this time on the domain J. That is, we have a function $\Psi : J \to \mathbb{R}$, where $\Psi(\mathbf{x})$ is the clique number of the graph corresponding to $\mathbf{x}$. Then Ψ is a (discrete) 1-Lipschitz function. That is if $\mathbf{x}, \mathbf{y} \in J$ are such that they differ in one coordinate, then $|\Psi(\mathbf{x}) - \Psi(\mathbf{y})| \leqslant 1$.[3] Further, Ψ satisfies the so-called *small certificates condition*. This means that whenever $\Psi(\mathbf{x}) \geqslant \ell$, there exists a set C of at most ℓ many coordinates such that $\Psi(\mathbf{y}) \geqslant \ell$ for each $\mathbf{y} \in J$ which agrees with $\mathbf{x}$ on each coordinate from C. (In other words, the values of $\mathbf{x}$ on coordinates from C alone certify that $\Psi(\mathbf{x}) \geqslant \ell$.)

[3]One could consider a stronger notion of 1-Lipschitzness, namely, to require that changing $\mathbf{x}$ on an E-coordinate by ε would change the value of our function by at most ε. This clearly is not true for Ψ. However, the weaker version is sufficient for our purposes.

Indeed, it is enough just to reveal the values at the indices of one maximum clique. Talagrand's inequality (see [18, Remark 2 following Talagrand's Inequality II, p. 81])[4] then states that there exists an absolute constant $\beta > 0$ such that for $t_n = (\mu(n, W))^{\frac{3}{4}}$, we have (for every large enough n) that

$$\mathbf{P}\left[|\omega(\mathbb{G}(n,W)) - \mu(n,W)| > t_n\right] \leqslant 2\exp\left(-\frac{\beta t_n^2}{\mu(n,W)}\right) = 2\exp\left(-\beta\sqrt{\mu(n,W)}\right) .$$

The conclusion immediately follows by letting n go to infinity. □

4. Biclique number

As a warm-up for the study of the key quantity (3), we first deal with its bipartite counterpart. To this end, we shall work with *bigraphons*. Bigraphons, introduced first in [16], arise as limits of balanced bipartite graphs. A bigraphon is a measurable function $U : \Omega_1 \times \Omega_2 \to [0, 1]$. Here, Ω_1 and Ω_2 are probability spaces which represent the two partition classes, and the value $U(x, y)$ represents the edge intensity between the parts corresponding to x and y. This suggests the sampling procedure for generating the inhomogeneous random bipartite graph $\mathbb{B}(n, U)$: we sample uniformly and independently at random points $x_1, \ldots, x_n$ from Ω_1 and $y_1, \ldots, y_n$ from Ω_2. In the bipartite graph $\mathbb{B}(n, U)$ with colour classes $\{a_i\}_{i=1}^n$ and $\{b_j\}_{j=1}^n$, we connect a_i with b_j with probability $U(x_i, y_j)$.

When we refer to a bipartite graph $H = (V, W; E)$ as a *bigraph*, we consider a distinguished order of the colour classes $V = \{v_1, \ldots, v_p\}$ and $W = \{w_1, \ldots, w_q\}$. In such a case we define the *bipartite density* of H in U by

$$t_{\mathrm{B}}(H, U) = \int_{x_1,\ldots,x_p} \int_{y_1,\ldots,y_q} \prod_{ij: v_i w_j \in E} U(x_i, y_j) . \tag{11}$$

Note that for a bigraph $H = (V, W; E)$ and its conjugate $H' = (W, V; E)$ the quantities $t_{\mathrm{B}}(H, U)$ and $t_{\mathrm{B}}(H', U)$ are not necessarily equal.

Next, we define the natural bipartite counterpart to the clique number. Given a bipartite graph $G = (A, B; E)$, we define its *biclique number* as the largest ℓ such that there exist sets $X \subseteq A$, $Y \subseteq B$, $|X| = |Y| = \ell$, that induce a complete bipartite graph. We denote the biclique number of G by $\omega_2(G)$.

The main result concerning the biclique number is the following.

Theorem 4.1. *Let $U : \Omega_1 \times \Omega_2 \to [0, 1]$ be a bigraphon whose essential supremum $p = \operatorname{ess\,sup} U$ is strictly between zero and one. Then we asymptotically almost surely have*

$$\omega_2(\mathbb{B}(n, U)) = (1 \pm o(1)) \cdot \frac{2}{\log 1/p} \cdot \log n .$$

As we will see, the upper bound in Theorem 4.1 is trivial. For the lower bound, we need to make a small detour to Sidorenko's conjecture.

4.1. Sidorenko's conjecture.

A famous conjecture of Simonovits and Sidorenko, "Sidorenko's conjecture", [21, 20] asserts that among all graphs of a given (large) order n and fixed edge density d, the density of a fixed bipartite graph is minimized for a typical sample from $\mathbb{G}(n, d)$. The conjecture can be particularly neatly phrased in the language of graphons — as observed already by Sidorenko a decade before the notion of graphons itself — in which case it asserts that

$$t_{\mathrm{B}}(H, U) \geqslant \left(\int_x \int_y U(x, y)\right)^{e(H)} , \tag{12}$$

for each bigraphon U and each bigraph H. Despite recent results [9, 4, 13, 12], the conjecture is wide open. We shall need the solution of Sidorenko's Conjecture for $H = K_{n,m}$ which was observed already by Sidorenko. We give a short self-contained and unoriginal proof here.

Proposition 4.2. *Suppose that U is an arbitrary bigraphon and $n, m \in \mathbb{N}$ are arbitrary. Then for the complete bigraph $K_{n,m}$ we have $t_{\mathrm{B}}(K_{n,m}, U) \geqslant \left(\int_x \int_y U(x, y)\right)^{nm}$.*

---

[4] Actually, as was communicated to us by Lutz Warnke and Mike Molloy, there is a typo in [18]. The effect of this typo, however, is only the value of the constant β below. Since we do not make β explicit, this typo is irrelevant.

Proof. We have

$$\begin{aligned}\int_{x_1,\dots,x_n}\int_{y_1,\dots,y_m}\prod_{i\in[n],j\in[m]}U(x_i,y_j) &= \int_{x_1,\dots,x_n}\Big(\int_y\prod_{i\in[n]}U(x_i,y)\Big)^m \geqslant \Big(\int_{x_1,\dots,x_n}\int_y\prod_{i\in[n]}U(x_i,y)\Big)^m\\ &= \Big(\int_y\int_{x_1,\dots,x_n}\prod_{i\in[n]}U(x_i,y)\Big)^m = \Big(\int_y\Big(\int_x U(x,y)\Big)^n\Big)^m\\ &\geqslant \Big(\int_x\int_y U(x,y)\Big)^{nm},\end{aligned}$$

where both inequalities follow by applications of Hölder's inequality. □

4.2. **Bicliques in almost constant bigraphons.** As a preliminary step for our proof of Theorem 4.1, we study bicliques in random bipartite graphs sampled from almost constant bigraphons. This condition is formalized by the following definition.

Definition 4.3. A bigraphon $U:\Omega_1\times\Omega_2\to[0,1]$ is (d,ε)-*constant* if $\int_x\int_y U(x,y)\geqslant d$ and $\operatorname{ess\,sup} U\leqslant d+\varepsilon$.

Proposition 4.4. *Let $0<d_1<d_2<1$ be given. Then for every $\alpha\in(0,1)$ there exists $\varepsilon\in(0,1)$ such that the following holds: Whenever we have $d\in(d_1,d_2)$ and a (d,ε)-constant bigraphon $U:\Omega_1\times\Omega_2\to[0,1]$ then for $G\sim\mathbb{B}(k,U)$ we have a.a.s. that*

$$\omega_2(G)\geqslant(1-\alpha)\cdot\frac{2}{\log 1/d}\cdot\log k\;.$$

Proof. Let $\alpha\in(0,1)$ be arbitrary. Suppose that $\varepsilon>0$ is sufficiently small (we will precise this later), $d\in(d_1,d_2)$ and $U:\Omega_1\times\Omega_2\to[0,1]$ is (d,ε)-constant. Suppose further that k is large.

Let X_k be the number of bicliques in $\mathbb{B}(k,U)$ whose two colour classes have size $\ell=(1-\alpha)\cdot\frac{2}{\log 1/d}\cdot\log k$. Multiplicities caused by automorphisms of $K_{\ell,\ell}$ are not counted. By Proposition 4.2 we have

$$\mathbf{E}[X_k]=\binom{k}{\ell}^2\cdot t_{\mathrm{B}}(K_{\ell,\ell},U)\geqslant\binom{k}{\ell}^2 d^{\ell^2}\geqslant\left(\frac{k}{\ell}\right)^{2\ell}d^{\ell^2}=\left(\frac{k^{2\alpha}}{\ell^2}\right)^{\ell}\;. \tag{13}$$

Next, we are going to show by a second moment argument that $X_k\approx\mathbf{E}[X_k]$ a.a.s. For $p,q=0,1,\dots,\ell$, we define the bigraph $K_{[\ell,p,q]}$ as a result of gluing two copies of $K_{\ell,\ell}$ along p vertices in the first colour class and q vertices in the second colour class. Alternatively, $K_{[\ell,p,q]}$ can be obtained by erasing edges of two disjoint copies of the bigraph $K_{\ell-p,\ell-q}$ from $K_{2\ell-p,2\ell-q}$. We have

$$e(K_{[\ell,p,q]})=2\ell^2-pq\;. \tag{14}$$

We have

$$\mathbf{E}[X_k^2]=\sum_{p=0}^{\ell}\sum_{q=0}^{\ell}\mathbf{E}[Y_{p,q}]\;, \tag{15}$$

where $Y_{p,q}$ counts the number of bigraphs $K_{[\ell,p,q]}$ which preserve the order of the colour classes. We expand the second moment as

$$\mathbf{E}[Y_{p,q}]=\binom{k}{\ell-p\,|\,\ell-p\,|\,p}\cdot\binom{k}{\ell-q\,|\,\ell-q\,|\,q}\cdot t_{\mathrm{B}}(K_{[\ell,p,q]},U)\;. \tag{16}$$

Claim 4.4.1. *For every $c>0$ there exists $\varepsilon_1>0$ such that the following holds: Whenever $d\in(d_1,d_2)$ and $U:\Omega_1\times\Omega_2\to[0,1]$ is (d,ε_1)-constant then for each $p,q\in[\ell]_0$, $p+q\geqslant c\log k$, we have*

$$\mathbf{E}[X_k]^2\geqslant\log^3 k\cdot\mathbf{E}[Y_{p,q}]\;.$$

Proof of Claim 4.4.1. Let $c>0$ be arbitrary. Suppose that $\varepsilon_1>0$ is sufficiently small, $d\in(d_1,d_2)$ and $U:\Omega_1\times\Omega_2\to[0,1]$ is (d,ε_1)-constant. Let $p,q\in[\ell]_0$ be such that $p+q\geqslant c\log k$ (and of course

$p+q \leqslant 2\ell$). Upper-bounding the terms in (16) (while bearing in mind that $\ell = \Theta(\log k)$), we get

$$
\begin{aligned}
\mathbf{E}[Y_{p,q}] &\leqslant k^{4\ell-p-q}(d+\varepsilon_1)^{e(K_{[\ell,p,q]})} \overset{(14)}{=} k^{4\ell-p-q}(d+\varepsilon_1)^{2\ell^2-pq} \\
\boxed{\text{AM-GM Ineq., } d+\varepsilon_1<1} \quad &\leqslant \left(k^2(d+\varepsilon_1)^{\ell}\right)^{2\ell}\left(k(d+\varepsilon_1)^{\frac{p+q}{4}}\right)^{-(p+q)} \\
\boxed{p+q\leqslant 2\ell \text{ and } d+\varepsilon_1<1} \quad &\leqslant \left(k^2(d+\varepsilon_1)^{\ell}\right)^{2\ell}\left(k(d+\varepsilon_1)^{\frac{\ell}{2}}\right)^{-(p+q)} \\
\boxed{k(d+\varepsilon_1)^{\frac{\ell}{2}} \geqslant k^{\alpha}} \quad &\leqslant \exp\left(4\ell\log k\left(1+(1-\alpha)\tfrac{\log(d+\varepsilon_1)}{\log 1/d}\right)\right)k^{-c\alpha\log k} \\
&= k^{4\alpha\ell}\exp\left(4\ell\log k\,(1-\alpha)\left(1-\tfrac{\log(d+\varepsilon_1)}{\log d}\right)-c\alpha\log^2 k\right) \\
\boxed{\varepsilon_1 \text{ sufficiently small}} \quad &\leqslant k^{4\alpha\ell}\exp\left(-\tfrac{1}{2}c\alpha\log^2 k\right)
\end{aligned}
\tag{17}
$$

It is now enough to compare this with (13). □

Claim 4.4.2. *There exist numbers $C, \varepsilon_2 > 0$ such that the following holds: Whenever $d \in (d_1, d_2)$ and $U : \Omega_1 \times \Omega_2 \to [0,1]$ is (d, ε_2)-constant then for each $p, q \in [\ell]_0$, $1 \leqslant p+q < C\log k$, we have*

$$\mathbf{E}[Y_{0,0}] \geqslant k^{\frac{1}{2}}\mathbf{E}[Y_{p,q}]\ .$$

Proof of Claim 4.4.2. Suppose that $\varepsilon_2 > 0$ is sufficiently small, $d \in (d_1, d_2)$ and $U : \Omega_1 \times \Omega_2 \to [0,1]$ is (d, ε_2)-constant. Let us compare the combinatorial coefficients corresponding to $\mathbf{E}[Y_{0,0}]$ and $\mathbf{E}[Y_{p,q}]$ in (16). We have

$$\frac{\binom{k}{\ell\,|\,\ell}^2}{\binom{k}{\ell-p\,|\,\ell-p\,|\,p}\cdot\binom{k}{\ell-q\,|\,\ell-q\,|\,q}} = k^{(1+o(1))(p+q)}\ , \tag{18}$$

where $o(1) \to 0$ as $k \to \infty$ uniformly for any choice of p and q.

It remains to compare the terms $t_{\mathrm{B}}(K_{[\ell,0,0]}, U)$ and $t_{\mathrm{B}}(K_{[\ell,p,q]}, U)$.

First, we claim that for each $i, j, h \in \mathbb{N}$ we have

$$t_{\mathrm{B}}(K_{i,h}, U)t_{\mathrm{B}}(K_{j,h}, U) \leqslant t_{\mathrm{B}}(K_{i+j,h}, U)\ . \tag{19}$$

Indeed, by Hölder's inequality, we have

$$
\begin{aligned}
t_{\mathrm{B}}(K_{i,h}, U) &= \int_T \left(\int_x \deg(x,T)\right)^i \leqslant \left(\int_T \left(\int_x \deg(x,T)\right)^{i+j}\right)^{\frac{i}{i+j}}\ , \\
t_{\mathrm{B}}(K_{j,h}, U) &= \int_T \left(\int_x \deg(x,T)\right)^j \leqslant \left(\int_T \left(\int_x \deg(x,T)\right)^{i+j}\right)^{\frac{j}{i+j}}\ ,
\end{aligned}
$$

where the integrations are over $T = (t_1, \ldots, t_h) \in (\Omega_2)^h$, and $x \in \Omega_1$, and $\deg(x,T) = \prod_{r=1}^h U(x, t_r)$. Thus

$$t_{\mathrm{B}}(K_{i,h}, U)t_{\mathrm{B}}(K_{j,h}, U) \leqslant \int_T \left(\int_x \deg(x,T)\right)^{i+j} = t_{\mathrm{B}}(K_{i+j,h}, U)\ ,$$

as we wanted.

A double application of (19) followed by an application of Proposition 4.2 gives

$$
\begin{aligned}
t_{\mathrm{B}}(K_{\ell,\ell}, U) &\geqslant t_{\mathrm{B}}(K_{\ell-p,\ell}, U)t_{\mathrm{B}}(K_{p,\ell}, U) \geqslant t_{\mathrm{B}}(K_{\ell-p,\ell-q}, U)t_{\mathrm{B}}(K_{\ell-p,q}, U)t_{\mathrm{B}}(K_{p,\ell}, U) \\
&\geqslant t_{\mathrm{B}}(K_{\ell-p,\ell-q}, U)d^{q\ell-pq}d^{p\ell}\ .
\end{aligned}
\tag{20}
$$

In the defining formula (11) for $t_{\mathrm{B}}(K_{[\ell,p,q]}, U)$ we upper-bound some factors in $\prod U(x_i, y_i)$ by $d+\varepsilon_2$, as in Figure 1. Observe that after removal of the $p\ell + q\ell - 2pq$ edges indicated in Figure 1, the graph $K_{[\ell,p,q]}$ decomposes into a disjoint union of $K_{\ell,\ell}$ and $K_{\ell-p,\ell-q}$. Note that for a disjoint union $H_1 \oplus H_2$ of two bigraphs H_1 and H_2 we have $t_{\mathrm{B}}(H_1 \oplus H_2, U) = t_{\mathrm{B}}(H_1, U)t_{\mathrm{B}}(H_2, U)$. Thus,

$$t_{\mathrm{B}}(K_{[\ell,p,q]}, U) \leqslant (d+\varepsilon_2)^{p\ell+q\ell-2pq}t_{\mathrm{B}}(K_{\ell,\ell}, U)t_{\mathrm{B}}(K_{\ell-p,\ell-q}, U). \tag{21}$$

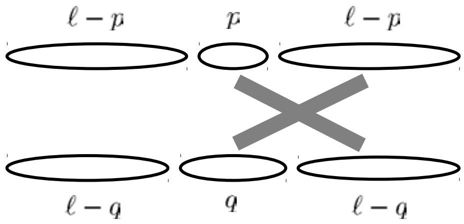

FIGURE 1. The $p\ell + q\ell - 2pq$ edges of $K_{[\ell,p,q]}$ corresponding to factors upper-bounded by $d + \varepsilon_2$ in (11).

Therefore,

$$\begin{aligned}
t_{\mathrm{B}}(K_{[\ell,0,0]},U) &= t_{\mathrm{B}}(K_{\ell,\ell},U)t_{\mathrm{B}}(K_{\ell,\ell},U) \overset{(20)}{\geqslant} t_{\mathrm{B}}(K_{\ell,\ell},U)t_{\mathrm{B}}(K_{\ell-p,\ell-q},U)d^{p\ell+q\ell-pq} \\
&\overset{(21)}{\geqslant} t_{\mathrm{B}}(K_{[\ell,p,q]},U)\left(\frac{d}{d+\varepsilon_2}\right)^{p\ell+q\ell}\left(\frac{(d+\varepsilon_2)^2}{d}\right)^{pq} \\
\boxed{\text{AM-GM Ineq., } \varepsilon_2 \text{ such that } \tfrac{(d+\varepsilon_2)^2}{d} \leqslant 1} \quad &\geqslant t_{\mathrm{B}}(K_{[\ell,p,q]},U)\left(1-\frac{\varepsilon_2}{d+\varepsilon_2}\right)^{p\ell+q\ell}\left(\frac{(d+\varepsilon_2)^2}{d}\right)^{(p+q)\cdot\frac{p+q}{4}} \\
&\geqslant t_{\mathrm{B}}(K_{[\ell,p,q]},U)\left(1-\frac{\varepsilon_2}{d+\varepsilon_2}\right)^{p\ell+q\ell} d^{(p+q)\cdot\frac{C\log k}{4}} .
\end{aligned} \tag{22}$$

Substituting (18) and (22) into (16) we get

$$\begin{aligned}
\frac{\mathbf{E}[Y_{0,0}]}{\mathbf{E}[Y_{p,q}]} &\geqslant \left(k^{1+o(1)}\left(1-\frac{\varepsilon_2}{d+\varepsilon_2}\right)^{\ell} d^{\frac{C\log k}{4}}\right)^{p+q} \\
\boxed{\text{for } \zeta \in (0,\tfrac{1}{2}) \text{ we have } 1-\zeta \geqslant \exp(-2\zeta)} \quad &\geqslant \left(k^{1+o(1)}\exp\left(-(1-\alpha)\frac{2\varepsilon_2}{d+\varepsilon_2}\cdot\frac{2}{\log 1/d}\cdot\log k\right) d^{\frac{C\log k}{4}}\right)^{p+q} \\
&= \left(k^{1+o(1)}\cdot k^{(1-\alpha)\frac{2\varepsilon_2}{d+\varepsilon_2}\cdot\frac{2}{\log d}}\cdot k^{\frac{C\log d}{4}}\right)^{p+q} \\
&\geqslant k^{\frac{1}{2}(p+q)} \geqslant k^{\frac{1}{2}} ,
\end{aligned}$$

for $C = 1/\log(1/d_1)$, and $\varepsilon_2 < \frac{1}{20}d_1\log(1/d_2)$. □

Let $C > 0$ and $\varepsilon_2 > 0$ be given by Claim 4.4.2. Let $\varepsilon_1 > 0$ be given by Claim 4.4.1 for $c = C$. Now if $\varepsilon < \min(\varepsilon_1, \varepsilon_2)$, we are ready to upper-bound the summands in (15). Note that we have $\Theta(\log^2 k)$ many of these summands. We get

$$\begin{aligned}
\mathbf{E}[X_k^2] &= \mathbf{E}[Y_{0,0}] + \sum_{\substack{p,q\in[l]_0 \\ 1\leqslant p+q<C\log k}} \mathbf{E}[Y_{p,q}] + \sum_{\substack{p,q\in[l]_0 \\ p+q\geqslant C\log k}} \mathbf{E}[Y_{p,q}] \\
&\leqslant \mathbf{E}[Y_{0,0}] + \Theta(\log^2 k)\mathbf{E}[Y_{0,0}]k^{-\frac{1}{2}} + \Theta(\log^2 k)\mathbf{E}[X_k]^2\log^{-3} k \\
\boxed{\mathbf{E}[Y_{0,0}] = (1+o(1))\mathbf{E}[X_k]^2} \quad &= (1+o(1))\mathbf{E}[X_k]^2.
\end{aligned}$$

Therefore we have $\mathbf{E}[X_k^2] = (1+o(1))\mathbf{E}[X_k]^2$, and it follows by Chebyshev's inequality that $\mathbf{P}[X_k > 0] = 1 - o(1)$. □

4.3. **Proof of Theorem 4.1.** The upper bound is easy, since it claims that $\omega_2(\mathbb{B}(n,U))$ is typically not bigger than the biclique number in the balanced bipartite Erdős–Rényi random graph $\mathbb{B}(n,p)$ (which clearly stochastically dominates $\mathbb{B}(n,U)$). For completeness, we include the calculations. We write $Y_k(\mathbb{B}(n,U))$ for the number of complete balanced bipartite graphs on $k+k$ vertices inside $\mathbb{B}(n,U)$. For $k = (1+\varepsilon)\cdot\frac{2}{\log 1/p}\cdot\log n$, we have,

$$\mathbf{E}[Y_k] \leqslant \binom{n}{k}\cdot\binom{n}{k}\cdot p^{k^2} \leqslant n^{2k}p^{k^2} = (n^2p^k)^k . \tag{23}$$

The statement now follows from Markov's Inequality, provided that we can show that $n^2p^k \to 0$. Indeed,

$$n^2p^k = n^2p^{\frac{2\log n}{\log 1/p}}p^{\frac{2\varepsilon\log n}{\log 1/p}} = p^{\frac{2\varepsilon\log n}{\log 1/p}} \to 0 \ .$$

We now turn to the lower bound. Let $\alpha \in (0,1)$ be arbitrary and let $\varepsilon_0 > 0$ be given by Proposition 4.4 for $d_1 = \frac{p}{2}$ and $d_2 = p$. Let $\varepsilon < \min\left(\varepsilon_0, \frac{p}{2}\right)$ be arbitrary. We denote by ν_i the measure given on Ω_i, $i = 1,2$. The definition of the essential supremum, together with Theorem 2.4, gives that there exist two measurable sets $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$ such that $\nu_1(A), \nu_2(B) > 0$ and $\int_x\int_y U(x,y) \geqslant (p-\varepsilon)\nu_1(A)\nu_2(B)$. We put $\delta = \min(\nu_1(A), \nu_2(B))$. By rescaling the measures ν_1, ν_2, we get probability measures ν_1^* on A and ν_2^* on B. Then we can view $U\restriction_{A\times B}$ as a bigraphon, which we denote by U^*. Note that U^* is $(p-\varepsilon,\varepsilon)$-constant (and thus also $(p-\varepsilon,\varepsilon_0)$-constant).

Consider now the sampling process to generate $B \sim \mathbb{B}(n,U)$ as described above. A standard concentration argument gives that with high probability, at least $\frac{\delta n}{2}$ points x_i sampled in Ω_1 lie in the set A, and at least $\frac{\delta n}{2}$ points y_j sampled in Ω_2 lie in the set B. In other words, with high probability we can find a copy of $\mathbb{B}(\frac{\delta n}{2}, U^*)$ inside $\mathbb{B}(n,U)$. Looking at the biclique number, we get that for $\ell = (1-\alpha)\cdot\frac{2}{\log 1/(p-\varepsilon)}\cdot\log(\delta n/2)$, we have

$$\mathbf{P}[\omega_2(\mathbb{B}(n,U)) \geqslant \ell] \geqslant \mathbf{P}\left[\omega_2(\mathbb{B}(\tfrac{\delta n}{2}, U^*)) \geqslant \ell\right] - o(1) \geqslant 1 - o(1) \ ,$$

where the last inequality follows from Proposition 4.4. Since $\log(\delta n/2) = (1+o(1))\log n$, and since $\alpha \in (0,1)$ and $\varepsilon < \min\left(\varepsilon_0, \frac{p}{2}\right)$ were arbitrary, the claim follows.

5. Formula for graphs with logarithmic clique number

In this section, we present the main result of the paper, Theorem 5.1. We show that if W is a graphon whose essential infimum is strictly positive then $\omega(\mathbb{G}(n,W)) \approx \kappa(W)\cdot\log n$, where $\kappa(W)$ is defined by

$$\kappa(W) = \sup\left\{\frac{2\|h\|_1^2}{\int_x\int_y h(x)h(y)\log\left({}^1/_{W(x,y)}\right)} \ : \ h \text{ is a histogram on } \Omega\right\} \ . \tag{24}$$

Here, we set $\frac{0}{0} = 0$ and $\frac{a}{0} = +\infty$ for $a \in \mathbb{R}\setminus\{0\}$.

Let us state the main result formally.

Theorem 5.1. *Suppose that W is a graphon whose essential infimum is strictly positive. Then*

- *if $\kappa(W) < +\infty$ then a.a.s. $\omega(\mathbb{G}(n,W)) = (1+o(1))\cdot\kappa(W)\cdot\log n$, and*
- *if $\kappa(W) = +\infty$ then a.a.s. $\omega(\mathbb{G}(n,W)) \gg \log n$.*

The proof of Theorem 5.1 is given in Section 7. In the remainder of this section we try to justify informally Theorem 5.1. While we believe that the derivation here captures the essence of the problem, the actual proof, presented in Section 7, is quite different. At the end of this section we comment on what fails in turning the heuristics into a rigorous argument.

We express $\kappa(W)$ in a way which will be more useful for our heuristics.

Fact 5.2. *We have*

$$\kappa(W) = \sup\left\{\|f\|_1 \ : \ f \text{ histogram on } \Omega, \ \Gamma(f,W) \geqslant 0\right\} \ , \tag{25}$$

where

$$\Gamma(f,W) = \underbrace{\int_x f(x)}_{(*)} + \underbrace{\frac{1}{2}\int_x\int_y f(x)f(y)\log W(x,y)}_{(**)} \ . \tag{26}$$

Proof. If there exists a nonzero histogram h such that $\int_u\int_v h(u)h(v)\log W(u,v) = 0$ then clearly both suprema in (24) and (25) are $+\infty$.

Let h be an arbitrary histogram in (24) and suppose that $\int_u\int_v h(u)h(v)\log\left({}^1/_{W(u,v)}\right) > 0$. We take $f = 2\|h\|_1\cdot h/\left(\int_u\int_v h(u)h(v)\log\left({}^1/_{W(u,v)}\right)\right)$. This way, the function f is a histogram on Ω, and $\|f\|_1$ equals to the term in the supremum in (24). So, to justify that the right-hand side of (25) is at least as big as that of (24), we only need to show that $\Gamma(f,W) \geqslant 0$. Indeed,

$$\Gamma(f,W) = \frac{2\|h\|_1}{\int\int h(u)h(v)\log({}^1/_{W(u,v)})}\cdot\int_x h(x) + \frac{1}{2}\cdot\left(\frac{2\|h\|_1}{\int\int h(u)h(v)\log({}^1/_{W(u,v)})}\right)^2\cdot\int_x\int_y h(x)h(y)\log W(x,y) = 0 \ .$$

On the other hand, let f be an arbitrary histogram appearing in (25) such that $\int_u \int_v f(u)f(v)\log\left(1/W(u,v)\right) > 0$. For $c \in \mathbb{R}$, let us denote by cf the c-multiple of the function f. Then the map $c \mapsto \Gamma(cf, W)$ is clearly a quadratic function with the limit $-\infty$ at $+\infty$. And since $\Gamma(f, W) \geqslant 0$, there is $c_0 \geqslant 1$ such that $\Gamma(c_0 f, W) = 0$. Now if we define $h = c_0 f$ then the corresponding term in (24) equals to $\|h\|_1 = c_0\|f\|_1 \geqslant \|f\|_1$. □

Motivated by (25), we say that a histogram f is *admissible* for a graphon W if $\Gamma(f, W) \geqslant 0$.

5.1. **First moment for a 2-step graphon.** Let us try to gain some intuition on Theorem 5.1 by looking at one of the simplest non-constant graphons. Let $W : \Omega^2 \to [0, 1]$ be represented on the unit interval Ω and defined by

$$W(x,y) = \begin{cases} p_{11} & \text{if } x, y \in \Omega_1, \\ p_{22} & \text{if } x, y \in \Omega_2, \text{ and} \\ p_{12} & \text{otherwise.} \end{cases} \tag{27}$$

Here, $p_{11}, p_{22}, p_{12} \in (0, 1)$ are arbitrary, and Ω is partitioned arbitrarily into two measurable sets Ω_1 and Ω_2 of positive measures β_1 and β_2.

Our aim is to determine for which numbers $c \in \mathbb{R}^+$ there typically exists a clique of order $c \log n$ in $G \sim \mathbb{G}(n, W)$, and for which c's there is typically none. So let us fix $c \in \mathbb{R}^+$ and let X_n count the number of cliques of order $c \log n$ in $G \sim \mathbb{G}(n, W)$. By Markov's Inequality, $\omega(G)$ will be typically smaller than $c \log n$ in the regime when $\mathbf{E}[X_n] \to 0$. On the other hand, it is plausible that a second moment argument will give that typically $\omega(G) \geqslant c \log n$ when $\mathbf{E}[X_n] \to +\infty$. With this belief — which is supported by the success of a second-moment argument in the proof of Theorem 4.1 — let us estimate $\mathbf{E}[X_n]$. Actually, we rather look at a refined quantity $Y_n^{\alpha_1,\alpha_2}(G)$ which is defined as the number of cliques in G that consist of $\alpha_1 \log n$ vertices whose representation on Ω lies in Ω_1, and $\alpha_2 \log n$ vertices that are represented in Ω_2. We have

$$\mathbf{E}[X_n] = \sum_{m=0}^{c\log n} \mathbf{E}\left[Y_n^{m/\log n, c-m/\log n}\right] .$$

We expect the quantities $Y_n^{\alpha_1,\alpha_2}$ to be either super-polynomially small or super-polynomially large. Since the sum above has only a $\Theta(\log n)$-many summands, we expect that

$$\mathbf{E}[X_n] \to +\infty \quad \text{if and only if} \quad \exists \alpha_1, \alpha_2 \geqslant 0 \text{ such that } \alpha_1 + \alpha_2 = c \text{ and } \mathbf{E}[Y_n^{\alpha_1,\alpha_2}] \to +\infty . \tag{28}$$

For a clique that contributes to $Y_n^{\alpha_1,\alpha_2}(G)$ to be present, $\binom{\alpha_1 \log n}{2}$ edges in the $(\Omega_1 \times \Omega_1)$-part of W, $\binom{\alpha_2 \log n}{2}$ edges in the $(\Omega_2 \times \Omega_2)$-part, and $\alpha_1\alpha_2 \log^2 n$ edges in the $(\Omega_1 \times \Omega_2)$-part must be present in the specific locations of perspective cliques or complete bipartite graphs. By the Law of Large Numbers, approximately $\beta_1 n$ points in the sampling process for $\mathbb{G}(n, W)$ end up in Ω_1 and approximately $\beta_2 n$ points end up in Ω_2. We get

$$\begin{aligned} \mathbf{E}[Y_n^{\alpha_1,\alpha_2}] &\approx \binom{\beta_1 n}{\alpha_1 \log n}\binom{\beta_2 n}{\alpha_2 \log n} p_{11}^{\binom{\alpha_1 \log n}{2}} p_{22}^{\binom{\alpha_2 \log n}{2}} p_{12}^{\alpha_1\alpha_2 \log^2 n} \\ &= \exp\left((1+o(1))\log^2 n \left(\alpha_1 + \alpha_2 + \frac{\alpha_1^2}{2}\log p_{11} + \frac{\alpha_2^2}{2}\log p_{22} + \alpha_1\alpha_2 \log p_{12}\right)\right) . \end{aligned} \tag{29}$$

Thus, whether $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}] \to 0$ or $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}] \to +\infty$ depends on whether

$$\underbrace{\alpha_1 + \alpha_2}_{(*)} + \underbrace{\frac{1}{2}\left(\alpha_1^2 \log p_{11} + \alpha_2^2 \log p_{22} + \alpha_1\alpha_2 \log p_{12} + \alpha_2\alpha_1 \log p_{12}\right)}_{(**)} \tag{30}$$

is negative or positive, respectively. It is straightforward to generalize this formula to graphons with more steps. Observe also that the values of β_1 and β_2 get lost in the transition between the first and the second line of (29), and are immaterial in the final (30) consequently (provided that they are positive). In particular, the step sizes β_i could have been "infinitesimally small". Thus, we can see a direct correspondence between $(*)$ and $(**)$ in (26) and (30), where the integration corresponds to passing to infinitesimal steps. In view of this, we will denote the term in (30) by $\Gamma(\alpha, W)$, where $\alpha = (\alpha_1, \alpha_2)$. The optimization over α_1 and α_2 in (28) corresponds to taking the supremum in (25).

This is why we call the functions f in (25) (or vectors, in case of step-graphons) histograms: they specify the densities of the vertices of the anticipated cliques over the space Ω.

Last, let us note that a physicist might refer to $(*)$ as the "entropy contribution", as it comes from the choice of the vertices of a clique, while $(**)$ could be referred to as the "energy" needed to include all required edges of that clique.

5.2. **Introducing the second moment to the example.** So far, our prediction was based on a first moment argument. Combined with Markov's Inequality this gives readily an upper bound on the typical clique number of $\mathbb{G}(n, W)$. We now want to complement the upper bound with a lower bound based on a second moment argument. Let us first recall that situation in the setting of the Erdős–Rényi random graphs $\mathbb{G}(n, p)$. There, a straightforward calculation for the random variable X_n counting cliques of order $c \log n$ (where $c > 0$ is fixed) gives that $\mathbf{E}[X_n]^2 = (1 + o(1))\mathbf{E}[X_n^2]$ if and only if $\mathbf{E}[X_n] \to +\infty$. Thus, the first and the second moment start working together at the same time.

The situation is more complicated in the model $\mathbb{G}(n, W)$. We will illustrate this on the graphon W from (27). Suppose that $\alpha_1, \alpha_2 \geqslant 0$ are such that (30) is positive, and we ask in hope whether

$$\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]^2 = (1 + o(1))\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2] \tag{31}$$

In (29), we provided asymptotitcs for $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]$. Thus, to understand whether we have (31), we need to calculate $\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2]$. We have

$$\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2] = \mathbf{E}\left[\left(\sum_K \mathbf{1}_{K \text{ induces a clique}}\right)^2\right] = \sum_{K,L} \mathbf{P}[K \text{ and } L \text{ induce cliques}] \,, \tag{32}$$

where K and L range over all sets of vertices with $\alpha_1 \log n$ vertices represented in Ω_1 and $\alpha_2 \log n$ vertices represented in Ω_2. Say, that K_1 are the vertices of K represented in Ω_1, and K_2, L_1, and L_2 are defined analogously. It is clear that $|K_1 \setminus L_1| = |L_1 \setminus K_1|$ and $|K_2 \setminus L_2| = |L_2 \setminus K_2|$. So for fixed sets K and L, we have

$$\mathbf{P}[K \text{ and } L \text{ induce cliques}] = p_{11}^{\binom{|K_1 \cup L_1|}{2} - |K_1 \setminus L_1|^2} \cdot p_{22}^{\binom{|K_2 \cup L_2|}{2} - |K_2 \setminus L_2|^2} \cdot p_{12}^{|K_1 \cup L_1| \cdot |K_2 \cup L_2| - 2|K_1 \setminus L_1||K_2 \setminus L_2|} \,.$$

Thus, grouping (32) depending on the values of $m_1 = |K_1 \cap L_1|$ and $m_2 = |K_2 \cap L_2|$ we get

$$\begin{aligned}\mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2] \approx \sum_{m_1=0}^{\alpha_1 \log n} \sum_{m_2=0}^{\alpha_2 \log n} & \binom{\beta_1 n}{m_1 \mid \alpha_1 \log n - m_1 \mid \alpha_1 \log n - m_1} \\ & \times \binom{\beta_2 n}{m_2 \mid \alpha_2 \log n - m_2 \mid \alpha_2 \log n - m_2} \\ & \times p_{11}^{\binom{2\alpha_1 \log n - m_1}{2} - (\alpha_1 \log n - m_1)^2} \cdot p_{22}^{\binom{2\alpha_2 \log n - m_2}{2} - (\alpha_2 \log n - m_2)^2} \\ & \times p_{12}^{(2\alpha_1 \log n - m_1)\cdot(2\alpha_2 \log n - m_2) - 2(\alpha_1 \log n - m_1)(\alpha_2 \log n - m_2)} \,,\end{aligned} \tag{33}$$

where the approximate equality represents the fact that we assumed on the right-hand side that exactly $\beta_i n$ vertices are represented in Ω_i. Let us write $\gamma_i = m_i / \log n$, and let us write $z_n^{\gamma_1,\gamma_2}$ for the individual summands on the right-hand side of (33). Routine manipulations give that

$$\begin{aligned}\frac{\log z_n^{\gamma_1,\gamma_2}}{\log^2 n} &\approx \gamma_1 + \gamma_2 + 2(\alpha_1 - \gamma_1) + 2(\alpha_2 - \gamma_2) \\ &\quad + (\alpha_1^2 - \tfrac{1}{2}\gamma_1^2) \log p_{11} + (\alpha_2^2 - \tfrac{1}{2}\gamma_2^2) \log p_{22} + (2\alpha_1\alpha_2 - \gamma_1\gamma_2) \log p_{12} \,.\end{aligned}$$

Thus, if we want the second moment (31) to work then comparing the calculations above with (29), we must have for each $\gamma_1 \in [0, \alpha_1]$ and $\gamma_2 \in [0, \alpha_2]$ that

$$\begin{aligned}&2\left(\alpha_1 + \alpha_2 + \frac{\alpha_1^2}{2} \log p_{11} + \frac{\alpha_2^2}{2} \log p_{22} + \alpha_1\alpha_2 \log p_{12}\right) \\ &\gtrapprox \gamma_1 + \gamma_2 + 2(\alpha_1 - \gamma_1) + 2(\alpha_2 - \gamma_2) + (\alpha_1^2 - \tfrac{1}{2}\gamma_1^2) \log p_{11} + (\alpha_2^2 - \tfrac{1}{2}\gamma_2^2) \log p_{22} + (2\alpha_1\alpha_2 - \gamma_1\gamma_2) \log p_{12} \,,\end{aligned}$$

which rewrites as $\Gamma(\gamma, W) \gtrapprox 0$. To summarize, to justify (25) (at least for step-functions), it would suffice to have the following.

Dream Lemma. *If W is a step-graphon with k steps and $\alpha \in [0,+\infty)^k$ is a vector with $\Gamma(\alpha, W) > 0$, then for all $\gamma \in \mathrm{Box}(\alpha)$ we have $\Gamma(\gamma, W) \geqslant 0$.*

This, however, does not hold in general. Indeed, take for example

$$p_{11} = e^{-3}, p_{12} = p_{22} = e^{-\frac{1}{4}}, \quad \alpha_1 = 1, \alpha_2 = 1, \quad \gamma_1 = 1, \gamma_2 = 0.$$

We have $\Gamma(\alpha, W) = \frac{1}{8} > 0$ but $\Gamma(\gamma, W) = -\frac{1}{2} < 0$. It is worth explaining what is happening in the example in words. The parameters p_{11} and α_1 are set so that asymptotically almost surely, $\mathbb{G}(\frac{n}{2}, p_{11})$ contains no cliques of order $\alpha_1 \log n$. However, in the rare cases when $\mathbb{G}(\frac{n}{2}, p_{11})$ (viewed as a subgraph of $\mathbb{G}(n, W)$) does contain such a clique, there are typically many ways of extending it on the Ω_2-part by $\alpha_2 \log n$ additional vertices, thus inflating $\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]$ substantially.

The above suggests a correction for (25) in that we should range only over those histograms f for which $\Gamma(g, W) \geqslant 0$ for all histograms $g \in \mathrm{Box}(f)$. Note also that the *necessity* of testing the admissibility condition over all sub-histograms of f has a clear combinatorial interpretation: If cliques with a given histogram typically appear, then for each given sub-histogram cliques with that sub-histogram must appear (just because a subset of a clique again induces a clique).

Now, after all the arguing why (25) should seem wrong, let us explain why it is actually all right. We show in Lemma 6.3 that for any histogram f attaining the supremum in (25) we automatically have that all sub-histograms are admissible (recall that a histogram h is admissible for a graphon W if $\Gamma(h, W) \geqslant 0$). If the supremum is not attained then for each histogram f almost attaining the supremum, we have $\Gamma(g, W) \gtrapprox 0$ for all sub-histograms g, which is sufficient for the argument.

5.3. **Failure of turning the above heuristics into a rigorous argument.** There are two types of errors that we introduced in the above argument. Firstly, the "little imprecisions" when we replaced a sum by its maximal term (such as in (28)) or when we used the $\lessapprox$-symbol. Each such step introduces an error of $o(1)$ to the quantities $\frac{\log X_n}{\log^2 n}$ and $\frac{\log Y_n^{\alpha_1,\alpha_2}}{\log^2 n}$. That means, that in actuality we can only conclude that

$$\mathbf{E}[Y_n^{\alpha_1,\alpha_2}]^2 = \exp\left(o(\log^2 n)\right) \mathbf{E}[(Y_n^{\alpha_1,\alpha_2})^2] ,$$

which is too crude for the second moment argument to work.

Secondly, the notion of a "set of vertices following a certain histogram" makes sense only in the stochastic block model, but not when we have a finite set of vertices in an uncountable probability space. Let us jump ahead and note that in the rigorous proof in Section 7 we, in a sense, are able to make use of histograms in the continuous setting. Namely, Lemma 2.2 allows us to discretize a given graphon in an appropriate sense, after which it does make sense to talk about histograms.

Let us remark that for the stochastic block model the first issue (which is the only in that case) can be dealt with by pedestrian calculations, thus yielding a routine proof of Theorem 5.1 for the special class of stochastic block models.

6. Tools for the proof of Theorem 5.1

In this section we prepare tools for lower-bound in Theorem 5.1. In Section 6.1 we provide alternative expressions for $\kappa(W)$. In Section 6.2 we state and prove Lemma 6.3 which was motivated in Section 5.2. In Section 6.3 we introduce a new graphon parameter $\xi(W)$. This parameter is motivated by controlling the second moment of the number of cliques of a given size. All the work in Section 6 steers towards deriving the two main results of this section, Lemma 6.8 and Lemma 6.10. The former lemma asserts that each graphon W contains a subgraphon U with $\xi(U) \approx \frac{1}{\kappa(W)}$. The latter asserts that $\omega(\mathbb{G}(n, U)) \gtrapprox \frac{1}{\xi(U)} \cdot \log n$. These two lemmas combine easily to give the proof of the lower bound in Theorem 5.1 (as is shown in Section 7).

6.1. **Alternative formulas for $\kappa(W)$.** In Fact 5.2 we gave an alternative formula for $\kappa(W)$. In Proposition 6.1 we give two further expressions. These expressions require the graphon W to be changed on a nullset. Note that we have the liberty of making such a change as the distribution of the model $\mathbb{G}(n, W)$ remains unaltered.

Proposition 6.1. *Suppose that $W:(0,1)^2\to[0,1]$ is an arbitrary graphon. Then W can be changed on a nullset in such a way that the following holds:*

We have

$$\exp\left(-\frac{2}{\kappa(W)}\right)=\lim_{r\to\infty}P_r\quad\textit{where}\quad P_r=\sup_{F\subseteq I,|F|=r}\left(\prod_{x,y\in F,x<y}W(x,y)\right)^{2/(|F|^2-|F|)}\quad\textit{and}\tag{34}$$

$$\frac{2}{\kappa(W)}=\inf_A\frac{1}{\lambda(A)^2}\int_{A\times A}\log {}^1\!/\!_W\ ,\tag{35}$$

where the supremum ranges over all measurable sets $A\subseteq(0,1)$ of positive measure.

Lastly,

$$\kappa(W)=\sup_{a\in\mathbb{R}_+,A\subseteq\Omega}\{a\lambda(A)\ :\ \Gamma(a\cdot\mathbf{1}_A,W)\geqslant 0\}\ .\tag{36}$$

More precisely, for each $\varepsilon>0$ and each set A from (35) satisfying $(1+\varepsilon)\frac{2}{\kappa(W)}\geqslant\frac{1}{\lambda(A)^2}\int_{A\times A}\log({}^1\!/\!_W)$ we have that for the number $a=(1-\varepsilon)\frac{\kappa(W)}{\lambda(A)}$ that $\Gamma(a\cdot\mathbf{1}_A,W)\geqslant 0$.

Proof. Let us replace the value of W in every point $(x,y)\in(0,1)^2$ that is not a point of approximate continuity by c. This is a change of measure zero by Theorem 2.5.

Let us deal with the first part of the statement postponing (36) to later.

Claim 6.1.1. *For each $r\in\{2,3,\ldots\}$ we have $\log P_r\geqslant-\frac{2}{\kappa(W)}$.*

Proof of Claim 6.1.1. Let h be an arbitrary function appearing in (24) (not constant zero). Fix $r\in\{2,3,\ldots\}$, and let $F\subseteq(0,1)$ be a random set consisting of r independent points sampled from $(0,1)$ according to the density function $d=h/\|h\|_1$. Then by linearity of expectation we have

$$\mathbf{E}_F\left[\tfrac{2}{r(r-1)}\sum_{x,y\in F,x<y}\log W(x,y)\right]=\mathbf{E}_{x,y\sim d}\left[\log W(x,y)\right]=\frac{1}{\|h\|_1^2}\int_x\int_y h(x)h(y)\log W(x,y)\ .$$

This shows that there exists a deterministic r-element set F for which

$$\frac{2}{r(r-1)}\sum_{x,y\in F,x<y}\log W(x,y)\geqslant\frac{1}{\|h\|_1^2}\int_x\int_y h(x)h(y)\log W(x,y)\ .$$

This concludes the proof of Claim 6.1.1. □

Let us denote the right-hand side of (35) as $-P$.

Claim 6.1.2. *For each $r\in\{2,3,\ldots\}$, we have that $P\geqslant\frac{r-1}{r}\log P_r+\frac{\log c}{r}$.*

Proof of Claim 6.1.2. Suppose that $r\in\{2,3,\ldots\}$ is given, and let $F=\{x_1<x_2<\ldots<x_r\}$ be an arbitrary set of points in $(0,1)$ as in (34). Let $\varepsilon\in(0,c/2)$ be arbitrary. Let $C=\log c-\log(c-\varepsilon)$. Firstly, note that the concavity of the logarithm gives that

$$\log(a-\varepsilon)\geqslant\log a-C\tag{37}$$

for each $a\in[c,\infty)$. Secondly, note that $C\searrow 0$ as $\varepsilon\searrow 0$.

Let us take $\delta>0$ such that for each $i\in[r]$, we have that the sets $S_i=[x_i-\delta,x_i+\delta]$ are pairwise disjoint, and such that the measure of each of the sets $D_{ij}=\{(x,y)\in S_i\times S_j: W(x_i,x_j)-\varepsilon>W(x,y)\}$ is at most $\varepsilon(2\delta)^2$. The latter property can be achieved since each point (x_i,x_j) is either a Lebesgue point of W, or it is a point attaining the infimum of W. Let $A = \bigcup_{i=1}^{r} S_i$. Then,

$$\begin{aligned}\frac{1}{\lambda(A)^2}\int_{A\times A}\log W &= \frac{1}{(2\delta r)^2}\sum_i\sum_j\left(\int_{(x,y)\in(S_i\times S_j)\setminus D_{ij}}\log W(x,y)+\int_{(x,y)\in D_{ij}}\log W(x,y)\right)\\ &\geqslant \frac{1}{(2\delta r)^2}\sum_i\sum_j\left(\int_{(x,y)\in(S_i\times S_j)\setminus D_{ij}}\log(W(x_i,x_j)-\varepsilon)+\int_{(x,y)\in D_{ij}}\log c\right)\\ \boxed{\text{(37) for 1}^{\text{st}}\text{ term; }\nu(D_{ij})\leqslant\varepsilon(2\delta)^2\text{ for 2}^{\text{nd}}\text{ term}}\quad &\geqslant \frac{1}{r^2}\sum_i\sum_j\log W(x_i,x_j)\ -C+\varepsilon\log c\\ &\geqslant \frac{r-1}{r}\cdot\frac{2}{r(r-1)}\sum_{i,j:i<j}\log W(x_i,x_j)\ +\frac{1}{r^2}\sum_i\log W(x_i,x_i)-C+\varepsilon\log c\\ &\geqslant \frac{r-1}{r}\cdot\frac{2}{r(r-1)}\sum_{i,j:i<j}\log W(x_i,x_j)\ -C+(\varepsilon+\tfrac{1}{r})\log c\,.\end{aligned}$$

Letting $\varepsilon \searrow 0$ (which means that also $C \searrow 0$), we get the claim. □

By Claim 6.1.1, we have $\liminf_{r\to\infty}\log P_r \geqslant -\frac{2}{\kappa(W)}$. By Claim 6.1.2, we have $P \geqslant \limsup_{r\to\infty}\log P_r$. Further, it is obvious that $-\frac{2}{\kappa(W)} \geqslant P$. Indeed, the supremum in (24) ranges over all histograms, of which indicators of measurable sets are just a particular case. The combination of the three above inequalities proves the fact.

So, it remains to deal with (36). Positive multiples of indicator functions are histograms, so (25) tells us that $\kappa(W) \geqslant \sup_{a\in\mathbb{R}_+,A\subseteq\Omega}\{a\lambda(A) : \Gamma(a\cdot\mathbf{1}_A, W)\geqslant 0\}$. It remains to deal with the opposite inequality. We shall prove this in the "more precisely" form. Let $\varepsilon > 0$ be arbitrary and take A such that $(1+\varepsilon)\frac{2}{\kappa(W)} \geqslant \frac{1}{\lambda(A)^2}\int_{A\times A}\log(1/W)$. Set $a = (1-\varepsilon)\frac{\kappa(W)}{\lambda(A)}$. We claim that the pair (a, A) is admissible for the supremum in (36). Indeed,

$$\begin{aligned}\Gamma(a\cdot\mathbf{1}_A, W) &= a\lambda(A) - \tfrac{1}{2}a^2\int_{A\times A} -\log W\\ &= (1-\varepsilon)\kappa(W)\left(1-\tfrac{1}{2}(1-\varepsilon)\kappa(W)\frac{1}{\lambda(A)^2}\int_{A\times A}\log(1/W)\right)\\ &\geqslant (1-\varepsilon)\kappa(W)\left(1-\tfrac{1}{2}(1-\varepsilon)\kappa(W)(1+\varepsilon)\frac{2}{\kappa(W)}\right) = \varepsilon^2(1-\varepsilon)\kappa(W) > 0\,.\end{aligned}$$

Since $a\cdot\lambda(A) = (1-\varepsilon)\kappa(W)$, and since $\varepsilon > 0$ was arbitrary, this finishes the proof. □

6.2. **Subhistograms of optimal histograms are admissible.** The main result of this section, Lemma 6.3 tells us that if f^* is a histogram almost attaining the supremum in (25) then $\Gamma(f, W)$ is almost positive for all subhistograms $f \leqslant f^*$. We showed that this particular case of the (false) Dream Lemma is needed for the second moment to work. The proof of Lemma 6.3 is technical, building on Lemma 6.2. It turns out that in those situations when the supremum in (25) is attained, Lemma 6.3 has a much shorter (but conceptually the same) proof. We offer this simplified statement in Lemma A.1 in the Appendix.

Lemma 6.2. *Suppose that W is an arbitrary graphon with $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$. Then there is a constant $K > 0$ depending only on the graphon W such that the following holds: Let g be an arbitrary histogram admissible for W and let $\delta \in (0,1)$. Suppose that $a \in (0,1)$ and that $g = g' + g''$ for some non-trivial histograms g' and g'' such that $\|g'\|_1 < \|g\|_1 - \delta$. Then either $\Gamma(g') \geqslant -a$, or there exists a histogram g^* which is admissible for W and for which we have*

$$\|g^*\|_1 \geqslant \|g\|_1 + K\delta^3 a^{\frac{5}{2}}\,. \tag{38}$$

Proof. Since we shall work exclusively with the graphon W, we write $\Gamma(\cdot)$ as a shortcut for $\Gamma(\cdot, W)$. Let us write $m^- = \operatorname{ess\,inf} W$ and $m^+ = \operatorname{ess\,sup} W$.

Let us fix numbers $a, \delta \in (0,1)$ and a decomposition $g = g' + g''$ of g into non-trivial histograms g', g'' such that $\|g'\|_1 < \|g\|_1 - \delta$. For $\varepsilon_1 \in (0,1)$ and $\varepsilon_2 > 0$, let us write $g^*(\varepsilon_1, \varepsilon_2) = (1-\varepsilon_1)g' + (1+\varepsilon_2)g''$. Let us also write

$$A = \|g'\|_1, \quad B = \|g''\|_1,$$

$$C=-\frac{1}{2}\int_{x,y} g'(x)g'(y)\log W(x,y),\quad D=-\frac{1}{2}\int_{x,y} g''(x)g''(y)\log W(x,y),\quad E=-\int_{x,y} g'(x)g''(y)\log W(x,y)$$

and note that $A,B,C,D,E>0$. We have $\Gamma(g')=A-C$ and $\Gamma(g)=A+B-C-D-E$. There is nothing to prove when $\Gamma(g')\geqslant -a$. Thus, assume otherwise. Then

$$A<C-a\,. \tag{39}$$

Upper-bounding C by $\frac{1}{2}\|g'\|_1^2\log(1/m^-)$ and using that $C>a$, we get

$$\|g'\|_1>\sqrt{\frac{2a}{\log\left({}^{1}/{}_{m^-}\right)}}\,. \tag{40}$$

For each $\varepsilon_1\in(0,1)$ and $\varepsilon_2>0$, the difference $\Gamma(g^*(\varepsilon_1,\varepsilon_2))-\Gamma(g)$ can be expressed as

$$\begin{aligned}(1-\varepsilon_1)A+(1+\varepsilon_2)B-(1-\varepsilon_1)^2C-(1+\varepsilon_2)^2D-(1-\varepsilon_1)(1+\varepsilon_2)E-(A+B-C-D-E)\\ =\varepsilon_1(-A+2C+E)+\varepsilon_2(B-2D-E)-\varepsilon_1^2C-\varepsilon_2^2D+\varepsilon_1\varepsilon_2E.\end{aligned}$$

In particular, if $\varepsilon_2=(1+\beta)\frac{A}{B}\varepsilon_1$ (where $\varepsilon_1\in(0,1)$ and $\beta>0$ will be determined later) then we have

$$\begin{aligned}&\Gamma\left(g^*(\varepsilon_1,(1+\beta)\tfrac{A}{B}\varepsilon_1)\right)-\Gamma(g)\\ =&\varepsilon_1\left(2C+E-\tfrac{2AD}{B}-\tfrac{AE}{B}\right)+\varepsilon_1\cdot\varepsilon_1\left(-C-\tfrac{A^2D}{B^2}+\tfrac{AE}{B}-2\tfrac{A^2}{B^2}\beta D+\beta\tfrac{AE}{B}-\tfrac{A^2}{B^2}\beta^2D\right)+\varepsilon_1\beta(A-\tfrac{2AD}{B}-\tfrac{AE}{B})\\ \geqslant&\varepsilon_1\underbrace{\left(2C+E-\tfrac{2AD}{B}-\tfrac{AE}{B}\right)}_{\mathtt{T1}}-\varepsilon_1\cdot\underbrace{\varepsilon_1\left(C+\tfrac{A^2D}{B^2}+2\tfrac{A^2}{B^2}\beta D+\tfrac{A^2}{B^2}\beta^2D\right)}_{\mathtt{T2}}-\varepsilon_1\underbrace{\beta(\tfrac{2AD}{B}+\tfrac{AE}{B})}_{\mathtt{T3}}\,.\end{aligned} \tag{41}$$

Let us expand the term `T1`.

$$\begin{aligned}2C+E-\frac{2AD}{B}-\frac{AE}{B}&\overset{(39)}{>}2A-\frac{2AD}{B}-\frac{AE}{B}+2a\\ &>2\frac{A}{B}(B-D-E)+2a\\ &>2\frac{A}{B}\left(B-D-E+(A-C)\right)+2a\\ &=2\frac{A}{B}\Gamma(g)+2a\geqslant 2a.\end{aligned}$$

Now, set $\varepsilon_1=\frac{a}{4}\min(\frac{1}{C},\frac{B^2}{2A^2D})$ and $\beta=\min(1,\frac{aB}{4AD},\frac{aB}{2AE})$. Routine calculations give that each of the terms `T2` and `T3` is smaller than a. Plugging the bounds above in (41), we get

$$\Gamma\left(g^*(\varepsilon_1,(1+\beta)\tfrac{A}{B}\varepsilon_1)\right)\geqslant\Gamma(g)\geqslant 0. \tag{42}$$

We have

$$C=\frac{1}{2}\int_{x,y} g'(x)g'(y)\log\left({}^{1}/{}_{W(x,y)}\right)\leqslant\frac{1}{2}\int_{x,y} g(x)g(y)\log\left({}^{1}/{}_{m^-}\right)\leqslant\frac{1}{2}\kappa(W)^2\log\frac{1}{m^-}$$

and similarly $D,E\leqslant\frac{1}{2}\kappa(W)^2\log\left({}^{1}/{}_{m^-}\right)$. We also have

$$\frac{B}{A}=\frac{\|g-g'\|_1}{\|g'\|_1}\geqslant\frac{\delta}{\|g\|_1-\delta}\geqslant\frac{\delta}{\kappa(W)-\delta}\geqslant\frac{\delta}{\kappa(W)}\,.$$

Therefore

$$\begin{aligned}\varepsilon_1\beta&\geqslant\min\left(\frac{a}{4C},\frac{aB^2}{8A^2D},\frac{a^2B}{16ACD},\frac{a^2B^3}{32A^3D^2},\frac{a^2B}{8ACE},\frac{a^2B^3}{16A^3DE}\right)\\ &\geqslant a^2\min\left(\frac{1}{4C},\frac{B^2}{8A^2D},\frac{B}{16ACD},\frac{B^3}{32A^3D^2},\frac{B}{8ACE},\frac{B^3}{16A^3DE}\right)\\ &\geqslant a^2\frac{1}{\log^2\left({}^{1}/{}_{m^-}\right)}\min\left(\frac{\log\left({}^{1}/{}_{m^-}\right)}{2\kappa(W)^2},\frac{\delta^2\log\left({}^{1}/{}_{m^-}\right)}{4\kappa(W)^4},\frac{\delta}{4\kappa(W)^5},\frac{\delta^3}{8\kappa(W)^7},\frac{\delta}{2\kappa(W)^5},\frac{\delta^3}{4\kappa(W)^7}\right)\\ &\geqslant a^2\delta^3\frac{1}{\log^2\left({}^{1}/{}_{m^-}\right)}\min\left(\frac{\log\left({}^{1}/{}_{m^-}\right)}{2\kappa(W)^2},\frac{\log\left({}^{1}/{}_{m^-}\right)}{4\kappa(W)^4},\frac{1}{4\kappa(W)^5},\frac{1}{8\kappa(W)^7},\frac{1}{2\kappa(W)^5},\frac{1}{4\kappa(W)^7}\right)\,.\end{aligned} \tag{43}$$

It follows that

$$\begin{aligned}\|g^*(\varepsilon_1,\varepsilon_2)\|_1 &= (1-\varepsilon_1)A + (1+(1+\beta)\tfrac{A}{B}\varepsilon_1)B = \|g\|_1 + \beta\varepsilon_1\|g'\|_1 \\ &\overset{(40),(43)}{>} \|g\|_1 + a^{\frac{5}{2}}\delta^3 \min\left(\frac{\log(1/m^-)}{2\kappa(W)^2}, \frac{\log(1/m^-)}{4\kappa(W)^4}, \frac{1}{4\kappa(W)^5}, \frac{1}{8\kappa(W)^7}, \frac{1}{2\kappa(W)^5}, \frac{1}{4\kappa(W)^7}\right)\sqrt{\frac{2}{\log^5(1/m^-)}}.\end{aligned}$$

This finishes the proof. □

Lemma 6.3. *Suppose that W is an arbitrary graphon with $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$. Then there exists a number $\gamma_0 > 0$ and a function $q : (0,\gamma_0) \to \mathbb{R}_+$ with $\lim_{\gamma\searrow 0} q(\gamma) = 0$ such that the following holds. Let f^* be an admissible histogram for W and let $\gamma \in (0,\gamma_0)$. Suppose that $\|f^*\|_1 \geqslant \kappa(W) - \gamma$. Then for every $f \in \mathrm{Box}(f^*)$ we have $\Gamma(f,W) \geqslant -q(\gamma)$.*

Proof. Let K be the constant from Lemma 6.2. Set γ_0 so that $\gamma_0 < K^2$. Suppose that $\gamma \in (0,\gamma_0)$ and f^* is an admissible histogram with $\|f^*\|_1 \geqslant \kappa(W) - \gamma$.

We write $\Gamma(\cdot)$ as a shortcut for $\Gamma(\cdot, W)$.

Let $f \in \mathrm{Box}(f^*)$ be non-trivial. Suppose first that $\|f\|_1 \geqslant \|f^*\|_1 - \sqrt[9]{\gamma}$. Using (26) we get

$$\begin{aligned}\Gamma(f) &= \int_x f(x) + \frac{1}{2}\int_x\int_y f(x)f(y)\log W(x,y) \\ &\geqslant \int_x f^*(x) - \sqrt[9]{\gamma} + \frac{1}{2}\int_x\int_y f^*(x)f^*(y)\log W(x,y) \\ &= \Gamma(f^*) - \sqrt[9]{\gamma} \geqslant -\sqrt[9]{\gamma}\,.\end{aligned} \tag{44}$$

Suppose next that $\|f\|_1 < \|f^*\|_1 - \sqrt[9]{\gamma}$. We apply Lemma 6.2 to

$$g = f^*, \quad g' = f, \quad g'' = f^* - f, \quad \delta = \sqrt[9]{\gamma}, \quad a = \frac{\sqrt[5]{\gamma}}{K^{\frac{2}{5}}}\,.$$

Then there is no histogram g^* admissible for W such that $\|g^*\|_1 \geqslant \|f^*\|_1 + K\delta^3 a^{\frac{5}{2}}$ as the right hand side equals $\|f^*\|_1 + \gamma^{\frac{5}{6}} > \|f^*\|_1 + \gamma \geqslant \kappa(W)$. So the lemma tells us that

$$\Gamma(f) \geqslant -\frac{\sqrt[5]{\gamma}}{K^{\frac{2}{5}}}\,. \tag{45}$$

Combining (44) with (45), it suffices to define the function q by

$$q(\gamma) = \max\left(\sqrt[9]{\gamma}, \frac{\sqrt[5]{\gamma}}{K^{\frac{2}{5}}}\right)\,, \quad q > 0\,,$$

since then it is clear that $\lim_{\gamma\searrow 0} q(\gamma) = 0$. □

6.3. **The graphon parameter $\xi(\cdot)$.** In Section 5.2 we outlined why the second moment argument for counting cliques should go through. For the actual execution of this step, however, we need to introduce a new graphon parameter. This parameter is a version of the the cut norm with an exotic scaling. Given an arbitrary graphon W represented on a probability space Ω we define

$$\xi(W) = \sup_{B\subseteq\Omega, \nu(B)>0} \frac{1}{2\nu(B)} \int\int_{(x,y)\in B\times B} \log(1/W(x,y))\,. \tag{46}$$

It is easy to show that $\frac{1}{\xi(W)} \leqslant \kappa(W)$, and more generally, $\frac{1}{\xi(U)} \leqslant \kappa(W)$ for each subgraphon U of W. This fact will not be needed in our proof of Theorem 5.1. However, since it is so basic we record it here.

Fact 6.4. *Let W be an arbitrary graphon on a probability space Ω, and let $U = W\restriction_{A\times A}$ be a subgraphon obtained by restricting W to a set A of positive measure. Then $\frac{1}{\xi(U)} \leqslant \kappa(W)$.*

Proof. By considering the set $B = A$ in (46) we see that

$$\xi(U) \geqslant \frac{1}{2}\int_{A\times A} \log(1/w)\mathrm{d}(\nu_A^2) \overset{(4)}{=} \frac{1}{2}\frac{1}{\nu(A)^2}\int_{A\times A}\log(1/w)\mathrm{d}(\nu^2) \overset{(35)}{\geqslant} \frac{1}{\kappa(W)}\,.$$

□

Our proof of the lower bound in Theorem 5.1 will, however, crucially rely on the opposite direction: In Lemma 6.8 we prove that every graphon W contains a subgraphon U for which $\frac{1}{\xi(U)} > \kappa(W) - \varepsilon$ (here, $\varepsilon > 0$ is arbitrarily small).

Indeed, after picking a subgraphon U for which $\frac{1}{\xi(U)} \approx \kappa(W)$ we continue with the proof of the lower bound in Theorem 5.1 as follows. We prove in Lemma 6.10 that asymptotically almost surely $\omega(\mathbb{G}(n,U)) \gtrapprox \frac{1}{\xi(U)} \log n$. The reason for focusing on $\frac{1}{\xi(U)}$ rather than on $\kappa(U)$ is that the former appears when controlling the second moment (see (58)). As described in (5), the above combination of Lemma 6.8 and Lemma 6.10 will conclude the desired proof of the lower bound in Theorem 5.1.

For the next lemma, note that if G is a finite graph then the value of $\xi(W_G)$ does not depend on the particular representation W_G of G.

Lemma 6.5. *Suppose that $c \in (0,1]$. Let W be a graphon with $\operatorname{ess\,inf} W \geqslant c$ and G an edge-weighted complete graph with all edge-weights $w(i,j)$ in the interval $[c,1]$. Consider the "negative logarithms of W and G", that is, an L^∞-graphon $W'(x,y) := \log(1/W(x,y))$ and a weighted graph G' with $V(G') = V(G)$ and weight function $w'(i,j) = \log(1/w(i,j))$. Then for an arbitrary $\gamma \in (0,1]$ we have*

$$|\xi(W) - \xi(W_G)| \leqslant \max\left(\tfrac{\gamma}{2}\log\tfrac{1}{c}, \tfrac{1}{\gamma}\delta_\square(W',G')\right) .$$

Proof. We shall prove the upper bound only for $\xi(W) - \xi(W_G)$. The upper bound on $\xi(W_G) - \xi(W)$ is done completely analogously. Suppose that W is represented on a probability space Ω. Looking at definition (46), we need to upper-bound

$$\underbrace{\frac{1}{\nu(A)}\int\int_{(x,y)\in A\times A, x<y} \log\left(1/W(x,y)\right)}_{\texttt{S1}} - \underbrace{\xi(W_G)}_{\texttt{S2}} \tag{47}$$

for each set $A \subseteq \Omega$ of positive measure. If $\nu(A) \leqslant \gamma$ then the integral is over a set of measure at most $\frac{1}{2}\nu^2(A) \leqslant \frac{\gamma}{2}\nu(A)$. Thus, the term S1 can be upper-bounded by $-\frac{\gamma}{2}\log(\operatorname{ess\,inf} W) \leqslant \frac{\gamma}{2}\log\frac{1}{c}$, as needed.

Suppose next that $\nu(A) > \gamma$. Suppose first that $\delta_\square(G', W') = 0$. Using the invertible measure preserving maps from (6), we know that for each $\varepsilon > 0$ there exists a graphon representation $W_{G'}$ of G' on Ω such that $d_\square(W_{G'}, W') < \varepsilon\nu(A)$. Then

$$\begin{aligned}
&\frac{1}{\nu(A)}\int\int_{(x,y)\in A\times A, x<y} \log\left(1/W(x,y)\right) - \frac{1}{\nu(A)}\int\int_{(x,y)\in A\times A, x<y} W_{G'}(x,y)\\
&= \frac{1}{2\nu(A)}\left(\int\int_{(x,y)\in A\times A} W'(x,y) - W_{G'}(x,y)\right)\\
&\leqslant \frac{1}{\nu(A)} \cdot d_\square(W', W_{G'}) \leqslant \varepsilon ,
\end{aligned}$$

as was needed.

Suppose next that $\delta_\square(G', W') > 0$. Using the invertible measure preserving maps from (6), we know that for each $\varepsilon > 0$ there exists a graphon representation $W_{G'}$ of G' on Ω such that $d_\square(W_{G'}, W') < (1+\varepsilon)\delta_\square(W_{G'}, W')$. We shall fix such a representation $W_{G'}$ for $\varepsilon = \frac{\nu(A)}{\gamma} - 1$. Then (47) can be bounded from above by

$$\begin{aligned}
&\frac{1}{\nu(A)}\int\int_{(x,y)\in A\times A, x<y} \log\tfrac{1}{W(x,y)} - \frac{1}{\nu(A)}\int\int_{(x,y)\in A\times A, x<y} W_{G'}(x,y)\\
&= \frac{1}{2\nu(A)}\left(\int\int_{(x,y)\in A\times A} W'(x,y) - W_{G'}(x,y)\right)\\
&\leqslant \frac{1}{\nu(A)} \cdot d_\square(W', W_{G'}) \leqslant \frac{1}{\gamma} \cdot \delta_\square(W', W_{G'}) ,
\end{aligned}$$

as was needed. □

Lemma 6.6. *Suppose that $c \in (0,1]$. Let W be a graphon with $\operatorname{ess\,inf} W \geqslant c$. Suppose that a sequence of integers $(k_n)_{n=1}^\infty$ has the property that $\sqrt{\log n} \leqslant k_n \leqslant \sqrt[3]{n}$. Suppose that $\varepsilon > 0$ is arbitrary.*

In the weighted random graph $G \sim \mathbb{H}(n, W)$ consider the family $\mathcal{H}$ of all sets $X \subseteq V(G)$ of size k_n which have the property that $|\xi(W) - \xi(W_{G[X]})| \geqslant \varepsilon$. Then asymptotically almost surely (as $n \to +\infty$) we have that $|\mathcal{H}| \leqslant \varepsilon\binom{n}{k_n}$.

For the proof we shall need the following well-known fact which we include here for reader's convenience.

Fact 6.7. *Let us place m balls independently at random into one of n bins. If $n \geqslant m^3$ then with probability at least $1 - 2n^{-1/3}$ each bin contains at most one ball.*

Proof. Let us first bound the probability that one distinguished ball is placed into a bin which contains some other balls. Recall that for each $n \geqslant 2$,

$$1 - \tfrac{1}{n} \geqslant \exp(-\tfrac{2}{n}) \ . \tag{48}$$

The mentioned probability is exactly

$$1 - (1 - \tfrac{1}{n})^{m-1} \overset{(48)}{\leqslant} 1 - \exp(-\tfrac{2(m-1)}{n}) \leqslant 1 - \exp(-\tfrac{2m}{n}) \leqslant 1 - \exp(-2n^{-2/3}) \leqslant 2n^{-2/3} \ .$$

The claim then follows by summing this error probability over all $m \leqslant n^{1/3}$ balls. □

Proof of Lemma 6.6. Let Ω be the probability space underlying W. Let $W' = \log {}^1\!/\!_W$ be the negative logarithm of W. Note that W' is upper-bounded by $\log {}^1\!/\!_c$. Sampling the random graph $G \sim \mathbb{H}(n, W)$ can be naturally coupled with sampling a random graph $G' \sim \mathbb{H}(n, W')$. So, for the first part of the argument, we shall analyze the graph G'.

Suppose first that n is fixed. Lemma 2.2 (consult also Remark 2.3) implies that with probability at least $1 - \exp(-\frac{n}{2\log n}) = 1 - o(1)$ we have $\delta_\square(G', W') \leqslant \frac{20 \log {}^1\!/\!_c}{\sqrt{\log n}}$. We shall prove the statement for each weighted graph G' satisfying this property (provided that n is sufficiently big). That means that we assume that G' is fixed, and G is its exponentiated version. In particular, all the probabilistic calculations below are only with respect to later randomized steps. Let $\mathcal{K}$ be the family of all subsets X of $V(G') = V(G)$ of size k_n for which $\delta_\square(G', G'[X]) \geqslant \frac{20 \log {}^1\!/\!_c}{\sqrt{\log k_n}}$.

Consider the graphon representation $W_{G'}$ of G' represented on a partition $A_1 \dot\cup A_2 \dot\cup \ldots \dot\cup A_n = \Omega$. Sample the graph $H \sim \mathbb{H}(k_n, W_{G'})$. If we condition on the event $\mathcal{E}$ that the k_n representatives of the vertices of H in the sampling procedure were selected from pairwise distinct "bins" $A_1 \dot\cup A_2 \dot\cup \ldots \dot\cup A_n$ then H is a uniformly random subgraph of G' of order k_n. Fact 6.7 gives that $\mathbf{P}[\mathcal{E}] \geqslant 1 - 2n^{-1/3}$. Thus,

$$\mathbf{P}\left[\delta_\square(H, W_{G'}) \geqslant \frac{20 \log {}^1\!/\!_c}{\sqrt{\log k_n}}\right] \geqslant \mathbf{P}\left[\ \delta_\square(H, W_{G'}) \geqslant \frac{20 \log {}^1\!/\!_c}{\sqrt{\log k_n}} \mid \mathcal{E}\right] \mathbf{P}[\mathcal{E}] \geqslant \frac{|\mathcal{K}|}{\binom{n}{k_n}}(1 - 2n^{-1/3}) \ .$$

Since $\mathbf{P}[\delta_\square(H, W_{G'}) \geqslant \frac{20 \log {}^1\!/\!_c}{\sqrt{\log k_n}}] < \exp(-\frac{k_n}{2 \log k_n})$ we get (for n sufficiently large) that $|\mathcal{K}| \leqslant \varepsilon\binom{n}{k_n}$. So, the lemma will follow provided that we prove that $\mathcal{H} \subseteq \mathcal{K}$, which we prove next.

Indeed, let $X \notin \mathcal{K}$ be an arbitrary vertex set of size k_n. Then $\delta_\square(G'[X], W') \leqslant \delta_\square(G'[X], G') + \delta_\square(W_{G'}, W') \leqslant \frac{20 \log {}^1\!/\!_c}{\sqrt{\log k_n}} + \frac{20 \log {}^1\!/\!_c}{\sqrt{\log n}} \leqslant \frac{40 \log {}^1\!/\!_c}{\sqrt{\log k_n}}$. Then Lemma 6.5 tells us that for each $\gamma \in (0, 1)$,

$$|\xi(W) - \xi(W_{G[X]})| \leqslant \max\left(\tfrac{\gamma}{2} \log {}^1\!/\!_c, \frac{1}{\gamma} \cdot \frac{40 \log {}^1\!/\!_c}{\sqrt{\log k_n}}\right) \ .$$

We take $\gamma = 1/\sqrt[4]{\log k_n}$, and see that the right-hand side is, for large enough n, smaller than ε. This proves that $X \notin \mathcal{H}$ and consequently concludes the lemma. □

Our next two lemmas are crucial in proving the lower bound in Theorem 5.1. The first lemma, Lemma 6.8, tells us that in every graphon W there exists a subgraphon U of W for which we have $\frac{1}{\xi(U)} \gtrapprox \kappa(W)$. The second lemma, Lemma 6.10, then tells us that in $\mathbb{G}(n, U)$ we can typically find cliques of order almost $\frac{1}{\xi(U)} \log n$.

Lemma 6.8. *Suppose that $W : \Omega^2 \to [0, 1]$ is a graphon with $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$. Then for every $\varepsilon > 0$ there exists a set $A \subseteq \Omega$ of positive measure such that for the subgraphon $U = W \restriction_{A \times A}$ we have $\frac{1}{\xi(U)} \geqslant \kappa(W) - \varepsilon$.*

Proof. Let us write $m = \log\left({}^1\!/_{\operatorname{ess\,inf} W}\right)$. Consider the number $\gamma_0 > 0$ and the function $q : (0, \gamma_0) \to \mathbb{R}_+$ given by Lemma 6.3 for the graphon W.

Let $\delta > 0$ be fixed such that $\delta\kappa(W) < \gamma_0$. We use (35) to find a set A of positive measure such that

$$(1+\delta)\frac{2}{\kappa(W)} \geqslant \frac{1}{\nu(A)^2}\int_{A\times A}\log({}^1\!/_W)\mathrm{d}(\nu^2)\ . \tag{49}$$

Consider now the subgraphon $U = W\restriction_{A\times A}$ on the probability space A endowed with the measure ν_A.

We now turn to obtaining the bound on $\xi(U)$. To this end we want to control each term in (46).

Claim 6.8.1. *Suppose that $B \subseteq A$ is an arbitrary set of positive measure. We have*

$$\frac{1}{1+\delta}\cdot\frac{1}{\nu_A(B)}\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu_A^2) \leqslant \frac{2}{\kappa(W)} + \sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}\cdot m\ . \tag{50}$$

Proof of Claim 6.8.1. In the following, we abbreviate $q = q(\delta\kappa(W))$. Suppose first that

$$\nu_A(B) < \sqrt{\frac{q}{(1-\delta)\kappa(W)}}\ .$$

Then

$$\frac{1}{\nu_A(B)}\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu_A^2) \leqslant \nu_A(B)m < \sqrt{\frac{q}{(1-\delta)\kappa(W)}}\cdot m\ ,$$

as needed.

So, it remains to consider the case

$$\nu_A(B) \geqslant \sqrt{\frac{q}{(1-\delta)\kappa(W)}}\ . \tag{51}$$

Let $c > 0$ be maximum such that $\Gamma(c\cdot\mathbf{1}_A, W) \geqslant 0$. That is, we have $\Gamma(c\cdot\mathbf{1}_A, W) = 0$ which can be rewritten as

$$c = \frac{2\nu(A)}{\int_{A\times A}\log({}^1\!/_W)\mathrm{d}(\nu^2)}\ . \tag{52}$$

The "more precisely" part of Proposition 6.1 tells us that

$$c \geqslant (1-\delta)\frac{\kappa(W)}{\nu(A)}\ . \tag{53}$$

Indeed, the assumption of Proposition 6.1 is satisfied by (49) (after a change of W on a null set).

Consider now the function $f = c\cdot\mathbf{1}_A$. We have $\|f\|_1 \geqslant (1-\delta)\kappa(W)$. Thus Lemma 6.3 tells us that $\Gamma(g, W) \geqslant -q$ for each subhistogram g of f. Let us apply this to the function $g := c\cdot\mathbf{1}_B$. Then

$$\Gamma(g,W) = c\nu(B) - \frac{1}{2}c^2\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu^2) \geqslant -q\ ,$$

yielding

$$\nu(B) \geqslant \frac{1}{2}c\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu^2) - \frac{q}{c} \overset{(52)}{=} \nu(A)\frac{\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu^2)}{\int_{A\times A}\log({}^1\!/_W)\mathrm{d}(\nu^2)} - \frac{q}{c}\ .$$

This can be rewritten as

$$\begin{aligned}
\frac{1}{\nu(A)}\int_{A\times A}\log({}^1\!/_W)\mathrm{d}(\nu^2) &\geqslant \frac{1}{\nu(B)}\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu^2) - \frac{q\int_{A\times A}\log({}^1\!/_W)\mathrm{d}(\nu^2)}{c\nu(A)\nu(B)}\\
&\geqslant \frac{1}{\nu(B)}\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu^2) - \frac{qm\nu(A)}{c\nu(B)}\\
&\overset{(53)}{\geqslant} \frac{1}{\nu(B)}\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu^2) - \frac{qm\nu(A)}{(1-\delta)\kappa(W)\nu_A(B)}\\
&\overset{(51)}{\geqslant} \frac{1}{\nu(B)}\int_{B\times B}\log({}^1\!/_W)\mathrm{d}(\nu^2) - \sqrt{\frac{q}{(1-\delta)\kappa(W)}}\cdot m\cdot\nu(A)\ .
\end{aligned} \tag{54}$$

Thus,

$$\begin{aligned}\frac{2}{\kappa(W)} &\overset{(49)}{\geqslant} \frac{1}{1+\delta}\frac{1}{\nu(A)^2}\int_{A\times A}\log(1/W)\mathrm{d}(\nu^2)\\ &\overset{(54)}{\geqslant} \frac{1}{1+\delta}\frac{1}{\nu(A)\nu(B)}\int_{B\times B}\log(1/W)\mathrm{d}(\nu^2) - \frac{1}{1+\delta}\sqrt{\frac{q}{(1-\delta)\kappa(W)}}\cdot m\\ &\overset{(4)}{\geqslant} \frac{1}{1+\delta}\frac{1}{\nu_A(B)}\int_{B\times B}\log(1/W)\mathrm{d}(\nu_A^2) - \sqrt{\frac{q}{(1-\delta)\kappa(W)}}\cdot m\ ,\end{aligned}$$

as required. □

The term $\sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}$ in (50) does not depend on the choice of the set A. Thus, it tends to zero as we let $\delta \searrow 0$. We conclude that for $\delta > 0$ sufficiently small, if we select A as in (49), we have

$$\frac{1}{2\nu_A(B)}\int_{B\times B}\log(1/W)\mathrm{d}(\nu_A^2) \leqslant \frac{1+\delta}{\kappa(W)} + \frac{1+\delta}{2}\cdot\sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}\cdot m$$

for each $B \subseteq A$ of positive measure. By (46), we have

$$\xi(U) \leqslant \frac{1+\delta}{\kappa(W)} + \frac{1+\delta}{2}\cdot\sqrt{\frac{q(\delta\kappa(W))}{(1-\delta)\kappa(W)}}\cdot m\ .$$

If $\delta > 0$ is sufficiently small then the right hand side (which tends to $\frac{1}{\kappa(W)}$ as $\delta \searrow 0$) is smaller than $\frac{1}{\kappa(W)-\varepsilon}$, and so

$$\frac{1}{\xi(U)} \geqslant \kappa(W) - \varepsilon\ ,$$

as was needed. □

For the proof of Lemma 6.10, we shall need the following observation.

Fact 6.9. *Let G be a finite edge-weighted complete graph with vertex set $[n]$ whose symmetric weight function $w : V(G)^2 \to [0,1]$ puts weight 1 on all self-loops. Then for every $C \subseteq [n]$ we have*

$$\prod_{\substack{i,j\in C\\ i<j}} w(i,j) \geqslant \exp(-\xi(W_G)n|C|)\ .$$

Proof. Consider a representation W_G of G on the unit interval I. Suppose that each vertex $v \in [n]$ is represented by a set $D_v \subseteq I$.

The case $C = \emptyset$ is trivial, so assume $C \neq \emptyset$. Consider the set $X = \bigcup_{v\in C} D_v$. Definition (46) gives that

$$\xi(W_G) \geqslant \frac{1}{\lambda(X)}\int_{x,y\in X\times X, x<y}\log(1/W_G(x,y)) = \frac{n}{|C|}\cdot\frac{1}{n^2}\sum_{\substack{i,j\in C\times C\\ i<j}}\log(1/w(i,j))\ .$$

The lemma follows after exponentiation. □

Lemma 6.10. *Suppose that W is a graphon with $\operatorname{ess\,inf} W > 0$. Suppose that $\alpha < 1/\xi(W)$. Then asymptotically almost surely, $\mathbb{G}(n,W)$ contains a clique of order $\alpha\log n$.*

Proof. Choose $\delta > 0$ such that $\alpha(\xi(W)+\delta) < 1$ and let us write

$$\gamma = 1 - \alpha(\xi(W)+\delta)\ . \tag{55}$$

Let $H \sim \mathbb{H}(n,W)$. Set $k := \alpha\log n$. Let $\mathcal{A}$ be be the family of all sets $A \subseteq V(H)$ of size k for which $|\xi(W) - \xi(W_{H[A]})| < \delta$. Lemma 6.6 tells us that with high probability, the graph H has the property that $|\mathcal{A}| \geqslant (1-\delta)\binom{n}{k}$. Condition on this event, and fix a realization of the weighted graph H with a weight function $w : \binom{V(H)}{2} \to [0,1]$ having the above property.

We shall now obtain from H an unweighted graph G by including each edge ij with probability $w(i,j)$. It is our task to show that with high probability, G contains a clique of order k. (Recall that this probability is only with respect to obtaining G from H.)

For each $A \in \mathcal{A}$ set up the indicator X_A of the event that $G[A]$ is a clique. Define $Y_A := \frac{X_A}{\mathbf{E}[X_A]}$ (note that the denominator is not zero because $\operatorname{ess\,inf} W > 0$). Let $Y = \sum_{A \in \mathcal{A}} Y_A$. To conclude the proof, we want to prove that for each $\varepsilon > 0$ (which we now consider fixed), we have

$$Y > 0 \text{ with probability at least } 1 - \varepsilon, \tag{56}$$

provided that n is sufficiently large.

We have $\mathbf{E}[Y_A] = 1$ for each $A \in \mathcal{A}$, and consequently $\mathbf{E}[Y] = |\mathcal{A}|$, which tends to infinity with $n \to +\infty$. Below we shall prove that $\mathbf{E}[Y^2] \leqslant (1+\varepsilon)\mathbf{E}^2[Y]$, which will establish (56) via the usual second-moment argument.

Claim 6.10.1. *We have* $\mathbf{E}[Y^2] < (1+\varepsilon)\mathbf{E}^2[Y]$.

Proof of Claim 6.10.1. For $\ell = 0, 1, \ldots, k$, let us write

$$M_\ell = \sum_{\substack{A,B \in \mathcal{A} \\ |A \cap B| = \ell}} \mathbf{E}[Y_A Y_B] \, .$$

Then we have $\mathbf{E}[Y^2] = \sum_{\ell=0}^{k} M_\ell$. So, it is our goal to bound each of the numbers M_ℓ. We have

$$M_0 \leqslant \binom{n}{k}^2 . \tag{57}$$

For $\ell > 0$ we have[5]

$$M_\ell = \sum_{\substack{A,B \in \mathcal{A} \\ |A \cap B| = \ell}} \mathbf{E}[Y_A Y_B] = \sum_{C \in \binom{V(G)}{\ell}} \sum_{\substack{A,B \in \mathcal{A} \\ A \cap B = C}} \frac{\mathbf{E}[X_A X_B]}{\mathbf{E}[X_A]\mathbf{E}[X_B]} \, .$$

Given two sets $A, B \in \mathcal{A}$, it is easy to see that

$$\mathbf{E}[X_A X_B] = \frac{\mathbf{E}[X_A]\mathbf{E}[X_B]}{\prod_{ij \in \binom{A \cap B}{2}} w(i,j)} \, .$$

Thus,

$$\begin{aligned}
M_\ell &= \sum_{C \in \binom{V(G)}{\ell}} \sum_{\substack{A,B \in \mathcal{A} \\ A \cap B = C}} \prod_{ij \in \binom{C}{2}} w^{-1}(i,j) \\
\boxed{\text{Fact 6.9 applied on the graph } H[A] \text{ and subset } C} \quad &\leqslant \sum_{C \in \binom{V(G)}{\ell}} \sum_{\substack{A,B \in \mathcal{A} \\ A \cap B = C}} \exp\left((\xi(W) + \delta)k\ell\right) \\
&\leqslant \binom{n}{\ell \mid k-\ell \mid k-\ell} \exp\left((\xi(W) + \delta)k\ell\right) \\
&= \frac{n!(n-k)!^2}{(n-2k+\ell)!n!^2} \cdot \frac{k!^2}{\ell!(k-\ell)!^2} \cdot \binom{n}{k}^2 \exp\left((\xi(W) + \delta)k\ell\right) \\
\boxed{n \text{ is sufficiently large}} \quad &\leqslant \left(\frac{2}{n}\right)^\ell \cdot k^{2\ell} \cdot \binom{n}{k}^2 \exp\left((\xi(W) + \delta)k\ell\right) \\
&= \left(\frac{2k^2 \exp\left((\xi(W) + \delta)k\right)}{n}\right)^\ell \cdot \binom{n}{k}^2 \\
&\overset{(55)}{\leqslant} \left(2^\ell k^{2\ell} \exp\left(-\gamma \ell \log n\right)\right) \cdot \binom{n}{k}^2 \\
\boxed{n \gg k} \quad &\leqslant \exp\left(-\tfrac{\gamma \ell \log n}{2}\right) \cdot \binom{n}{k}^2 .
\end{aligned} \tag{58}$$

[5] the calculations below are also valid in the case $\ell = 0$, but we shall not use them in that case

Recall that $\mathbf{E}[Y] = |\mathcal{A}| \geqslant (1-\delta)\binom{n}{k}$. Thus

$$\begin{aligned}\frac{\mathbf{E}[Y^2]}{\mathbf{E}^2[Y]} &\leqslant \frac{\sum_{\ell=0}^{k} M_\ell}{\mathbf{E}^2[Y]} \\ &\overset{(57)}{\leqslant} \frac{1}{(1-\delta)^2} + \frac{\sum_{\ell=1}^{k} M_\ell}{(1-\delta)^2\binom{n}{k}^2} \\ &\overset{(58)}{\leqslant} \frac{1}{(1-\delta)^2} + \frac{1}{(1-\delta)^2}\sum_{\ell=1}^{k}\exp\left(-\tfrac{\gamma\ell\log n}{2}\right) \\ &\leqslant \frac{1}{(1-\delta)^2}\sum_{\ell=0}^{\infty}\exp\left(-\tfrac{\gamma\ell\log n}{2}\right) .\end{aligned}$$

Note that the quotient $\exp\left(-\frac{\gamma\log n}{2}\right)$ of the above geometric series tends to 0 as $n\to\infty$, and so the sum of the series tends to 1. Thus for sufficiently large n and for sufficiently small $\delta>0$ we get $\frac{\mathbf{E}[Y^2]}{\mathbf{E}^2[Y]} < 1+\varepsilon$, as was needed. □

Claim 6.10.1 tells us that $\mathbf{Var}[Y] \leqslant \varepsilon\mathbf{E}^2[Y]$. Therefore, (56) follows from Chebyshev's Inequality. □

7. Proof of Theorem 5.1

Let $c = \operatorname{ess\,inf} W$. Suppose that W is represented on the unit interval $I=(0,1)$ equipped with the Lebesgue measure λ. Let us replace the value of W in every point $(x,y)\in(0,1)^2$ that is not a point of approximate continuity by c. This is a change of measure zero by Fact 2.5. In particular, $\kappa(W)$ does not change, nor does the distribution of the model $\mathbb{G}(n,W)$.

7.1. **Upper bound.** Let $\varepsilon\in(0,\kappa(W)/4)$ be arbitrary. Let n be sufficiently large. We want to show that a.a.s. $G\sim\mathbb{G}(n,W)$ contains no clique of order $k=(\kappa(W)+\varepsilon)\log n$. Let $X_n(G)$ count such cliques. We have

$$\mathbf{E}[X_n(G)] = \int_{x_1}\int_{x_2}\cdots\int_{x_n}\sum_{A\in\binom{[n]}{k}}\prod_{i,j\in A, i<j} W(x_i,x_j) .$$

This summation has $\binom{n}{k} < n^k = \exp\left((\kappa(W)+\varepsilon)\log^2 n\right)$ terms. By (34), each of these terms is bounded by $P_k^{\frac{k(k-1)}{2}}$ where $\lim_{k\to\infty} P_k = \exp\left(-\frac{2}{\kappa(W)}\right)$. So if n is sufficiently large then each term is bounded by

$$\exp\left(-\frac{2}{\kappa(W)}\cdot\frac{k(k-1)}{2}+\varepsilon\right) .$$

Thus,

$$\begin{aligned}\mathbf{E}[X_n(G)] &\leqslant \exp\left((\kappa(W)+\varepsilon)\log^2 n - \frac{2}{\kappa(W)}\cdot\frac{k(k-1)}{2}+\varepsilon\right) \\ &= \exp\left(\left(-\varepsilon-\frac{\varepsilon^2}{\kappa(W)}\right)\log^2 n + \left(1+\frac{\varepsilon}{\kappa(W)}\right)\log n + \varepsilon\right) \to 0 ,\end{aligned}$$

as n goes to infinity. Markov's inequality concludes the proof.

7.2. **Lower bound.** We shall assume that $\operatorname{ess\,sup} W < 1$. Let us justify this step. Suppose that W is an arbitrary graphon. We can then take a sequence of graphons $W_1, W_2, \ldots$, where $W_j = \max(W, 1-\frac{1}{j})$ (pointwise). Then (35) tells us that $\kappa(W_j)\to\kappa(W)$ (even in the case $\kappa(W)=+\infty$). Thus, it suffices to prove a lower bound for each of the graphons W_j.

Let $\varepsilon>0$ be arbitrary. We apply Lemma 6.8 to find a set $A\subseteq\Omega$ of positive measure such that for the subgraphon $U = W\restriction_{A\times A}$ we have $\frac{1}{\xi(U)} \geqslant \kappa(W)-\varepsilon$. Lemma 6.10 then tells us that asymptotically almost surely, $\omega(\mathbb{G}(n,U)) \geqslant (\kappa(W)-2\varepsilon)\log n$. Since there is a coupling of $G=\mathbb{G}(n,W)$ and $G'=\mathbb{G}(\frac{\nu(A)n}{2},U)$ such that G asymptotically almost surely contains a copy of G', we obtain that (cf. (5)),

$$\omega(\mathbb{G}(n,W)) \geqslant (\kappa(W)-3\varepsilon)\log n \text{ asymptotically almost surely.}$$

Since $\varepsilon>0$ was arbitrary, this completes the proof of Theorem 5.1.

8. Graphons with bounded clique number

Given a graphon W, we denote the expectation of $\omega(\mathbb{G}(n,W))$ by $\mu(n,W) := \mathbf{E}[\omega(\mathbb{G}(n,W))]$. Clearly, for each graphon W, and each $n \in \mathbb{N}$, $\omega(\mathbb{G}(n,W))$ is stochastically dominated by $\omega(\mathbb{G}(n+1,W))$. As a consequence, the sequence $\mu(1,W), \mu(2,W), \ldots$ is nondecreasing. We say that W *has bounded clique number* if $\lim_{n\to\infty} \mu(n,W) < +\infty$. Note that one example of graphons of bounded clique number are graphons W for which $t(H,W) = 0$ for on finite graph H.

The following result is the missing piece in our proof of Theorem 3.5.

Lemma 8.1. *Let W be a graphon, and $L = \lim_{n\to\infty} \mu(n,W)$. Then $L = \sup\{k \in \mathbb{N} : t(K_k, W) > 0\}$. In addition, if L is finite then $\lim_{n\to\infty} \mathbf{P}[\omega(\mathbb{G}(n,W)) = L] = 1$.*

Proof. The statement follows from the following claim. Suppose that W is a graphon. Then for each $\ell \in \mathbb{N}$ we have

$$\lim_{n\to\infty} \mathbf{P}[\omega(\mathbb{G}(n,W)) \geqslant \ell] \in \{0,1\} . \tag{59}$$

Indeed, suppose that for some ℓ and n we have that $\mathbf{P}[\omega(\mathbb{G}(n,W)) \geqslant \ell] = \delta > 0$. Then, for each k, we have that $\mathbf{P}[\omega(\mathbb{G}(n,W)) \geqslant \ell] \geqslant 1 - (1-\delta)^k$. Consequently, $\lim_{n\to\infty} \mathbf{P}[\omega(\mathbb{G}(n,W)) \geqslant \ell] = 1$. □

Lemma 8.1 provides some information about graphons of bounded clique number. However, the fact that $t(K_{k+1}, W) = 0$ is not a very explicit description of a structure of the graphon W. To gain better understanding of the structure of graphons of bounded clique number we need to look at some examples.

One example of graphons of bounded clique number are *k-partite graphons*. These are graphons $W : \Omega \times \Omega \to [0,1]$ for which there exists a measurable partition $\Omega = \Omega_1 \dot\cup \Omega_2 \dot\cup \ldots \dot\cup \Omega_k$ such that for each $i \in [k]$, $W \restriction_{\Omega_i \times \Omega_i} = 0$ almost everywhere. In the following example, we show that the structure of graphons with bounded clique number can be more complicated. We consider a sequence of triangle-free graphs $G_1, G_2, \ldots$ whose chromatic numbers grow to infinity (it is a standard exercise that such graphs indeed exist). Let $W_1, W_2, \ldots$ be their graphon representations. We now glue these graphons into one graphon W. Clearly, $\omega(\mathbb{G}(n,W)) \leqslant 2$ with probability one, but W is not k-partite for any k. Here, we show that the structure of graphons of bounded clique number cannot be much more complicated. We call a graphon $W : \Omega \times \Omega \to [0,1]$ *countably-partite*, if there exists a measurable partition $\Omega = \Omega_1 \dot\cup \Omega_2 \dot\cup \ldots$ such that for each $i \in \mathbb{N}$, $W \restriction_{\Omega_i \times \Omega_i} = 0$ almost everywhere.

Let us recall that while this result was inspired by our investigations of $\omega(\mathbb{G}(n,W))$ it is independent of the rest of the paper. Recall that by the above, a graphon W has bounded clique number if and only $t(H,W) = 0$ for some finite graph H.

Theorem 8.2. *Every graphon of bounded clique number is countably-partite.*

Proof. Let $W : \Omega \times \Omega \to [0,1]$ be a graphon of bounded clique number. By Lemma 8.1 we know that $L = \sup\{k \in \mathbb{N} : t(K_k, W) > 0\}$ is a finite (natural) number. First of all, we will show that there is a set $B \subseteq \Omega$ of positive measure such that $W \restriction_{B \times B} = 0$ almost everywhere. We may assume that $L \geqslant 2$ (the case $L = 1$ is trivial). For every $(L-1)$-tuple $\boldsymbol{x} = (x_1, \ldots, x_{L-1}) \in \Omega^{L-1}$, let us denote $Q_{\boldsymbol{x}} = \{y \in \Omega \colon \prod_{i=1}^{L-1} W(x_i, y) > 0\}$. It follows from the equality $t(K_{L+1}, W) = 0$ that for every (up to a set of measure zero) $\boldsymbol{x} \in [0,1]^{L-1}$ such that $\prod_{i<j} W(x_i, x_j) > 0$, we have $W \restriction_{Q_{\boldsymbol{x}} \times Q_{\boldsymbol{x}}} = 0$ almost everywhere. But since $t(K_L, W) > 0$, the set of all $\boldsymbol{x} \in \Omega^{L-1}$ such that $\prod_{i<j} W(x_i, x_j) > 0$ and $\nu(Q_{\boldsymbol{x}}) > 0$, has positive measure. So it suffices to set $B = Q_{\boldsymbol{x}}$ for a suitable $\boldsymbol{x} \in \Omega^{L-1}$.

Next, observe that for every $A \subseteq \Omega$ of positive measure, there is $B \subseteq A$ of positive measure such that $W \restriction_{B \times B} = 0$ almost everywhere. This follows by the previous considerations applied on the subgraphon $W^* = W \restriction_{A \times A}$ (for which we still have that $\sup\{k \in \mathbb{N} : t(K_k, W^*) > 0\} < +\infty$).

Finally, let $W' \colon (0,1)^2 \to [0,1]$ be a representation of the graphon W on $(0,1)$. Then the statement follows by an application of Lemma 2.6. □

9. Concluding remarks

9.1. **Sharpening the results.** As mentioned in Section 1, Matula, Grimmett and McDiarmid proved for $p \in (0,1)$ an asymptotic concentration of $\omega(\mathbb{G}(n,p))$ on two consecutive values for which they provided an explicit formula. It is possible that when, say, $0 < \operatorname{ess\,inf} W \leqslant \operatorname{ess\,sup} W < 1$, then $\omega(\mathbb{G}(n, W))$ is asymptotically concentrated on two consecutive values.

9.2. **Sparse inhomogeneous random graphs.** The random graphs $\mathbb{G}(n, W)$ are inhomogeneous counterparts to $\mathbb{G}(n, p)$. To get a counterpart to sparse random graphs $\mathbb{G}(n, p_n)$, $p_n \to 0$, one introduces rescaling $\mathbb{G}(n, p_n \cdot W)$. In this setting, W need not be bounded from above anymore. An impressive example of work concerning sparse inhomogeneous random graphs is [1] in which the existence and the size of the giant component in $\mathbb{G}(n, \frac{1}{n} \cdot W)$ was determined.

Note that our remark from Section 1.1 is no longer valid: the problem of maximum clique and maximum independent set in $\mathbb{G}(n, p_n \cdot W)$ is genuinely different. It turns out that the more interesting one is that of the independent set. For the Erdős–Rényi random graph $\mathbb{G}(n, p_n)$, the problem of determining the independence number is essentially solved by the above mentioned work [17, 8], and by work of Frieze [7] down to the range $p_n \gg \frac{1}{n}$. Note that the regime $p_n \ll \frac{1}{\sqrt{n}}$ is more subtle as the second moment argument does not work, and indeed Frieze's contribution was in establishing concentration of the count of large independent sets by alternative means. The regime $p_n = C/n$ seems to require methods from statistical physics. In a related model of random regular graphs, these methods have already provided an answer, [5].

It would be of interest to see whether the methods we developed in this paper can give an answer also for the independence number in sparser inhomogeneous random graphs. We believe that they do. Note that the key Lemma 2.2 does have a sparse counterpart, [2, Theorem 2.14].

Acknowledgments

JH thanks Lutz Warnke for suggesting the idea of the proof of Theorem 3.5.

References

[1] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures Algorithms*, 31(1):3–122, 2007.
[2] C. Borgs, J. T. Chayes, H. Cohn, and Y. Zhao. An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions, 2014. arXiv:1401.2906.
[3] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.
[4] D. Conlon, J. Fox, and B. Sudakov. An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.*, 20(6):1354–1366, 2010.
[5] J. Ding, A. Sly, and N. Sun. Maximum independent sets on random regular graphs, 2013.
[6] M. Fabian, P. Habala, P. Hájek, V. Montesinos, and V. Zizler. *Banach space theory*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. The basis for linear and nonlinear analysis.
[7] A. M. Frieze. On the independence number of random graphs. *Discrete Math.*, 81(2):171–175, 1990.
[8] G. R. Grimmett and C. J. H. McDiarmid. On colouring random graphs. *Math. Proc. Cambridge Philos. Soc.*, 77:313–324, 1975.
[9] H. Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*, 175:125–150, 2010.
[10] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: first steps. *Social Networks*, 5(2):109–137, 1983.
[11] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.
[12] J. L. X. Li and B. Szegedy. On the logarithimic calculus and Sidorenko's conjecture. Preprint (arXiv:1107.1153).
[13] L. Lovász. Subgraph densities in signed graphons and the local Simonovits-Sidorenko conjecture. *Electron. J. Combin.*, 18(1):Paper 127, 21, 2011.
[14] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.
[15] L. Lovász and B. Szegedy. Szemerédi's Lemma for the analyst. *J. Geom. and Func. Anal*, 17:252–270, 2007.
[16] L. Lovász and B. Szegedy. Regularity partitions and the topology of graphons. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 415–446. János Bolyai Math. Soc., Budapest, 2010.
[17] D. W. Matula. The largest clique size in a random graph. Technical report, Department of Computer Science, Southern Methodist University, 1976.
[18] M. Molloy and B. Reed. *Graph colouring and the probabilistic method*, volume 23 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2002.
[19] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.
[20] A. Sidorenko. A correlation inequality for bipartite graphs. *Graphs Combin.*, 9(2):201–204, 1993.
[21] M. Simonovits. Extremal graph problems, degenerate extremal problems, and supersaturated graphs. In *Progress in graph theory (Waterloo, Ont., 1982)*, pages 419–437. Academic Press, Toronto, ON, 1984.

Appendix A. A simplified version of Lemma 6.3

Here we provide a weaker version of Lemma 6.3 which deals with the case that the supremum in (25) is attained. While this version is not sufficient for our purposes we decided to offer it to the reader because its proof is based on the same idea yet is stripped off technicalities.

Lemma A.1. *Suppose that W is an arbitrary graphon, and suppose that f^* is an admissible histogram for W for which $\|f^*\|_1 = \kappa(W)$. Then every subhistogram of f^* is admissible for W.*

Proof. The statement follows immediately from Claim 6.2 (which is a simplified version of Lemma 6.2).

We abbreviate $\Gamma(\cdot, W)$ as $\Gamma(\cdot)$. Also, when we say "admissible", we mean with respect to W.

Claim A.1.1. *Assume that g is an arbitrary admissible histogram. Suppose further that $g = g' + g''$ for some histograms g' and g''. Then either g' is admissible, or there exist $\varepsilon_1, \varepsilon_2 \in (0, 1)$ such that for $g^* = (1-\varepsilon_1)g' + (1+\varepsilon_2)g''$ we have that g^* is admissible, and $\|g^*\|_1 > \|g\|_1$.*

Proof of Claim A.1.1. For $\varepsilon_1, \varepsilon_2 \in (0, 1)$, let us write $g^*(\varepsilon_1, \varepsilon_2) = (1-\varepsilon_1)g' + (1+\varepsilon_2)g''$. We also write

$$A = \|g'\|_1, \quad B = \|g''\|_1,$$

$$C = -\frac{1}{2}\int_{x,y} g'(x)g'(y)\log W(x,y), \quad D = -\frac{1}{2}\int_{x,y} g''(x)g''(y)\log W(x,y), \quad E = -\int_{x,y} g'(x)g''(y)\log W(x,y)$$

Note that $A, B, C, D, E \geqslant 0$. For any $\varepsilon_1, \varepsilon_2 \in (0, 1)$, the difference $\Gamma(g^*(\varepsilon_1, \varepsilon_2)) - \Gamma(g)$ can be expressed as

$$\begin{aligned}(1-\varepsilon_1)A + (1+\varepsilon_2)B - (1-\varepsilon_1)^2C - (1+\varepsilon_2)^2D - (1-\varepsilon_1)(1+\varepsilon_2)E - (A+B-C-D-E)\\ = \varepsilon_1(-A+2C+E) + \varepsilon_2(B-2D-E) - \varepsilon_1^2C - \varepsilon_2^2D + \varepsilon_1\varepsilon_2E.\end{aligned}$$

In particular, if $\varepsilon_2 = \frac{A}{B}\varepsilon_1$ then we have

$$\Gamma\left(g^*\left(\varepsilon_1, \frac{A}{B}\varepsilon_1\right)\right) - \Gamma(g) = \varepsilon_1\left(2C + E - \frac{2AD}{B} - \frac{AE}{B}\right) + \varepsilon_1^2\left(-C - \frac{A^2D}{B^2} + \frac{AE}{B}\right). \tag{60}$$

Now let us assume that $\Gamma(g') < 0$, i.e. $A < C$. Then we have

$$\begin{aligned}2C + E - \frac{2AD}{B} - \frac{AE}{B} &> 2A - \frac{2AD}{B} - \frac{AE}{B}\\ &> 2\frac{A}{B}(B-D-E)\\ &> 2\frac{A}{B}\left(B-D-E+(A-C)\right)\\ &= 2\frac{A}{B}\Gamma(g) \geqslant 0.\end{aligned} \tag{61}$$

By (60) and (61), there is $\varepsilon_1 > 0$ (which we fix now) small enough such that $\varepsilon_1, \frac{A}{B}\varepsilon_1 \in (0, 1)$ and

$$\Gamma\left(g^*\left(\varepsilon_1, \frac{A}{B}\varepsilon_1\right)\right) > \Gamma(g)\,(\geqslant 0)\,. \tag{62}$$

Since the function $(\varepsilon_1, \varepsilon_2) \mapsto \Gamma\left(g^*\left(\varepsilon_1, \varepsilon_2\right)\right)$ is obviously continuous, we can find $\varepsilon_2 \in \left(\frac{A}{B}\varepsilon_1, 1\right)$ such that $\Gamma\left(g^*\left(\varepsilon_1, \varepsilon_2\right)\right)$ is still nonnegative. Then we have

$$\|g^*\left(\varepsilon_1, \varepsilon_2\right)\|_1 = (1-\varepsilon_1)A + (1+\varepsilon_2)B > (1-\varepsilon_1)A + \left(1 + \frac{A}{B}\varepsilon_1\right)B = A + B = \|g\|_1.$$

This finishes the proof. □

□

Institute of Mathematics, Czech Academy of Sciences. Žitná 25, 110 00, Praha, Czech Republic. The Institute of Mathematics of the Czech Academy of Sciences is supported by RVO:67985840.

E-mail address: dolezal@math.cas.cz

E-mail address: honzahladky@gmail.com

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

E-mail address: A.Mathe@warwick.ac.uk