

EXPONENTIAL RANDOM GRAPH MODELS WITH SOFT CLIQUE CONSTRAINTS

YASMIN TOUSINEJAD AND VERA KOPONEN

ABSTRACT. Let $r \geq 3$ be fixed, and let $\mathbf{G}_n$ be the set of all simple graphs with vertex set $[n] = \{1, \dots, n\}$. We consider an exponential random graph model which gives higher probability to $G \in \mathbf{G}_n$ than to $H \in \mathbf{G}_n$ if G has fewer r -cliques than H . But all graphs in $\mathbf{G}_n$ have positive probability. The degree to which graphs with fewer r -cliques are given higher probability is determined by a positive weight w . We prove that, asymptotically almost surely as $n \rightarrow \infty$, a random graph from $\mathbf{G}_n$ has a vertex partition into $r - 1$ parts of roughly equal size, the density of edges between the parts is close to $1/2$, and for every $\varepsilon > 0$ the density of edges within any part is less than ε . The asymptotic structural properties are independent of the weight w as long as it is positive. We also extend the result to the context of several clique sizes, each one with its own weight.

1. INTRODUCTION

Studies of large graphs, or other structures, subject to certain hard constraints have a long history in the field of random structures. The classical result of Erdős, Kleitman and Rothschild [12] states that the proportion of bipartite graphs among all triangle-free graphs with vertices $1, \dots, n$ tends to 1 as $n \rightarrow \infty$. More generally, Kolaitis, Prömel and Rothschild [19] proved that, for every fixed $r \geq 3$, the proportion of $(r - 1)$ -partite graphs among all K_r -free graphs with vertices $1, \dots, n$ tends to 1 as $n \rightarrow \infty$. We can say that asymptotically almost surely (a.a.s.) a K_r -free graph is $(r - 1)$ -partite. Later studies have considered the a.a.s. structure of H -free graphs for other “forbidden” subgraphs H (e.g. [2, 3, 15]), and yet other investigations have considered forbidding certain *induced* subgraphs (e.g. [4, 17, 24, 25]). Some studies have considered the a.a.s. structure under constraints other than forbidding a *finite* number of subgraph or induced subgraph configurations (e.g. [6, 5, 20, 31]), and some have considered structures other than graphs (e.g. [18, 29]).

In all mentioned examples, the probability distribution on structures with vertex set $[n] := \{1, \dots, n\}$ that satisfy the given constraints is the uniform one. One can reformulate the problem by considering all graphs (or some other kind of structure) with vertex set $[n]$, assigning probability 0 to graphs that do not satisfy the constraints and equal probability to the other graphs. An alternative is to not assign probability 0 to graphs that violate the constraints, but to instead consider a probability distribution on graphs with vertex set $[n]$ with the following property: A graph with more violations of the constraints is less likely than a graph with fewer violations of the constraints, but graphs with violations can still have positive probability. Since the constraints may be violated in this setting, it is more appropriate to call them *soft constraints*. This idea is present in so-called *exponential random graph models (ERGMs)* [27, 9]. The same idea also underlies probability distributions defined by so-called *Markov logic networks (MLNs)* [26], which are formal specification models in the field of *Statistical Relational Artificial Intelligence (SRAI)* [8, 11, 13]. MLNs can be defined for finite relational structures, not just graphs, but when restricted to graphs, they are special cases of ERGMs. Not much

is known about *structural* properties of large random structures defined by fixed-weight MLNs. The investigations most directly related to this topic that we are aware of are [21, 30].

Our main result concerns an ERGM, or equivalently an MLN, on random graphs which “penalizes” (for any fixed $r \geq 3$) r -cliques, that is, copies of the complete graph on r vertices, denoted K_r . In other words, a graph with more r -cliques will be less likely than a graph with fewer r -cliques, but all graphs have positive probability. The strength of the penalization is expressed by a positive weight w . The precise definition of the distribution is given in Definition 2.2. Informally speaking, our main result, Theorem 2.7, shows that, for every fixed positive weight w and every $\varepsilon > 0$, with probability tending to 1 as the number n of vertices tends to infinity, the following holds for a random graph:

- (1) There is a partition of the vertex set into $r - 1$ parts, each of size close to $\frac{n}{r-1}$.
- (2) The density of edges between any two different parts is close to $\frac{1}{2}$.
- (3) For every $\varepsilon > 0$, if n is large enough then the density of edges within any part is less than ε .

Note that the asymptotic almost sure structure of a random graph does *not* depend on the weight w (as long as $w > 0$).

We also prove (Corollary 6.6) that if $r \geq 4$ then the probability that a random graph is $(r - 2)$ -partite tends to 0 exponentially fast as the number of vertices tends to infinity.

A further consequence of our results is a stochastic block model limit for fixed induced subgraphs (Corollary 6.2). Informally speaking, for every fixed m , as n tends to infinity, the subgraph induced by $[m]$ converges in distribution to a stochastic block model with $r - 1$ classes such that each of the m vertices is assigned independently and uniformly to one of these classes. Given these assignments, there are no edges within a class, and each possible edge between different classes is included independently with probability $\frac{1}{2}$. The limiting distribution depends on r , but not on w .

We can refine the probability model considered so that we consider some integers $3 \leq r_1 < \dots < r_k$ and for each r_i we penalize an r_i -clique with weight $w_i > 0$. (The larger w_i is, the more strongly r_i -cliques are penalized.) We show (Theorem 6.7) that, for all choices of weights $w_1, \dots, w_k > 0$, the asymptotic behaviour of random graphs under this model is determined by r_1 . More precisely, with probability tending to 1 as the number n of vertices tends to infinity, a random graph will satisfy conditions (1) – (3) above if we let $r := r_1$.

Our main results are proved by using graphons [7, 22, 16], which are analytic objects used to represent limits of sequences of dense graphs. The article is organized as follows. Just below we introduce the general notation and terminology that will be used. In Section 2 we define the probability model and state the main results (in Theorem 2.7). Section 3 defines the concept of graphon, as well as related concepts, and proves some basic results that will be used later. In Section 4 we work with the concepts of entropy of graphons, homomorphism density of K_r in graphons, and the cut distance of a graphon from a “balanced $(r - 1)$ -partite graphon”, and show how they are related. In Section 5 we prove Theorem 2.7. Section 6 derives some consequences of Theorem 2.7 which were briefly described above.

General notation and terminology. By $\mathbb{N}$ we denote the set of positive integers. For $n \in \mathbb{N}$, $[n] := \{1, \dots, n\}$. For $n \in \mathbb{N}$, let $\mathbf{G}_n$ be the set of all undirected graphs without loops (i.e. simple graphs) with vertex set $[n]$. For every finite graph G , let $E(G)$ denote its edge set and let $\chi(G)$ denote its chromatic number.

If $n \in \mathbb{N}$, $G \in \mathbf{G}_n$, and $A \subseteq [n]$, set

$$e_G(A) := |\{\{u, v\} \in E(G) : u, v \in A\}|.$$

If in addition $B \subseteq [n]$ is disjoint from A , define

$$e_G(A, B) := |\{\{u, v\} \in E(G) : |\{u, v\} \cap A| = 1, |\{u, v\} \cap B| = 1\}|.$$

For $n \in \mathbb{N}$ and graphs $G, H \in \mathbf{G}_n$, define the n^2 -normalized edge-edit distance between G and H by

$$d_{\text{edit}}(G, H) := \frac{1}{n^2} |E(G) \Delta E(H)|.$$

Here Δ denotes symmetric difference. Thus $d_{\text{edit}}(G, H)$ is the number of edge insertions and deletions needed to turn G into H , normalized by n^2 .

For $n \in \mathbb{N}$ and $p \in [0, 1]$, the notation

$$\mathbf{H}_n \sim G(n, p)$$

means that $\mathbf{H}_n$ is a random element of $\mathbf{G}_n$ sampled by including an unordered pair $\{i, j\}$ as an edge with probability p , independently of whether other pairs have been included or not. In particular, $\mathbf{H}_n \sim G(n, \frac{1}{2})$ means that every graph in $\mathbf{G}_n$ has probability $2^{-\binom{n}{2}}$, so $\mathbf{H}_n$ is uniformly distributed on $\mathbf{G}_n$.

For every integer $s \geq 2$, K_s denotes a complete graph with s vertices. For every finite graph G , let $K_s(G)$ denote the number of *unordered* copies of K_s in G , in other words, the number of subgraphs of G that are isomorphic to K_s .

2. THE PROBABILITY MODEL AND THE MAIN RESULTS

Throughout this article $r \geq 3$ is a fixed integer and $w > 0$ is a fixed real number interpreted as the “*weight*” that will be associated to every r -tuple of vertices of a graph that does not form an r -clique.

2.1. The probability distribution.

Definition 2.1. For every $n \in \mathbb{N}$ and $G \in \mathbf{G}_n$, let

$$(2.1) \quad N_r(G) := n^r - r! K_r(G),$$

so $N_r(G)$ is the number of ordered r -tuples $(v_1, \dots, v_r) \in [n]^r$ (with repetitions allowed) that do not consist of r distinct vertices that form an r -clique in G .

Definition 2.2. For each $n \in \mathbb{N}$ and $G \in \mathbf{G}_n$, we define its *unnormalized mass* by

$$\mu_{n,r}(G) := \exp(w N_r(G)).$$

We then define the *partition function* (or *normalizing constant*) by

$$Z_{n,r} := \sum_{J \in \mathbf{G}_n} \mu_{n,r}(J).$$

Finally, we define the *probability measure* of the model by

$$\mathbb{P}_{n,r}(A) := \frac{1}{Z_{n,r}} \sum_{G \in A} \mu_{n,r}(G) \quad \text{for all } A \subseteq \mathbf{G}_n.$$

For a single graph $G \in \mathbf{G}_n$, we also set

$$\mathbb{P}_{n,r}(G) := \mathbb{P}_{n,r}(\{G\}) = \frac{\mu_{n,r}(G)}{Z_{n,r}}.$$

Since w is fixed in the model, we suppress the dependence on w in the notation $\mu_{n,r}$, $Z_{n,r}$, and $\mathbb{P}_{n,r}$. Note that if $G, H \in \mathbf{G}_n$ and $K_r(G) < K_r(H)$, then $\mathbb{P}_{n,r}(G) > \mathbb{P}_{n,r}(H)$. The distribution $\mathbb{P}_{n,r}$ is exactly the probability distribution on $\mathbf{G}_n$ obtained from a Markov logic network [26] with one soft constraint of the form

$$\bigvee_{1 \leq i < j \leq r} \neg E(x_i, x_j) \quad \text{with weight } w,$$

where $\neg E(x_i, x_j)$ expresses that there is no edge between x_i and x_j .

Remark 2.3. (On counting only tuples without repetitions) In Definition 2.2 we count *all* ordered r -tuples (also those with repetitions)

$$(x_1, \dots, x_r) \in [n]^r.$$

Another natural convention is to count only those ordered tuples whose entries are pairwise distinct. Under that convention, the total number of allowed ordered tuples is

$$n(n-1) \cdots (n-r+1),$$

and the falsifying tuples are still exactly the $r!$ orderings of the vertices of each copy of K_r . So the number of satisfied groundings becomes

$$N_r^{\text{dist}}(G) = n(n-1) \cdots (n-r+1) - r! K_r(G).$$

If we use the same weight w , then the corresponding unnormalized mass is

$$\mu_{n,r}^{\text{dist}}(G) := \exp(w N_r^{\text{dist}}(G)) = \exp(w [n(n-1) \cdots (n-r+1) - r! K_r(G)]).$$

Compare this with our present convention,

$$\mu_{n,r}(G) = \exp(w [n^r - r! K_r(G)]).$$

Subtracting the exponents gives

$$\mu_{n,r}(G) = \exp(w [n^r - n(n-1) \cdots (n-r+1)]) \mu_{n,r}^{\text{dist}}(G).$$

The factor

$$\exp(w [n^r - n(n-1) \cdots (n-r+1)])$$

is *deterministic* in the sense that it depends only on n , r , and w , and not on the graph G . Therefore this factor changes the partition function, but it disappears when we normalize to obtain probabilities. Indeed, if

$$Z_{n,r}^{\text{dist}} := \sum_{J \in \mathbf{G}_n} \mu_{n,r}^{\text{dist}}(J),$$

then

$$Z_{n,r} = \exp(w [n^r - n(n-1) \cdots (n-r+1)]) Z_{n,r}^{\text{dist}},$$

and hence

$$\mathbb{P}_{n,r}(G) = \frac{\mu_{n,r}(G)}{Z_{n,r}} = \frac{\mu_{n,r}^{\text{dist}}(G)}{Z_{n,r}^{\text{dist}}}.$$

So, for the same clause weight w , the two counting conventions define exactly the same probability measure on $\mathbf{G}_n$. Thus the choice between these two conventions does not affect which graphs are more or less likely.

2.2. The equivalent exponential random graph model. The distribution $\mathbb{P}_{n,r}$ can be defined in the setting of exponential random graph models [27].

Definition 2.4. Fix $n, k \in \mathbb{N}$. An *Exponential Random Graph Model (ERGM)* on $\mathbf{G}_n$ is a probability distribution of the form

$$\mathbb{P}_{n,\vartheta}(G) := \frac{\exp(\sum_{i=1}^k \vartheta_i T_i(G))}{Z_n^{\text{ERGM}}(\vartheta)} \quad \text{for all } G \in \mathbf{G}_n,$$

where $T_1, \dots, T_k$ are real-valued functions on $\bigcup_{n \in \mathbb{N}} \mathbf{G}_n$, $\vartheta = (\vartheta_1, \dots, \vartheta_k) \in \mathbb{R}^k$ is a parameter vector, and

$$Z_n^{\text{ERGM}}(\vartheta) := \sum_{J \in \mathbf{G}_n} \exp\left(\sum_{i=1}^k \vartheta_i T_i(J)\right)$$

is the normalizing constant. All else being equal, a positive value of ϑ_i rewards larger values of $T_i(G)$ in the exponent, while a negative value penalizes them.

Note that if $k = 1$, $\vartheta_1 = w$, and $T_1(G) = N_r(G)$, then

$$\mathbb{P}_{n,r}(G) = \mathbb{P}_{n,\vartheta}(G) \quad \text{for all } G \in \mathbf{G}_n.$$

Hence $\mathbb{P}_{n,r}$ is an ERGM.

2.3. The reduced partition function. It will be convenient to work with the quantities defined below.

Definition 2.5. (a) For all $n \in \mathbb{N}$, let

$$(2.2) \quad \begin{aligned} \tilde{Z}_{n,r} &:= \tilde{Z}_{n,r}(w) := \sum_{J \in \mathbf{G}_n} \exp(-r! w K_r(J)) \quad \text{and} \\ F_{n,r} &:= F_{n,r}(w) := \frac{1}{n^2} \ln \tilde{Z}_{n,r}. \end{aligned}$$

We call $\tilde{Z}_{n,r}$ the *reduced partition function* and $F_{n,r}(w)$ the *reduced free energy*.

(b) For a graph $G \in \mathbf{G}_n$, we call

$$\exp(-r! w K_r(G))$$

the *reduced mass* of G . Thus $\tilde{Z}_{n,r}$ is the sum of the reduced masses of all graphs in $\mathbf{G}_n$.

The partition function $Z_{n,r}$ sums the unnormalized masses of all graphs in $\mathbf{G}_n$. As the next lemma shows, its logarithm contains the graph-independent term wn^r . This term changes the normalizing constant but not the probability measure, so we will often remove it and work with the reduced partition function and reduced free energy.

Lemma 2.6. *For every $n \in \mathbb{N}$ and every $G \in \mathbf{G}_n$,*

$$\mu_{n,r}(G) = \exp(w n^r) \exp(-r! w K_r(G)).$$

Moreover,

$$Z_{n,r} = \exp(w n^r) \tilde{Z}_{n,r},$$

and

$$(2.3) \quad \mathbb{P}_{n,r}(G) = \frac{\exp(-r! w K_r(G))}{\sum_{J \in \mathbf{G}_n} \exp(-r! w K_r(J))} = \frac{\exp(-r! w K_r(G))}{\tilde{Z}_{n,r}}.$$

Proof. Insert (2.1) into the definition of $\mu_{n,r}(G)$:

$$\mu_{n,r}(G) = \exp(w [n^r - r! K_r(G)]).$$

Now factor out the deterministic term e^{wn^r} . The formulas for $Z_{n,r}$ and $\mathbb{P}_{n,r}$ follow immediately. $\square$

2.4. Main results. Recall that we have fixed an integer $r \geq 3$ and a real $w > 0$. The notation

$$\mathbf{G}_n \sim \mathbb{P}_{n,r}$$

means that $\mathbf{G}_n$ is a random graph from $\mathbf{G}_n$ under the distribution $\mathbb{P}_{n,r}$. (So if $J \in \mathbf{G}_n$ then $\mathbb{P}(\mathbf{G}_n = J) = \mathbb{P}_{n,r}(J)$.) Recall also notation introduced in Section 1. Some of the notions appearing in the main result below are defined only later: A *graphon* is defined in Definition 3.1. For a graph $G \in \mathbf{G}_n$, its *empirical graphon* W_G is defined in Definition 3.2. The *cut distance* $\delta_\square$ between two graphons is defined in Definition 3.6, and the cut distance from a graphon to a family is defined in Definition 3.7. The family B_q^* is defined in Definition 4.1; informally, it consists of the balanced q -block graphons that are equal to 0 on the diagonal blocks and equal to $\frac{1}{2}$ on the off-diagonal blocks.

Theorem 2.7. Fix $r \geq 3$, set $q = r - 1$. Let $\mathbf{G}_n \sim \mathbb{P}_{n,r}$. Then the following hold.

(i) The reduced free energy satisfies

$$\lim_{n \rightarrow \infty} F_{n,r}(w) = \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

(ii) For every $\varepsilon > 0$, there are constants $c = c(\varepsilon, r) > 0$ and $N = N(\varepsilon, r, w)$ such that, for every $n \geq N$,

$$\mathbb{P}(\delta_{\square}(W_{\mathbf{G}_n}, B_q^*) \geq \varepsilon) \leq e^{-cn^2}.$$

(iii) For every $\varepsilon > 0$, there are constants $c = c(\varepsilon, r) > 0$ and $n_0 = n_0(\varepsilon, r, w)$ such that, for every $n \geq n_0$, with probability at least $1 - e^{-cn^2}$, the graph $\mathbf{G}_n$ has a partition

$$[n] = V_1 \sqcup \cdots \sqcup V_q$$

with the following properties:

(a) Each part is almost balanced:

$$||V_i| - n/q| \leq \varepsilon n \quad \text{for every } i \in [q].$$

(b) Each between-part edge density is close to $\frac{1}{2}$:

$$\left| \frac{e_{\mathbf{G}_n}(V_i, V_j)}{|V_i||V_j|} - \frac{1}{2} \right| \leq \varepsilon \quad \text{for all distinct } i, j \in [q].$$

(c) The total number of edges inside the parts satisfies

$$\sum_{i=1}^q e_{\mathbf{G}_n}(V_i) \leq \frac{\varepsilon n^2}{10q}.$$

(d) The total density of edges inside the parts is at most ε :

$$\frac{\sum_{i=1}^q e_{\mathbf{G}_n}(V_i)}{\sum_{i=1}^q \binom{|V_i|}{2}} \leq \varepsilon.$$

In particular, if $\mathbf{G}'_n$ is obtained from $\mathbf{G}_n$ by deleting all edges whose endpoints lie in the same part, then $\mathbf{G}'_n$ is q -partite and

$$d_{\text{edit}}(\mathbf{G}_n, \mathbf{G}'_n) \leq \frac{\varepsilon}{10q} \leq \frac{\varepsilon}{2}.$$

2.5. Related work. In the introduction we mentioned several studies of graphs, or other relational structures, subject to “hard” constraints. We also mentioned [21, 30] about structures subject to “soft” constraints defined by Markov logic networks. As we use graphons to prove our main results we like to mention [1] which considers Markov logic networks and a multi-relational analogue of a graphon in the context when one considers several different edge relations. However, the results in [1], which generalize the basic theory of graphons, are not of help in the present context as far as we can see.

In [9] Chatterjee and Diaconis consider ERGM distributions on $\mathbf{G}_n$ of the form

$$\mathbb{P}_n^T(G) = \frac{\exp(n^2 T(G))}{\sum_{J \in \mathbf{G}_n} \exp(n^2 T(J))}$$

where T is a function from $\bigcup_{n \in \mathbb{N}} \mathbf{G}_n$ into the reals. If we rewrite the distribution $\mathbb{P}_{n,r}$ so that, for an appropriate choice of T and all n , $\mathbb{P}_{n,r} = \mathbb{P}_n^T$ then, for all n and all $G \in \mathbf{G}_n$, we must have

$$T(G) := \frac{w N_r(G)}{n^2}.$$

Recall that $N_r(G)$ is the number of r -tuples of vertices of G that do not form an r -clique. So if $G \in \mathbf{G}_n$ does not have an r -clique (e.g. if G is $(r-1)$ -partite), then $N_r(G) = n^r$

where $r \geq 3$. It follows that $\max_{G \in \mathbf{G}_n} T(G)$ is unbounded as $n \rightarrow \infty$. But the results of [9] assume that $\max_{G \in \mathbf{G}_n} T(G)$ is bounded as $n \rightarrow \infty$. Therefore their results do not apply to the present context.

3. GRAPHON PRELIMINARIES

3.1. Basic definitions.

Definition 3.1 (Graphon). A *graphon* is a measurable function

$$W : [0, 1]^2 \rightarrow [0, 1]$$

such that $W(x, y) = W(y, x)$ for almost every $(x, y) \in [0, 1]^2$. Let $\mathcal{W}$ denote the family of these measurable functions.

Let λ denote *Lebesgue measure* on $[0, 1]$. For each integer $m \geq 1$, let λ^m denote the *product Lebesgue measure* on $[0, 1]^m$.

Graphons are the standard limit objects for dense graphs. They also define random graph models. Let W be a graphon and let $n \in \mathbb{N}$. Choose $X_1, \dots, X_n$ independently and uniformly from $[0, 1]$. Conditional on $X_1, \dots, X_n$, put each edge $\{i, j\}$ with $i < j$ into the graph independently with probability

$$W(X_i, X_j).$$

Definition 3.2 (Empirical graphon). Let $n \in \mathbb{N}$, and let G be a graph with vertex set $[n]$ and adjacency matrix $(a_{ij})_{1 \leq i, j \leq n}$, where

$$a_{ij} = a_{ji} \quad \text{and} \quad a_{ii} = 0.$$

Split $[0, 1]$ into n intervals of equal length by setting

$$I_i = \left[\frac{i-1}{n}, \frac{i}{n} \right) \quad \text{for } 1 \leq i < n, \quad I_n = \left[\frac{n-1}{n}, 1 \right].$$

The *empirical graphon* of G is the graphon $W_G : [0, 1]^2 \rightarrow [0, 1]$ defined as follows. If $x \in I_i$ and $y \in I_j$, then

$$W_G(x, y) := a_{ij}.$$

Thus W_G records the adjacency matrix of G on the n^2 rectangles $I_i \times I_j$. It is measurable, symmetric, and takes only the values 0 and 1. Also, $W_G = 0$ on each rectangle $I_i \times I_i$, because $a_{ii} = 0$.

Definition 3.3 (Homomorphism density). Let H be a finite simple graph with vertex set $V(H) = \{1, \dots, k\}$. For a graphon W , the *homomorphism density* of H in W is

$$t(H, W) := \int_{[0, 1]^k} \prod_{\{i, j\} \in E(H)} W(x_i, x_j) dx_1 \cdots dx_k.$$

Remark 3.4. For $H = K_r$ this becomes

$$t(K_r, W) = \int_{[0, 1]^r} \prod_{1 \leq a < b \leq r} W(x_a, x_b) dx_1 \cdots dx_r.$$

The integrand is nonnegative. So if $t(K_r, W) = 0$, then the product inside the integral must vanish almost everywhere.

From now on, a graphon W is called a *graphon of zero K_r -density* if

$$t(K_r, W) = 0,$$

that is, if its K_r -homomorphism density is zero.

Definition 3.5 (Cut norm). For an integrable function $U : [0, 1]^2 \rightarrow \mathbb{R}$, the cut norm is

$$\|U\|_{\square} := \sup_{\substack{S, T \subseteq [0, 1] \\ S, T \text{ measurable}}} \left| \int_{S \times T} U(x, y) dx dy \right|.$$

Definition 3.6 (Cut distance). A measurable map $\phi : [0, 1] \rightarrow [0, 1]$ is *measure-preserving* if

$$\lambda(\phi^{-1}(A)) = \lambda(A) \quad \text{for every Lebesgue-measurable } A \subseteq [0, 1].$$

A *measure-preserving bijection* is a bijection $\phi : [0, 1] \rightarrow [0, 1]$ such that both ϕ and ϕ^{-1} are Lebesgue-measurable and ϕ is measure-preserving.

If W is a graphon and ϕ is a measure-preserving map, define

$$W^{\phi}(x, y) := W(\phi(x), \phi(y)).$$

Thus W^{ϕ} is obtained from W by applying ϕ to both variables. When ϕ is a measure-preserving bijection, we call W^{ϕ} a *relabeling* of W .

For graphons W_1 and W_2 , their *cut distance* is

$$\delta_{\square}(W_1, W_2) := \inf_{\phi} \|W_1 - W_2^{\phi}\|_{\square},$$

where the infimum is taken over all measure-preserving bijections $\phi : [0, 1] \rightarrow [0, 1]$.

Relabeling by a measure-preserving bijection does not change the cut norm. Indeed, let $F : [0, 1]^2 \rightarrow \mathbb{R}$ be a measurable function with

$$\int_{[0, 1]^2} |F(x, y)| dx dy < \infty,$$

and let ϕ be a measure-preserving bijection. Define

$$F^{\phi}(x, y) := F(\phi(x), \phi(y)).$$

Then

$$\|F^{\phi}\|_{\square} = \|F\|_{\square}.$$

This follows by changing variables in the integral over $S \times T$. Since ϕ is measure-preserving and bijective, the sets $\phi(S)$ and $\phi(T)$ run through the same measurable subsets of $[0, 1]$ as S and T . Thus the supremum in the cut norm is unchanged.

Definition 3.7. Let W be a graphon, and let $\mathcal{F}$ be a nonempty family of graphons. The *cut distance from W to $\mathcal{F}$* is

$$\delta_{\square}(W, \mathcal{F}) := \inf_{V \in \mathcal{F}} \delta_{\square}(W, V).$$

The cut distance $\delta_{\square}$ is a pseudometric on $\mathcal{W}$. Indeed, let $U, V, W \in \mathcal{W}$. To see symmetry, let ϕ be a measure-preserving bijection. By invariance of the cut norm under relabeling,

$$\|U - V^{\phi}\|_{\square} = \|U^{\phi^{-1}} - V\|_{\square} = \|V - U^{\phi^{-1}}\|_{\square}.$$

Taking the infimum over ϕ gives $\delta_{\square}(U, V) = \delta_{\square}(V, U)$. For the triangle inequality, if ϕ and ψ are measure-preserving bijections, then

$$\begin{aligned} \|U - W^{\psi \circ \phi}\|_{\square} &\leq \|U - V^{\phi}\|_{\square} + \|V^{\phi} - (W^{\psi})^{\phi}\|_{\square} \\ &= \|U - V^{\phi}\|_{\square} + \|V - W^{\psi}\|_{\square}. \end{aligned}$$

Taking the infimum over ϕ and ψ proves the triangle inequality. Nonnegativity and $\delta_{\square}(U, U) = 0$ are immediate.

Notation 3.8 (Reduced graphon space). Define

$$W \sim V \iff \delta_{\square}(W, V) = 0,$$

and let

$$\widetilde{\mathcal{W}} := \mathcal{W} / \sim, \quad \pi : \mathcal{W} \rightarrow \widetilde{\mathcal{W}}$$

be the quotient space and quotient map. We use $\widetilde{W} := \pi(W)$ and, for $n \in \mathbb{N}$ and $G \in \mathbf{G}_n$, $\widetilde{W}_G := \pi(W_G)$. For reduced graphons define

$$d_{\square}(\widetilde{U}, \widetilde{V}) := \delta_{\square}(U, V),$$

where U and V are arbitrary representatives. This definition does not depend on the chosen representatives. Indeed, if $U' \sim U$ and $V' \sim V$, then

$$\delta_{\square}(U', V') \leq \delta_{\square}(U', U) + \delta_{\square}(U, V) + \delta_{\square}(V, V') = \delta_{\square}(U, V).$$

The same argument with (U, V) and (U', V') interchanged gives

$$\delta_{\square}(U, V) \leq \delta_{\square}(U', V').$$

Hence

$$\delta_{\square}(U', V') = \delta_{\square}(U, V).$$

Thus $d_{\square}$ is well defined on $\widetilde{\mathcal{W}}$. It is a metric because nonnegativity, symmetry, and the triangle inequality are inherited from $\delta_{\square}$. Also,

$$d_{\square}(\widetilde{U}, \widetilde{V}) = 0$$

holds exactly when $U \sim V$, which means $\widetilde{U} = \widetilde{V}$. The metric space $(\widetilde{\mathcal{W}}, d_{\square})$ is compact; see [22, Thm. 9.23]. For $\mathcal{F} \subseteq \mathcal{W}$, we use

$$\pi(\mathcal{F}) := \{\pi(W) : W \in \mathcal{F}\} \subseteq \widetilde{\mathcal{W}}$$

for its quotient image.

Remark 3.9 (Zero cut distance). For graphons U and V , the condition

$$\delta_{\square}(U, V) = 0$$

means that U and V represent the same graph limit. Equivalently, there exist measure-preserving maps $\phi, \psi : [0, 1] \rightarrow [0, 1]$ such that

$$U^{\phi} = V^{\psi} \quad \text{almost everywhere;}$$

see [22, Cor. 10.34 and Cor. 10.35(a)].

The maps in this characterization need not be bijections. Consequently, the condition $\delta_{\square}(U, V) = 0$ is weaker than almost-everywhere equality after a single measure-preserving relabeling.

By Notation 3.8, graphons at zero cut distance represent the same point of $\widetilde{\mathcal{W}}$. This quotient space is the reduced graphon space used in [10, Sec. 1.3].

Notation 3.10. For an integrable function $U : [0, 1]^2 \rightarrow \mathbb{R}$, the L^1 -norm is defined as

$$\|U\|_1 := \int_{[0, 1]^2} |U(x, y)| \, dx \, dy.$$

Notation 3.11 (Indicator functions). For any set or event A , $\mathbf{1}_A$ denotes its indicator. Thus $\mathbf{1}_A(z) = 1$ when $z \in A$ and $\mathbf{1}_A(z) = 0$ otherwise.

Lemma 3.12. *We will use the following two forms of the Cauchy-Schwarz inequality.*

(i) *If $S \subseteq [0, 1]^2$ is measurable and $f : S \rightarrow \mathbb{R}$ is measurable with*

$$\int_S f^2 < \infty,$$

then

$$\left(\int_S |f| \right)^2 \leq \lambda^2(S) \int_S f^2.$$

(ii) *If $m \geq 1$ is an integer and $x_1, \dots, x_m \in \mathbb{R}$, then*

$$\left(\sum_{a=1}^m x_a \right)^2 \leq m \sum_{a=1}^m x_a^2.$$

Equivalently,

$$\sum_{a=1}^m x_a^2 \geq \frac{1}{m} \left(\sum_{a=1}^m x_a \right)^2.$$

Consequently, if $q \in \mathbb{N}$, $n \in \mathbb{Z}_{\geq 0}$, $m_1, \dots, m_q \in \mathbb{Z}_{\geq 0}$, and

$$\sum_{a=1}^q m_a = n,$$

then

$$\sum_{a=1}^q \binom{m_a}{2} = \frac{1}{2} \left(\sum_{a=1}^q m_a^2 - n \right) \geq \frac{1}{2} \left(\frac{n^2}{q} - n \right).$$

Proof. For (i), apply [28, Thm. 11.35] to the functions $|f|$ and 1 on S . We get

$$\int_S |f| \leq \left(\int_S f^2 \right)^{1/2} \left(\int_S 1 \right)^{1/2}.$$

Since

$$\int_S 1 = \lambda^2(S),$$

it follows that

$$\left(\int_S |f| \right)^2 \leq \lambda^2(S) \int_S f^2.$$

For (ii), apply [28, Thm. 1.35] to the two lists

$$x_1, \dots, x_m \quad \text{and} \quad 1, \dots, 1.$$

This gives

$$\left(\sum_{a=1}^m x_a \right)^2 \leq \left(\sum_{a=1}^m x_a^2 \right) \left(\sum_{a=1}^m 1 \right) = m \sum_{a=1}^m x_a^2.$$

Dividing by m , we obtain

$$\sum_{a=1}^m x_a^2 \geq \frac{1}{m} \left(\sum_{a=1}^m x_a \right)^2.$$

Now apply (ii) with $m = q$ and $x_a = m_a$. Since

$$\sum_{a=1}^q m_a = n,$$

we have

$$\sum_{a=1}^q m_a^2 \geq \frac{n^2}{q}.$$

Therefore

$$\sum_{a=1}^q \binom{m_a}{2} = \frac{1}{2} \sum_{a=1}^q m_a(m_a - 1) = \frac{1}{2} \left(\sum_{a=1}^q m_a^2 - n \right) \geq \frac{1}{2} \left(\frac{n^2}{q} - n \right).$$

□

Lemma 3.13. *Let $F : [0, 1]^2 \rightarrow \mathbb{R}$ be bounded and measurable, and let $f, g : [0, 1] \rightarrow [0, 1]$ be measurable. Then*

$$\left| \int_{[0,1]^2} F(x, y) f(x) g(y) dx dy \right| \leq \|F\|_{\square}.$$

Proof. By [16, Eq. (4.2) and Rem. 4.1],

$$\|F\|_{\square} = \sup_{\substack{\phi, \psi : [0,1] \rightarrow [0,1] \\ \phi, \psi \text{ measurable}}} \left| \int_{[0,1]^2} F(x, y) \phi(x) \psi(y) dx dy \right|.$$

Taking $\phi = f$ and $\psi = g$ gives

$$\left| \int_{[0,1]^2} F(x, y) f(x) g(y) dx dy \right| \leq \|F\|_{\square}.$$

□

Lemma 3.14 (Clique density is Lipschitz in cut distance). *Let $r \geq 2$. For all graphons U and V ,*

$$|t(K_r, U) - t(K_r, V)| \leq \binom{r}{2} \|U - V\|_{\square}.$$

Consequently,

$$|t(K_r, U) - t(K_r, V)| \leq \binom{r}{2} \delta_{\square}(U, V).$$

In particular, the map $W \mapsto t(K_r, W)$ is continuous with respect to cut distance.

Proof. By [16, Rem. 2.1], every Lebesgue-measurable graphon on $[0, 1]^2$ is equal almost everywhere to a Borel-measurable one. Such a replacement changes neither homomorphism densities nor cut norms. We may therefore replace U and V by Borel-measurable representatives and continue to denote them by U and V . In particular, whenever all but one of the variables are fixed below, the resulting one-variable function is measurable.

Let

$$m := \binom{r}{2}, \quad D := U - V.$$

Choose once and for all an ordering of the $\binom{r}{2}$ edges of K_r . Equivalently, choose an ordering of the pairs

$$(i_1, j_1), \dots, (i_m, j_m), \quad 1 \leq i_{\ell} < j_{\ell} \leq r,$$

so that every pair appears exactly once.

Vertex s carries the single variable $x_s \in [0, 1]$.

The symbols i_{ℓ} and j_{ℓ} record which two vertices form the ℓ -th edge in the chosen list.

For example, when $r = 4$, one possible ordering is

$$\begin{aligned} (i_1, j_1) &= (1, 2), & (i_2, j_2) &= (1, 3), & (i_3, j_3) &= (1, 4), \\ (i_4, j_4) &= (2, 3), & (i_5, j_5) &= (2, 4), & (i_6, j_6) &= (3, 4). \end{aligned}$$

By Remark 3.4,

$$t(K_r, W) = \int_{[0,1]^r} \prod_{\ell=1}^m W(x_{i_\ell}, x_{j_\ell}) dx_1 \cdots dx_r.$$

This means that we choose numbers

$$x_1, \dots, x_r \in [0, 1],$$

one number for each vertex of the clique. For every edge $\{i, j\}$ of the clique, we include the factor $W(x_i, x_j)$. Then we multiply all these factors together. Finally, we average over all choices of $(x_1, \dots, x_r) \in [0, 1]^r$. Since the set $[0, 1]^r$ has total measure 1, the integral here is literally an average.

Applying this formula to U and to V , we obtain

$$t(K_r, U) - t(K_r, V) = \int_{[0,1]^r} \left(\prod_{\ell=1}^m U(x_{i_\ell}, x_{j_\ell}) - \prod_{\ell=1}^m V(x_{i_\ell}, x_{j_\ell}) \right) dx_1 \cdots dx_r.$$

To compare the two products, it is helpful to change the edge factors one at a time. For a fixed point $(x_1, \dots, x_r) \in [0, 1]^r$, use

$$U_\ell := U(x_{i_\ell}, x_{j_\ell}), \quad V_\ell := V(x_{i_\ell}, x_{j_\ell}).$$

Now define

$$P_0 := \prod_{\ell=1}^m V_\ell,$$

and for $s = 1, \dots, m$ define

$$P_s := \left(\prod_{\ell=1}^s U_\ell \right) \left(\prod_{\ell=s+1}^m V_\ell \right).$$

Thus P_0 is the product built entirely from V , while P_m is the product built entirely from U .

Therefore

$$\prod_{\ell=1}^m U_\ell - \prod_{\ell=1}^m V_\ell = P_m - P_0 = \sum_{s=1}^m (P_s - P_{s-1}).$$

For each s , only the s -th factor changes when one passes from P_{s-1} to P_s , so

$$P_s - P_{s-1} = \left(\prod_{\ell < s} U_\ell \right) (U_s - V_s) \left(\prod_{\ell > s} V_\ell \right).$$

Substituting back the definitions of U_ℓ and V_ℓ , we get

$$\prod_{\ell=1}^m U(x_{i_\ell}, x_{j_\ell}) - \prod_{\ell=1}^m V(x_{i_\ell}, x_{j_\ell}) = \sum_{s=1}^m \left(\prod_{\ell < s} U(x_{i_\ell}, x_{j_\ell}) \right) D(x_{i_s}, x_{j_s}) \left(\prod_{\ell > s} V(x_{i_\ell}, x_{j_\ell}) \right).$$

If $s = 1$ or $s = m$, one of the products is empty; by convention, an empty product is equal to 1.

Substituting this into the integral and using the triangle inequality, we obtain

$$|t(K_r, U) - t(K_r, V)| \leq \sum_{s=1}^m I_s,$$

where

$$(3.1) \quad I_s := \left| \int_{[0,1]^r} \left(\prod_{\ell < s} U(x_{i_\ell}, x_{j_\ell}) \right) D(x_{i_s}, x_{j_s}) \left(\prod_{\ell > s} V(x_{i_\ell}, x_{j_\ell}) \right) dx_1 \cdots dx_r \right|.$$

We now estimate one such term.

Fix $s \in \{1, \dots, m\}$ and use

$$(i_s, j_s) = (a, b).$$

Let

$$\hat{x} = (x_t)_{t \in [r] \setminus \{a, b\}}, \quad d\hat{x} := \prod_{t \in [r] \setminus \{a, b\}} dx_t,$$

where the coordinates are taken in increasing order. Thus $\hat{x}$ denotes all variables except x_a and x_b . When $r = 2$, $[0, 1]^0$ is interpreted as a one-point probability space, so integration with respect to $d\hat{x}$ means evaluation at that point.

Because every graphon takes values in $[0, 1]$, the integrand in (3.1) is measurable and bounded in absolute value by 1. Thus it is integrable on $[0, 1]^r$. We can therefore write the integral by first averaging over the two variables x_a and x_b , and then averaging over the remaining $r - 2$ variables, which we collect in $\hat{x}$.

The distinguished edge $\{a, b\}$ contributes the factor

$$D(x_a, x_b).$$

Every other edge of K_r falls into exactly one of the following three types.

An edge of the form $\{a, c\}$ with $c \notin \{a, b\}$ contributes a factor that depends on x_a but not on x_b .

An edge of the form $\{b, c\}$ with $c \notin \{a, b\}$ contributes a factor that depends on x_b but not on x_a .

An edge of the form $\{c, d\}$ with $c, d \notin \{a, b\}$ contributes a factor that depends on neither x_a nor x_b ; once $\hat{x}$ is fixed, it is constant.

Because K_r is a simple graph, there is only one edge joining a and b . So the factor

$$D(x_a, x_b)$$

is the only factor that depends on both variables at the same time.

Figure 1 illustrates this factorization in the case K_4 .

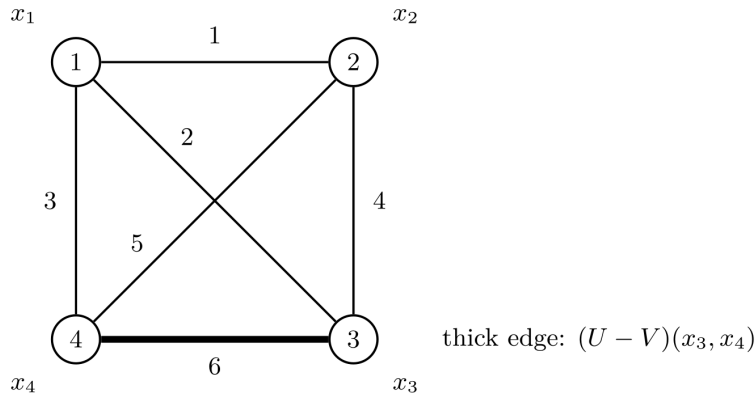


FIGURE 1. In the integral defining $t(K_4, W)$, vertex i carries the variable x_i , and the edge $\{i, j\}$ contributes the factor $W(x_i, x_j)$. If the distinguished edge is $\{3, 4\}$, then that edge contributes the factor $(U - V)(x_3, x_4)$. Once x_1 and x_2 are fixed, the remaining factors split into a function of x_3 , a function of x_4 , and a constant.

Therefore, for each fixed choice of $\hat{x}$, the whole integrand in (3.1) can be written in the form

$$C(\hat{x}) D(x_a, x_b) A_{\hat{x}}(x_a) B_{\hat{x}}(x_b),$$

where $A_{\widehat{x}}, B_{\widehat{x}} : [0, 1] \rightarrow [0, 1]$ are measurable and $C(\widehat{x}) \in [0, 1]$. More concretely, $A_{\widehat{x}}$ is the product of all remaining factors coming from edges that touch a but not b , $B_{\widehat{x}}$ is the product of all remaining factors coming from edges that touch b but not a , and $C(\widehat{x})$ is the product of all remaining factors coming from edges that touch neither a nor b . Empty products are interpreted as 1. These functions take values in $[0, 1]$ because every graphon factor does.

Using this factorization and averaging first in x_a and x_b , we get

$$(3.2) \quad \begin{aligned} I_s &= \left| \int_{[0,1]^{r-2}} \int_{[0,1]^2} C(\widehat{x}) D(x_a, x_b) A_{\widehat{x}}(x_a) B_{\widehat{x}}(x_b) dx_a dx_b d\widehat{x} \right| \\ &\leq \int_{[0,1]^{r-2}} \left| \int_{[0,1]^2} D(x_a, x_b) (C(\widehat{x}) A_{\widehat{x}}(x_a)) B_{\widehat{x}}(x_b) dx_a dx_b \right| d\widehat{x}. \end{aligned}$$

For each fixed $\widehat{x}$, define

$$f_{\widehat{x}}(x_a) := C(\widehat{x}) A_{\widehat{x}}(x_a), \quad g_{\widehat{x}}(x_b) := B_{\widehat{x}}(x_b).$$

Then $f_{\widehat{x}}, g_{\widehat{x}} : [0, 1] \rightarrow [0, 1]$ are measurable, so Lemma 3.13 applies and gives

$$\left| \int_{[0,1]^2} D(x_a, x_b) f_{\widehat{x}}(x_a) g_{\widehat{x}}(x_b) dx_a dx_b \right| \leq \|D\|_{\square}.$$

Substituting this bound into (3.2) yields

$$I_s \leq \int_{[0,1]^{r-2}} \|D\|_{\square} d\widehat{x} = \|D\|_{\square},$$

because the set $[0, 1]^{r-2}$ has total measure 1.

This bound is the same for every $s \in [m]$. Hence

$$|t(K_r, U) - t(K_r, V)| \leq \sum_{s=1}^m I_s \leq \sum_{s=1}^m \|D\|_{\square} = m \|D\|_{\square} = \binom{r}{2} \|U - V\|_{\square}.$$

This proves the first inequality.

We now pass from cut norm to cut distance. Let ϕ be any measure-preserving bijection of $[0, 1]$. By Definition 3.6,

$$V^{\phi}(x, y) = V(\phi(x), \phi(y)).$$

Define

$$\Phi : [0, 1]^r \rightarrow [0, 1]^r, \quad \Phi(x_1, \dots, x_r) := (\phi(x_1), \dots, \phi(x_r)).$$

Since ϕ preserves Lebesgue measure on $[0, 1]$, the map Φ preserves the product measure on $[0, 1]^r$.

Therefore

$$t(K_r, V^{\phi}) = t(K_r, V).$$

Applying the inequality already proved to U and V^{ϕ} , we obtain

$$|t(K_r, U) - t(K_r, V)| = |t(K_r, U) - t(K_r, V^{\phi})| \leq \binom{r}{2} \|U - V^{\phi}\|_{\square}.$$

Now take the infimum over all measure-preserving bijections ϕ . By the definition of $\delta_{\square}$,

$$|t(K_r, U) - t(K_r, V)| \leq \binom{r}{2} \delta_{\square}(U, V).$$

Thus $W \mapsto t(K_r, W)$ is Lipschitz with respect to cut distance, and in particular it is continuous. $\square$

3.2. Entropy. Entropy tells us, on the logarithmic scale, how many ways there are to choose edges and non-edges so that the edge density has a fixed value. The relevant one-variable function is the binary entropy

$$h(p) := -p \ln p - (1-p) \ln(1-p), \quad p \in [0, 1],$$

with the convention $0 \ln 0 := 0$. If pN is an integer, then the number of 0–1 strings of length N with exactly pN ones is $\binom{N}{pN}$, and for large N the logarithm of this number has leading term $Nh(p)$. Thus $h(p)$ measures the entropy contribution per possible edge when the edge density is p . In particular,

$$h(0) = h(1) = 0, \quad 0 \leq h(p) \leq \ln 2, \quad h\left(\frac{1}{2}\right) = \ln 2.$$

So the entropy is zero when the presence or absence of an edge is completely determined, and it is largest at $p = \frac{1}{2}$, where the two possibilities are most evenly balanced.

Definition 3.15 (Graphon entropy). For a graphon W , define

$$(3.3) \quad \text{Ent}(W) := \frac{1}{2} \int_{[0,1]^2} h(W(x, y)) \, dx \, dy.$$

The factor $\frac{1}{2}$ is needed because the square $[0, 1]^2$ counts both (x, y) and (y, x) , while an edge is an unordered pair.

Lemma 3.16. For every $p \in [0, 1]$,

$$(3.4) \quad h(p) \leq \ln 2 - 2\left(p - \frac{1}{2}\right)^2.$$

Proof. Set

$$\psi(p) := \ln 2 - h(p) - 2\left(p - \frac{1}{2}\right)^2.$$

For $p \in (0, 1)$,

$$h'(p) = \ln \frac{1-p}{p},$$

so

$$\psi'(p) = \ln \frac{p}{1-p} - 4\left(p - \frac{1}{2}\right).$$

Also,

$$\psi''(p) = \frac{1}{p(1-p)} - 4 \geq 0,$$

because $p(1-p) \leq 1/4$. Thus ψ is convex on $(0, 1)$ and extends continuously to $[0, 1]$. Since

$$\psi\left(\frac{1}{2}\right) = 0, \quad \psi'\left(\frac{1}{2}\right) = 0,$$

the point $p = 1/2$ is a global minimizer. Hence $\psi(p) \geq 0$ for every $p \in [0, 1]$, which gives (3.4). $\square$

Lemma 3.17. If $U, V \in \mathcal{W}$ satisfy $\delta_{\square}(U, V) = 0$, then

$$\text{Ent}(U) = \text{Ent}(V).$$

Consequently,

$$\text{Ent}(\widetilde{W}) := \text{Ent}(W), \quad \widetilde{W} = \pi(W),$$

is a well-defined function on $\widetilde{\mathcal{W}}$.

Proof. By Remark 3.9, there are measure-preserving maps $\phi, \psi : [0, 1] \rightarrow [0, 1]$ such that $U^\phi = V^\psi$ almost everywhere. The product maps $\phi \times \phi$ and $\psi \times \psi$ preserve λ^2 , hence

$$\text{Ent}(U) = \text{Ent}(U^\phi) = \text{Ent}(V^\psi) = \text{Ent}(V).$$

□

Graphon entropy controls the number of finite graphs with prescribed large-scale structure. More precisely, for a cut-closed family $\mathcal{F}$, the n^2 -scale upper bound is governed by $\sup_{W \in \mathcal{F}} \text{Ent}(W)$, rather than by the entropy of an arbitrarily chosen member of $\mathcal{F}$. In Section 5, Lemma 5.9 uses such an entropy supremum to bound the number of graphs whose empirical graphons lie in a given family. We combine this result with Lemma 4.12 in the proof of Theorem 2.7(ii) to show that the probability of staying a fixed cut distance away from B_q^* decreases exponentially in n^2 . Lemma 4.5 shows that, among graphons with zero K_r -density, entropy is maximized exactly by the graphons at zero cut distance from B_q^* .

Lemma 3.18. *For every $n \in \mathbb{N}$, every graph G with vertex set $[n]$, and every integer $r \geq 2$,*

$$t(K_r, W_G) = \frac{r! K_r(G)}{n^r}.$$

Proof. Let $(a_{ij})_{1 \leq i, j \leq n}$ be the adjacency matrix of G , and let $I_1, \dots, I_n$ be the intervals from Definition 3.2. Splitting $[0, 1]^r$ into the boxes

$$I_{v_1} \times \dots \times I_{v_r}, \quad v_1, \dots, v_r \in [n],$$

gives

$$(3.5) \quad t(K_r, W_G) = \frac{1}{n^r} \sum_{v_1, \dots, v_r \in [n]} \prod_{1 \leq s < t \leq r} a_{v_s, v_t}.$$

Since every entry of the adjacency matrix is either 0 or 1, the summand corresponding to $(v_1, \dots, v_r)$ is equal to 1 exactly when the vertices $v_1, \dots, v_r$ are pairwise distinct and form a copy of K_r in G . Indeed, if $v_s = v_t$ for some $s < t$, then the product contains the factor

$$a_{v_s, v_t} = a_{v_s, v_s} = 0,$$

because G has no loops.

Thus the sum in (3.5) counts the ordered r -tuples obtained by ordering the vertices of copies of K_r in G . Every unordered copy has exactly $r!$ such orderings, and therefore

$$(3.6) \quad \sum_{v_1, \dots, v_r \in [n]} \prod_{1 \leq s < t \leq r} a_{v_s, v_t} = r! K_r(G).$$

Combining (3.5) and (3.6) gives

$$t(K_r, W_G) = \frac{r! K_r(G)}{n^r},$$

as required. □

Remark 3.19. For $n \geq r$, Lemma 3.18 gives

$$t(K_r, W_G) = \frac{n(n-1) \cdots (n-r+1)}{n^r} \frac{K_r(G)}{\binom{n}{r}}.$$

Thus the graphon homomorphism-density normalization differs from the usual clique density based on unordered copies by the factor

$$\frac{n(n-1) \cdots (n-r+1)}{n^r}.$$

For fixed r ,

$$\lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-r+1)}{n^r} = 1.$$

4. ENTROPY MAXIMIZERS AND STABILITY

In this section we define a family of “balanced q -partite graphons” (recall $q := r - 1$) and prove results which relate the distance of a graphon W to a balanced q -partite graphon to the homomorphism density of K_r in W and the entropy of W . In particular, we show that the balanced q -partite graphons are the graphons that maximize entropy among the graphons with K_r -homomorphism density 0. This will be used in the proof of Theorem 2.7 in Section 5.

4.1. The family B_q^* and the balanced Turán graphons.

Definition 4.1. Let $q \geq 2$. The family B_q^* consists of all graphons $W : [0, 1]^2 \rightarrow [0, 1]$ for which there exists a measurable partition

$$[0, 1] = A_1 \sqcup \cdots \sqcup A_q, \quad \lambda(A_i) = \frac{1}{q} \quad \text{for every } i \in [q],$$

such that

$$W(x, y) = 0 \quad \text{for almost every } (x, y) \in A_i \times A_i \quad \text{for every } i,$$

and

$$W(x, y) = \frac{1}{2} \quad \text{for almost every } (x, y) \in A_i \times A_j \quad \text{for all distinct } i, j \in [q].$$

Equivalently, the values of W on the q^2 block products are given by the following block-value matrix:

$$\begin{bmatrix} 0 & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & 0 & \cdots & \frac{1}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2} & \frac{1}{2} & \cdots & 0 \end{bmatrix}$$

with blocks of equal size.

So B_q^* is the graphon analogue of the family of balanced complete q -partite graphs, except that the off-diagonal blocks have value $\frac{1}{2}$ instead of 1.

Notation 4.2 (Balanced complete q -partite graphons). Define

$$\mathcal{U}_q^* := \left\{ \mathbf{1}_{\bigcup_{\substack{1 \leq i, j \leq q \\ i \neq j}} A_i \times A_j} : \begin{array}{l} A_1, \dots, A_q \subseteq [0, 1] \text{ are measurable,} \\ [0, 1] = A_1 \sqcup \cdots \sqcup A_q, \\ \lambda(A_i) = \frac{1}{q} \text{ for every } i \in [q] \end{array} \right\}.$$

These are the balanced complete q -partite $\{0, 1\}$ -valued graphons. By construction,

$$B_q^* = \left\{ W \in \mathcal{W} : W = \frac{1}{2} U \text{ almost everywhere for some } U \in \mathcal{U}_q^* \right\}.$$

Notation 4.3 (Canonical representative of B_q^*). We fix one balanced partition of $[0, 1]$ by setting

$$A_a := \left[\frac{a-1}{q}, \frac{a}{q} \right) \quad \text{for } a = 1, \dots, q-1, \quad A_q := \left[\frac{q-1}{q}, 1 \right].$$

Thus

$$[0, 1] = A_1 \sqcup \cdots \sqcup A_q, \quad \lambda(A_a) = \frac{1}{q} \quad \text{for every } a \in [q].$$

Define the canonical element $W_q^\star \in B_q^\star$ by

$$W_q^\star(x, y) := \begin{cases} 0, & \text{if } x, y \in A_a \text{ for some } a \in [q], \\ \frac{1}{2}, & \text{if } x \in A_a, y \in A_b \text{ for some distinct } a, b \in [q]. \end{cases}$$

Equivalently,

$$W_q^\star = \frac{1}{2} \mathbf{1} \bigcup_{\substack{1 \leq a, b \leq q \\ a \neq b}} A_a \times A_b.$$

Every graphon in $B_q^\star$ is equal almost everywhere to a relabeling of $W_q^\star$. Indeed, suppose that $V \in B_q^\star$ corresponds to a balanced partition $[0, 1] = B_1 \sqcup \dots \sqcup B_q$. For each $i \in [q]$, both B_i and A_i are measurable subsets of $[0, 1]$ with Lebesgue measure $1/q$. Equip these sets with their normalized restricted Lebesgue measures. By [16, Cor. A.11], there is a measure-preserving bijection

$$\phi_i : B_i \rightarrow A_i.$$

Since the same normalization factor q is used on both sets, ϕ_i also preserves the original Lebesgue measure. Pasting $\phi_1, \dots, \phi_q$ gives a measure-preserving bijection $\phi : [0, 1] \rightarrow [0, 1]$ such that

$$\phi(B_i) = A_i \quad \text{up to null sets for every } i \in [q].$$

Therefore

$$V = (W_q^\star)^\phi \quad \text{almost everywhere.}$$

In particular,

$$\delta_\square(V, W_q^\star) = 0 \quad \text{for every } V \in B_q^\star.$$

Define

$$\widetilde{W}_q^\star := \pi(W_q^\star), \quad \widetilde{B}_q^\star := \pi(B_q^\star).$$

Then

$$\widetilde{B}_q^\star = \{\widetilde{W}_q^\star\}.$$

Consequently, for every graphon W ,

$$\delta_\square(W, B_q^\star) = \delta_\square(W, W_q^\star).$$

Lemma 4.4. *Fix $r \geq 3$ and set $q = r - 1$. Let U be a $\{0, 1\}$ -valued graphon and assume*

$$t(K_r, U) = 0.$$

Then

$$\int_{[0,1]^2} U(x, y) dx dy \leq 1 - \frac{1}{q}.$$

Moreover, equality holds if and only if

$$\delta_\square(U, \mathcal{U}_q^\star) = 0.$$

Proof. This is the K_r -free case of the graphon Turán theorem; see [22, Cor. 16.11]. In the present notation, its equality case is exactly the condition

$$\delta_\square(U, \mathcal{U}_q^\star) = 0.$$

□

Lemma 4.5 (Entropy maximizers among graphons of zero K_r -density). *Fix $r \geq 3$ and set $q = r - 1$. Then*

$$\sup\{\text{Ent}(W) : t(K_r, W) = 0\} = \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

A graphon W with $t(K_r, W) = 0$ attains this supremum if and only if

$$\delta_\square(W, B_q^\star) = 0.$$

Equivalently, the reduced maximizer set is the singleton $\tilde{B}_q^\star = \{\tilde{W}_q^\star\}$.

Remark 4.6. Lemma 4.5 can also be deduced from [9, Thm. 8.2]. In that theorem, take $H = K_r$ and $p = 1/2$. Since $\chi(K_r) = r$, the graphon g appearing there is the balanced $(r-1)$ -partite $\{0, 1\}$ -valued graphon. Hence $pg = \frac{1}{2}g$ is a representative of the reduced class $\tilde{W}_q^\star$; equivalently,

$$\pi(pg) = \tilde{W}_q^\star.$$

The off-diagonal support of g has measure $1 - 1/q$. Therefore

$$\text{Ent}\left(\frac{1}{2}g\right) = \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right),$$

which also gives the value of the constrained maximum.

For $p = 1/2$, the Erdős-Rényi rate function $I_{\frac{1}{2}}$, defined in (5.9) below, satisfies

$$I_{\frac{1}{2}}(W) = \frac{\ln 2}{2} - \text{Ent}(W).$$

Thus minimizing $I_{\frac{1}{2}}$ under the constraint $t(K_r, W) = 0$ is the same as maximizing $\text{Ent}(W)$ under the same constraint. We nevertheless give a direct proof below, using the graphon Turán theorem, because it is short and makes the equality cases explicit.

Proof of Lemma 4.5. Let W be a graphon satisfying

$$t(K_r, W) = 0.$$

Set

$$S := \{(x, y) \in [0, 1]^2 : W(x, y) > 0\}, \quad U := \mathbf{1}_S.$$

Since W is measurable and symmetric almost everywhere, the function U is also measurable and symmetric almost everywhere. Also, U only takes the values 0 and 1. Hence U is a graphon.

We first show that U has zero K_r -density. By Remark 3.4, the integrand defining $t(K_r, W)$ is nonnegative. Since $t(K_r, W) = 0$, we have

$$\prod_{1 \leq a < b \leq r} W(x_a, x_b) = 0$$

for almost every $(x_1, \dots, x_r) \in [0, 1]^r$. If

$$\prod_{1 \leq a < b \leq r} U(x_a, x_b) = 1,$$

then every factor $U(x_a, x_b)$ is equal to 1. By the definition of U , this means that every corresponding factor $W(x_a, x_b)$ is strictly positive. Thus

$$\prod_{1 \leq a < b \leq r} W(x_a, x_b) > 0.$$

This can happen only on a null set. Therefore

$$t(K_r, U) = 0.$$

We now bound the entropy of W . Since $W = 0$ on S^c and $h(0) = 0$, Definition 3.15 gives

$$\text{Ent}(W) = \frac{1}{2} \int_S h(W(x, y)) dx dy.$$

By Lemma 3.16, we have $h(p) \leq \ln 2$ for every $p \in [0, 1]$. Hence

$$(4.1) \quad \text{Ent}(W) \leq \frac{\ln 2}{2} \lambda^2(S) = \frac{\ln 2}{2} \int_{[0, 1]^2} U(x, y) dx dy.$$

Since U is a $\{0, 1\}$ -valued graphon and $t(K_r, U) = 0$, Lemma 4.4 gives

$$(4.2) \quad \int_{[0,1]^2} U(x, y) dx dy \leq 1 - \frac{1}{q}.$$

Combining (4.1) and (4.2), we get

$$\text{Ent}(W) \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

Therefore

$$\sup\{\text{Ent}(W) : t(K_r, W) = 0\} \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

Next we show that this upper bound is attained. Consider the canonical graphon $W_q^\star$ from Notation 4.3. The graphon $W_q^\star$ has value 0 on the diagonal blocks and value $1/2$ on the off-diagonal blocks. Since $q = r - 1$, any choice of r points in $[0, 1]$ contains two points in the same canonical block. For that pair, the corresponding factor in the K_r -integrand is 0. Hence

$$(4.3) \quad t(K_r, W_q^\star) = 0.$$

The diagonal blocks have total measure $1/q$, and the off-diagonal blocks have total measure $1 - 1/q$. Using Definition 3.15, together with

$$h(0) = 0 \quad \text{and} \quad h\left(\frac{1}{2}\right) = \ln 2,$$

we obtain

$$(4.4) \quad \text{Ent}(W_q^\star) = \frac{1}{2} \left(1 - \frac{1}{q}\right) h\left(\frac{1}{2}\right) = \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

Thus

$$\sup\{\text{Ent}(W) : t(K_r, W) = 0\} = \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

It remains to identify the graphons that attain this value.

Suppose first that W is an entropy maximizer under the constraint

$$t(K_r, W) = 0.$$

Use the same notation as above,

$$S = \{(x, y) \in [0, 1]^2 : W(x, y) > 0\}, \quad U = \mathbf{1}_S.$$

Since W attains the value in (4.4), both inequalities (4.1) and (4.2) must be equalities.

Equality in (4.2), by the equality case in Lemma 4.4, gives

$$\delta_\square(U, \mathcal{U}_q^\star) = 0.$$

Equality in (4.1) gives

$$h(W(x, y)) = \ln 2 \quad \text{for almost every } (x, y) \in S.$$

By Lemma 3.16, equality $h(p) = \ln 2$ is possible only when $p = 1/2$. Hence

$$W(x, y) = \frac{1}{2} \quad \text{for almost every } (x, y) \in S.$$

Also, by the definition of S , we have $W = 0$ almost everywhere on S^c . Therefore

$$W = \frac{1}{2} U \quad \text{almost everywhere.}$$

By Notation 4.2, the family $B_q^\star$ consists exactly of the graphons $\frac{1}{2}V$ with $V \in \mathcal{U}_q^\star$, up to almost-everywhere equality. Since

$$\delta_\square(U, \mathcal{U}_q^\star) = 0,$$

we get

$$\delta_{\square}\left(\frac{1}{2}U, B_q^{\star}\right) = 0.$$

Indeed, for every $\eta > 0$, choose $V_{\eta} \in \mathcal{U}_q^{\star}$ such that

$$\delta_{\square}(U, V_{\eta}) < 2\eta.$$

Then $\frac{1}{2}V_{\eta} \in B_q^{\star}$, and by homogeneity of the cut norm,

$$\delta_{\square}\left(\frac{1}{2}U, \frac{1}{2}V_{\eta}\right) \leq \frac{1}{2}\delta_{\square}(U, V_{\eta}) < \eta.$$

Since $W = \frac{1}{2}U$ almost everywhere, it follows that

$$\delta_{\square}(W, B_q^{\star}) = 0.$$

Conversely, suppose that

$$\delta_{\square}(W, B_q^{\star}) = 0.$$

By Notation 4.3, this is equivalent to

$$\delta_{\square}(W, W_q^{\star}) = 0.$$

Using Lemma 3.14 and (4.3), we get

$$t(K_r, W) = t(K_r, W_q^{\star}) = 0.$$

Using Lemma 3.17 and (4.4), we get

$$\text{Ent}(W) = \text{Ent}(W_q^{\star}) = \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

Thus W satisfies the constraint $t(K_r, W) = 0$ and attains the maximum entropy.

We have shown that the entropy maximizers are exactly the graphons W satisfying

$$\delta_{\square}(W, B_q^{\star}) = 0.$$

Finally, Notation 4.3 gives

$$\widetilde{B}_q^{\star} = \{\widetilde{W}_q^{\star}\},$$

so the reduced maximizer set is the singleton $\{\widetilde{W}_q^{\star}\}$. □

Remark 4.7. By Lemma 3.18, the reduced mass of $G \in \mathbf{G}_n$ is

$$\exp(-r! w K_r(G)) = \exp(-n^2 L_n t(K_r, W_G)), \quad L_n := w n^{r-2}.$$

The corresponding graphon functional is therefore

$$T_n(W) = -L_n t(K_r, W).$$

For fixed $L > 0$, let

$$\widetilde{\mathcal{F}}_L^{\star} := \pi \left(\arg \max_{W \in \mathcal{W}} \{\text{Ent}(W) - L t(K_r, W)\} \right).$$

For the ERGM with zero edge coefficient and clique coefficient $-L$, [9, Thm. 3.2] gives exponential concentration around $\widetilde{\mathcal{F}}_L^{\star}$. The concentration rate is not given uniformly in L . Moreover, [9, Thm. 7.1] gives

$$\lim_{L \rightarrow \infty} \sup_{\widetilde{W} \in \widetilde{\mathcal{F}}_L^{\star}} d_{\square}(\widetilde{W}, \widetilde{W}_q^{\star}) = 0.$$

These results describe an iterated limit in which one first lets $n \rightarrow \infty$ with L fixed and then lets $L \rightarrow \infty$. In the fixed-weight MLN considered here,

$$L_n = w n^{r-2}, \quad \lim_{n \rightarrow \infty} L_n = +\infty.$$

Thus n and L_n grow together. Since the concentration rate for fixed L is not uniform in L , the fixed-coefficient theorem does not directly give a bound of the form e^{-cn^2} with $c > 0$ independent of n along this sequence.

4.2. Entropy gap away from the maximizers. The next two lemmas show that once we stay a fixed cut distance away from the reduced maximizer class represented by B_q^* , the entropy drops.

Lemma 4.8. *Fix $r \geq 3$ and set $q = r - 1$. For every $\varepsilon > 0$ there exists $\gamma_r(\varepsilon) > 0$ such that the following holds.*

If U is a $\{0, 1\}$ -valued graphon and satisfies

$$t(K_r, U) = 0, \quad \delta_\square(U, \mathcal{U}_q^*) \geq \varepsilon,$$

then

$$\int_{[0,1]^2} U \leq \left(1 - \frac{1}{q}\right) - \gamma_r(\varepsilon).$$

Proof. Fix $\varepsilon > 0$ and define

$$m_\varepsilon := \sup \left\{ \int_{[0,1]^2} U : \begin{array}{l} U \text{ is a } \{0, 1\}\text{-valued graphon,} \\ t(K_r, U) = 0, \\ \delta_\square(U, \mathcal{U}_q^*) \geq \varepsilon \end{array} \right\}.$$

If the set inside the supremum is empty, take

$$\gamma_r(\varepsilon) = \frac{1}{2} \left(1 - \frac{1}{q}\right).$$

The conclusion then holds automatically. Hence assume that the set is nonempty.

By Lemma 4.4,

$$m_\varepsilon \leq 1 - \frac{1}{q}.$$

We claim that the inequality is strict.

Assume instead that

$$m_\varepsilon = 1 - \frac{1}{q}.$$

By the definition of the supremum, there exists a sequence (U_n) of $\{0, 1\}$ -valued graphons such that

$$t(K_r, U_n) = 0, \quad \delta_\square(U_n, \mathcal{U}_q^*) \geq \varepsilon, \quad \lim_{n \rightarrow \infty} \int_{[0,1]^2} U_n = 1 - \frac{1}{q}.$$

Let $\tilde{U}_n \in \tilde{\mathcal{W}}$ be the equivalence class of U_n . By Notation 3.8, the metric space $(\tilde{\mathcal{W}}, d_\square)$ is compact; see [22, Thm. 9.23]. Hence, after passing to a subsequence, we may assume that

$$\lim_{n \rightarrow \infty} d_\square(\tilde{U}_n, \tilde{U}) = 0$$

for some $\tilde{U} \in \tilde{\mathcal{W}}$. Choose any representative $U \in \tilde{U}$. Since the metric on $\tilde{\mathcal{W}}$ is induced by $\delta_\square$, this means exactly that

$$\lim_{n \rightarrow \infty} \delta_\square(U_n, U) = 0.$$

By Lemma 3.14, the map $t(K_r, \cdot)$ is continuous with respect to cut distance. Therefore

$$\lim_{n \rightarrow \infty} t(K_r, U_n) = t(K_r, U) = 0.$$

Also, the limit still stays at least ε away from $\mathcal{U}_q^*$. Indeed,

$$\delta_\square(U, \mathcal{U}_q^*) \geq \delta_\square(U_n, \mathcal{U}_q^*) - \delta_\square(U_n, U),$$

and letting $n \rightarrow \infty$ gives

$$(4.5) \quad \delta_{\square}(U, \mathcal{U}_q^{\star}) \geq \varepsilon.$$

We also need the edge density to pass to the limit. For graphons W and V , and for any measure-preserving bijection ϕ ,

$$\int_{[0,1]^2} V = \int_{[0,1]^2} V^{\phi}$$

by a change of variables. Therefore

$$\left| \int_{[0,1]^2} (W - V) \right| = \left| \int_{[0,1]^2} (W - V^{\phi}) \right|.$$

By Definition 3.5, we may choose $S = T = [0, 1]$ in the supremum defining the cut norm, so

$$\left| \int_{[0,1]^2} (W - V^{\phi}) \right| \leq \|W - V^{\phi}\|_{\square}.$$

Taking the infimum over all measure-preserving bijections ϕ gives

$$\left| \int_{[0,1]^2} (W - V) \right| \leq \delta_{\square}(W, V).$$

Apply this to $(W, V) = (U_n, U)$ and let $n \rightarrow \infty$:

$$(4.6) \quad \int_{[0,1]^2} U = \lim_{n \rightarrow \infty} \int_{[0,1]^2} U_n = 1 - \frac{1}{q}.$$

Now define the positivity set of U by

$$S := \{(x, y) \in [0, 1]^2 : U(x, y) > 0\}, \quad V := \mathbf{1}_S.$$

The function V is measurable, $\{0, 1\}$ -valued, and symmetric almost everywhere; hence it is a graphon. Moreover,

$$0 \leq U \leq V \leq 1.$$

We show first that V is a graphon of zero K_r -density. If

$$\prod_{1 \leq a < b \leq r} V(x_a, x_b) = 1,$$

then each factor $U(x_a, x_b)$ is strictly positive, so

$$\prod_{1 \leq a < b \leq r} U(x_a, x_b) > 0.$$

But the integrand defining $t(K_r, U)$ is nonnegative and has integral 0. So this can happen only on a set of measure 0. Hence

$$t(K_r, V) = 0.$$

Because $U \leq V$, equation (4.6) gives

$$\int_{[0,1]^2} V \geq \int_{[0,1]^2} U = 1 - \frac{1}{q}.$$

On the other hand, V is a $\{0, 1\}$ -valued graphon of zero K_r -density, so Lemma 4.4 gives

$$\int_{[0,1]^2} V \leq 1 - \frac{1}{q}.$$

Thus

$$\int_{[0,1]^2} V = 1 - \frac{1}{q}.$$

By the equality case in Lemma 4.4, this implies

$$\delta_{\square}(V, \mathcal{U}_q^*) = 0.$$

Finally,

$$V - U \geq 0$$

pointwise and

$$\int_{[0,1]^2} (V - U) = \int_{[0,1]^2} V - \int_{[0,1]^2} U = 0.$$

A nonnegative integrable function with integral 0 is zero almost everywhere. So

$$U = V \quad \text{almost everywhere.}$$

Therefore

$$\delta_{\square}(U, \mathcal{U}_q^*) = 0,$$

contradicting the earlier bound $\delta_{\square}(U, \mathcal{U}_q^*) \geq \varepsilon$ in (4.5).

This contradiction shows that

$$m_{\varepsilon} < 1 - \frac{1}{q}.$$

Now define

$$\gamma_r(\varepsilon) := \left(1 - \frac{1}{q}\right) - m_{\varepsilon} > 0.$$

This is exactly the desired gap. $\square$

Lemma 4.9 (Entropy gap away from B_q^*). *Fix $r \geq 3$ and set $q = r - 1$. For every $\varepsilon > 0$ there exists $c_{r,\varepsilon} > 0$ such that every graphon W satisfying*

$$t(K_r, W) = 0, \quad \delta_{\square}(W, B_q^*) \geq \varepsilon$$

also satisfies

$$\text{Ent}(W) \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right) - c_{r,\varepsilon}.$$

Proof. Fix $\varepsilon > 0$ and let W be a graphon with

$$t(K_r, W) = 0, \quad \delta_{\square}(W, B_q^*) \geq \varepsilon.$$

As before, set

$$S := \{(x, y) \in [0, 1]^2 : W(x, y) > 0\}, \quad U := \mathbf{1}_S.$$

The function U is measurable, $\{0, 1\}$ -valued, and symmetric almost everywhere; hence it is a graphon.

Since $t(K_r, W) = 0$, Remark 3.4 implies that

$$\prod_{1 \leq a < b \leq r} W(x_a, x_b) = 0$$

for almost every $(x_1, \dots, x_r) \in [0, 1]^r$. If

$$\prod_{1 \leq a < b \leq r} U(x_a, x_b) = 1,$$

then every factor $W(x_a, x_b)$ is strictly positive, and therefore

$$\prod_{1 \leq a < b \leq r} W(x_a, x_b) > 0.$$

This can happen only on a null set. Hence

$$t(K_r, U) = 0.$$

Moreover, $W = 0$ on S^c , while $h(u) \leq \ln 2$ for every $u \in [0, 1]$. Therefore

$$\text{Ent}(W) = \frac{1}{2} \int_S h(W(x, y)) dx dy \leq \frac{\ln 2}{2} \lambda^2(S) = \frac{\ln 2}{2} \int_{[0,1]^2} U.$$

We now split into two cases.

Case 1: $\delta_\square(U, \mathcal{U}_q^\star) \geq \varepsilon/2$.

Then Lemma 4.8 gives

$$\int_{[0,1]^2} U \leq \left(1 - \frac{1}{q}\right) - \gamma_r(\varepsilon/2).$$

Hence

$$\text{Ent}(W) \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right) - \frac{\ln 2}{2} \gamma_r(\varepsilon/2).$$

Thus, in Case 1, the claimed bound holds with entropy loss $\frac{\ln 2}{2} \gamma_r(\varepsilon/2)$.

Case 2: $\delta_\square(U, \mathcal{U}_q^\star) < \varepsilon/2$.

Choose $U^\star \in \mathcal{U}_q^\star$ such that

$$\delta_\square(U, U^\star) < \frac{\varepsilon}{2}.$$

Set

$$V^\star := \frac{1}{2} U^\star.$$

By Notation 4.2, the family $B_q^\star$ consists exactly of graphons that are equal almost everywhere to $\frac{1}{2}U$ for some $U \in \mathcal{U}_q^\star$. Since $U^\star \in \mathcal{U}_q^\star$ and $V^\star = \frac{1}{2}U^\star$, it follows that $V^\star \in B_q^\star$. Since $\delta_\square(W, B_q^\star) \geq \varepsilon$, we have

$$\delta_\square(W, V^\star) \geq \varepsilon.$$

Also, by homogeneity of the cut distance,

$$\delta_\square\left(\frac{1}{2}U, V^\star\right) = \delta_\square\left(\frac{1}{2}U, \frac{1}{2}U^\star\right) = \frac{1}{2} \delta_\square(U, U^\star) < \frac{\varepsilon}{4}.$$

So the triangle inequality gives

$$\delta_\square\left(W, \frac{1}{2}U\right) \geq \delta_\square(W, V^\star) - \delta_\square\left(\frac{1}{2}U, V^\star\right) > \frac{3\varepsilon}{4}.$$

Since

$$\delta_\square\left(W, \frac{1}{2}U\right) \leq \left\|W - \frac{1}{2}U\right\|_\square \leq \left\|W - \frac{1}{2}U\right\|_1,$$

it follows that

$$\left\|W - \frac{1}{2}U\right\|_1 > \frac{3\varepsilon}{4}.$$

Because $U = \mathbf{1}_S$ and $W = 0$ on S^c , this means

$$\int_S \left|W - \frac{1}{2}\right| > \frac{3\varepsilon}{4}.$$

Since U is $\{0, 1\}$ -valued and is a graphon of zero K_r -density, Lemma 4.4 gives

$$\lambda^2(S) = \int_{[0,1]^2} U \leq 1 - \frac{1}{q}.$$

By Lemma 3.12(i),

$$\left(\int_S \left|W - \frac{1}{2}\right|\right)^2 \leq \lambda^2(S) \int_S \left(W - \frac{1}{2}\right)^2.$$

Therefore

$$\int_S \left(W - \frac{1}{2}\right)^2 \geq \frac{\left(\frac{3\varepsilon}{4}\right)^2}{1 - \frac{1}{q}}.$$

For every $(x, y) \in S$, we also have $W(x, y) \in [0, 1]$. Hence Lemma 3.16 gives

$$h(W(x, y)) \leq \ln 2 - 2 \left(W(x, y) - \frac{1}{2} \right)^2 \quad \text{for every } (x, y) \in S.$$

Using this inequality in the entropy identity above, we obtain

$$\begin{aligned} \text{Ent}(W) &\leq \frac{1}{2} \int_S \left(\ln 2 - 2 \left(W(x, y) - \frac{1}{2} \right)^2 \right) dx dy \\ &= \frac{\ln 2}{2} \lambda^2(S) - \int_S \left(W(x, y) - \frac{1}{2} \right)^2 dx dy \\ &\leq \frac{\ln 2}{2} \left(1 - \frac{1}{q} \right) - \frac{9}{16} \frac{\varepsilon^2}{(1 - \frac{1}{q})}. \end{aligned}$$

Here the last inequality uses the two estimates already proved above:

$$\lambda^2(S) \leq 1 - \frac{1}{q} \quad \text{and} \quad \int_S \left(W(x, y) - \frac{1}{2} \right)^2 dx dy \geq \frac{9}{16} \frac{\varepsilon^2}{(1 - \frac{1}{q})}.$$

Combine the two cases and define

$$c_{r,\varepsilon} := \min \left\{ \frac{\ln 2}{2} \gamma_r(\varepsilon/2), \frac{9}{16} \frac{\varepsilon^2}{(1 - \frac{1}{q})} \right\} > 0.$$

Then

$$\text{Ent}(W) \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q} \right) - c_{r,\varepsilon}.$$

□

The graphon removal lemma in [14, Lem. 2.3] is stated with the strict assumption $t(K_r, W) < \delta$. We will use the following non-strict version, obtained by applying that lemma with a smaller threshold.

Lemma 4.10 (Graphon K_r removal). *Fix $r \geq 3$. For every $\eta > 0$ there exists $\delta = \delta_{K_r}(\eta) > 0$ such that whenever a graphon W satisfies*

$$t(K_r, W) \leq \delta,$$

there exists a graphon W' with

$$t(K_r, W') = 0, \quad \|W - W'\|_1 \leq \eta.$$

Proof. Apply [14, Lem. 2.3] with $F = K_r$ and error η . This gives a number $\delta_0 > 0$ such that, whenever

$$t(K_r, W) < \delta_0,$$

there is a graphon W' satisfying

$$t(K_r, W') = 0 \quad \text{and} \quad \|W - W'\|_1 \leq \eta.$$

Now define

$$\delta_{K_r}(\eta) := \frac{\delta_0}{2}.$$

If $t(K_r, W) \leq \delta_{K_r}(\eta)$, then

$$t(K_r, W) \leq \frac{\delta_0}{2} < \delta_0.$$

Thus W satisfies the strict hypothesis of [14, Lem. 2.3], applied with $F = K_r$ and error η . Hence there exists a graphon W' such that

$$t(K_r, W') = 0 \quad \text{and} \quad \|W - W'\|_1 \leq \eta.$$

This is the required graphon.

□

Lemma 4.11 (Entropy is continuous in L^1). *For graphons U and V , and for every $\alpha > 0$,*

$$(4.7) \quad |\text{Ent}(U) - \text{Ent}(V)| \leq \frac{1}{2} \omega(\alpha) + \frac{\ln 2}{2} \frac{\|U - V\|_1}{\alpha},$$

where

$$\omega(\alpha) := \sup\{|h(x) - h(y)| : x, y \in [0, 1], |x - y| \leq \alpha\}.$$

In particular, Ent is uniformly continuous on $\mathcal{W}$ with respect to $\|\cdot\|_1$. For every $\tau > 0$ there exists $\eta > 0$ such that, for all graphons U and V ,

$$\|U - V\|_1 \leq \eta \quad \implies \quad |\text{Ent}(U) - \text{Ent}(V)| \leq \tau.$$

Proof. Fix $\alpha > 0$ and define the two sets

$$A := \{(x, y) \in [0, 1]^2 : |U(x, y) - V(x, y)| \leq \alpha\},$$

$$B := \{(x, y) \in [0, 1]^2 : |U(x, y) - V(x, y)| > \alpha\}.$$

Then

$$\begin{aligned} |\text{Ent}(U) - \text{Ent}(V)| &= \frac{1}{2} \left| \int_{[0,1]^2} (h(U) - h(V)) \right| \\ &\leq \frac{1}{2} \int_A |h(U) - h(V)| + \frac{1}{2} \int_B |h(U) - h(V)|. \end{aligned}$$

On the set A we have

$$|h(U) - h(V)| \leq \omega(\alpha),$$

so

$$\int_A |h(U) - h(V)| \leq \omega(\alpha).$$

On the set B we use only the bound $0 \leq h \leq \ln 2$, and since $B \subseteq [0, 1]^2$, we get

$$\int_B |h(U) - h(V)| \leq (\ln 2) \lambda^2(B).$$

Also, since $|U - V| > \alpha$ on B ,

$$\|U - V\|_1 \geq \alpha \lambda^2(B), \quad \text{so} \quad \lambda^2(B) \leq \frac{\|U - V\|_1}{\alpha}.$$

Putting these estimates together yields (4.7).

Finally, h is uniformly continuous on the compact interval $[0, 1]$, and therefore

$$\lim_{\alpha \rightarrow 0} \omega(\alpha) = 0.$$

Fix $\tau > 0$. Choose $\alpha > 0$ such that

$$\frac{1}{2} \omega(\alpha) \leq \frac{\tau}{2},$$

and then choose $\eta > 0$ such that

$$\frac{\ln 2}{2} \frac{\eta}{\alpha} \leq \frac{\tau}{2}.$$

If $\|U - V\|_1 \leq \eta$, then (4.7) gives

$$|\text{Ent}(U) - \text{Ent}(V)| \leq \tau.$$

This proves the uniform continuity statement. $\square$

Lemma 4.12. *Fix $r \geq 3$ and set $q = r - 1$. For every $\varepsilon > 0$ there exist $\delta_{r,\varepsilon} > 0$ and $c'_{r,\varepsilon} > 0$ such that every graphon W satisfying*

$$t(K_r, W) \leq \delta_{r,\varepsilon} \quad \text{and} \quad \delta_{\square}(W, B_q^{\star}) \geq \varepsilon,$$

also satisfies

$$\text{Ent}(W) \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right) - c'_{r,\varepsilon}.$$

Proof. Fix $\varepsilon > 0$. Apply Lemma 4.9 with $\varepsilon/2$. This gives a constant $c_{r,\varepsilon/2} > 0$ such that every V that is a graphon of zero K_r -density and satisfies

$$\delta_{\square}(V, B_q^{\star}) \geq \frac{\varepsilon}{2}$$

satisfies

$$\text{Ent}(V) \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right) - c_{r,\varepsilon/2}.$$

Choose $\eta > 0$ so small that

$$\eta \leq \frac{\varepsilon}{2}$$

and

$$\sup_{\substack{U, V \in \mathcal{W} \\ \|U - V\|_1 \leq \eta}} |\text{Ent}(U) - \text{Ent}(V)| \leq \frac{1}{2} c_{r,\varepsilon/2}.$$

This is possible by Lemma 4.11.

Now apply Lemma 4.10 with this η . We obtain $\delta_{r,\varepsilon} > 0$ such that every graphon W with

$$t(K_r, W) \leq \delta_{r,\varepsilon}$$

admits a graphon W' of zero K_r -density satisfying

$$\|W - W'\|_1 \leq \eta.$$

Suppose that

$$t(K_r, W) \leq \delta_{r,\varepsilon} \quad \text{and} \quad \delta_{\square}(W, B_q^{\star}) \geq \varepsilon.$$

Choose a corresponding graphon W' such that $\|W - W'\|_1 \leq \eta$. Since

$$\delta_{\square}(W, W') \leq \|W - W'\|_{\square} \leq \|W - W'\|_1 \leq \eta,$$

the triangle inequality gives

$$\delta_{\square}(W', B_q^{\star}) \geq \delta_{\square}(W, B_q^{\star}) - \delta_{\square}(W, W') \geq \varepsilon - \eta \geq \frac{\varepsilon}{2}.$$

So Lemma 4.9 implies

$$\text{Ent}(W') \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right) - c_{r,\varepsilon/2}.$$

Also,

$$|\text{Ent}(W) - \text{Ent}(W')| \leq \frac{1}{2} c_{r,\varepsilon/2}$$

by our choice of η . Therefore

$$\text{Ent}(W) \leq \text{Ent}(W') + \frac{1}{2} c_{r,\varepsilon/2} \leq \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right) - \frac{1}{2} c_{r,\varepsilon/2}.$$

Thus the lemma holds with

$$c'_{r,\varepsilon} := \frac{1}{2} c_{r,\varepsilon/2} > 0.$$

□

5. PROOF OF THEOREM 2.7

Recall that we have fixed $r \geq 3$ and $w > 0$ and we continue to let $q := r - 1$. We first prove some technical results in sections 5.1 – 5.3. Then we prove parts (i), (ii), and (iii) of Theorem 2.7, in this order, in sections 5.4 – 5.7.

Notation 5.1 (Balanced q -partite entropy level). We write

$$(5.1) \quad s_q := \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

Corollary 2.6 shows that the graph-independent factor e^{wn^r} disappears after normalization. Thus the probability measure $\mathbb{P}_{n,r}$ is determined by the reduced masses

$$\exp(-r! w K_r(G)), \quad G \in \mathbf{G}_n.$$

By Lemma 3.18,

$$t(K_r, W_G) = \frac{r! K_r(G)}{n^r}.$$

Consequently,

$$(5.2) \quad \exp(-r! w K_r(G)) = \exp(-w t(K_r, W_G) n^r).$$

This identity connects the finite-graph penalty with the graphon arguments in this section.

5.1. A general vanishing lemma. The next lemma records the counting estimate that the graphon K_r -homomorphism density $t(K_r, W_{\mathbf{G}_n}) = r! K_r(\mathbf{G}_n)/n^r$ goes to zero in probability for every fixed $r \geq 3$ and every fixed $w > 0$, where $\mathbf{G}_n \sim \mathbb{P}_{n,r}$.

Lemma 5.2. Fix $r \geq 3$ and $w > 0$. For every $\varepsilon > 0$ and every $n \in \mathbb{N}$, define

$$\mathcal{A}_{n,\varepsilon}^{(r)} := \{G \in \mathbf{G}_n : K_r(G) \geq \varepsilon n^r\}.$$

Then

$$\mathbb{P}_{n,r}(\mathcal{A}_{n,\varepsilon}^{(r)}) \leq |\mathbf{G}_n| e^{-r! w \varepsilon n^r} = \exp\left(\binom{n}{2} \ln 2 - r! w \varepsilon n^r\right).$$

In particular, for every $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n,r}(\mathcal{A}_{n,\varepsilon}^{(r)}) = 0.$$

Equivalently, if $\mathbf{G}_n \sim \mathbb{P}_{n,r}$, then for every $\delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(t(K_r, W_{\mathbf{G}_n}) \geq \delta) = 0,$$

because this event is $\{\mathbf{G}_n \in \mathcal{A}_{n,\delta/r!}^{(r)}\}$. Thus $t(K_r, W_{\mathbf{G}_n})$ converges to 0 in probability.

Proof. By (2.3),

$$\mathbb{P}_{n,r}(G) = \frac{e^{-r! w K_r(G)}}{\tilde{Z}_{n,r}}.$$

The denominator is at least 1, because a graph with no edges has no copy of K_r and contributes 1. So for every event $\mathcal{A} \subseteq \mathbf{G}_n$,

$$\mathbb{P}_{n,r}(\mathcal{A}) = \sum_{G \in \mathcal{A}} \mathbb{P}_{n,r}(G) \leq \sum_{G \in \mathcal{A}} e^{-r! w K_r(G)}.$$

Apply this with $\mathcal{A} = \mathcal{A}_{n,\varepsilon}^{(r)}$. On this event we have $K_r(G) \geq \varepsilon n^r$, hence

$$\mathbb{P}_{n,r}(\mathcal{A}_{n,\varepsilon}^{(r)}) \leq |\mathcal{A}_{n,\varepsilon}^{(r)}| e^{-r! w \varepsilon n^r} \leq |\mathbf{G}_n| e^{-r! w \varepsilon n^r}.$$

Since $|\mathbf{G}_n| = 2^{\binom{n}{2}}$, we have

$$|\mathbf{G}_n| e^{-r! w \varepsilon n^r} = \exp\left(\left(\binom{n}{2} \ln 2 - r! w \varepsilon n^r\right)\right).$$

Because $r \geq 3$, the term $-r! w \varepsilon n^r$ is of strictly higher order than $\binom{n}{2} \ln 2 = O(n^2)$, so the exponent has limit $-\infty$. $\square$

5.2. Graphs with a fixed positive clique density. A fixed positive lower bound on $t(K_r, W_G)$ gives a negative term of order n^r in the exponent. The total number of labeled graphs is only $\exp(O(n^2))$. Since $r \geq 3$, the penalty is stronger than this counting term.

Lemma 5.3 (Fixed positive clique density is negligible). *Fix $0 < \delta \leq 1$. For every $M > 0$, there exists $N = N(M, \delta, r, w)$ such that, for every $n \geq N$,*

$$(5.3) \quad \sum_{\substack{G \in \mathbf{G}_n \\ t(K_r, W_G) \geq \delta}} \exp(-r! w K_r(G)) \leq \exp(-Mn^2).$$

Proof. If $t(K_r, W_G) \geq \delta$, then (5.2) gives

$$\exp(-r! w K_r(G)) \leq \exp(-w\delta n^r).$$

Since $|\mathbf{G}_n| = 2^{\binom{n}{2}}$, the total contribution of graphs satisfying $t(K_r, W_G) \geq \delta$ is bounded by

$$(5.4) \quad \sum_{\substack{G \in \mathbf{G}_n \\ t(K_r, W_G) \geq \delta}} \exp(-r! w K_r(G)) \leq \exp(E_n),$$

where

$$(5.5) \quad E_n := \binom{n}{2} \ln 2 - w\delta n^r.$$

We now compare E_n with the n^2 scale. By (5.5),

$$\frac{E_n}{n^2} = \frac{\binom{n}{2}}{n^2} \ln 2 - w\delta n^{r-2}.$$

Since

$$\frac{\binom{n}{2}}{n^2} = \frac{n(n-1)}{2n^2} = \frac{1}{2} - \frac{1}{2n},$$

we get

$$(5.6) \quad \frac{E_n}{n^2} = \frac{\ln 2}{2} - \frac{\ln 2}{2n} - w\delta n^{r-2}.$$

Because $r \geq 3$ and $w\delta > 0$,

$$\lim_{n \rightarrow \infty} \left(\frac{\ln 2}{2} - \frac{\ln 2}{2n} - w\delta n^{r-2} \right) = -\infty.$$

Therefore, by (5.6), there is $N = N(M, \delta, r, w)$ such that, for every $n \geq N$,

$$\frac{E_n}{n^2} \leq -M.$$

Equivalently,

$$E_n \leq -Mn^2.$$

Combining this with (5.4), we obtain, for every $n \geq N$,

$$\sum_{\substack{G \in \mathbf{G}_n \\ t(K_r, W_G) \geq \delta}} \exp(-r! w K_r(G)) \leq \exp(E_n) \leq \exp(-Mn^2).$$

This proves the lemma. $\square$

Thus graphs with $t(K_r, W_G) \geq \delta$ have negligible total reduced mass on the n^2 scale, for every fixed $\delta > 0$. The main contribution must come from graphs with small empirical K_r -density. Lemma 4.5 identifies the largest entropy at zero K_r -density, and Lemma 4.12 gives a strict entropy loss away from B_q^* when the K_r -density is sufficiently small.

5.3. Counting cut-closed families. We next convert entropy bounds for graphons into counting bounds for labeled graphs.

Definition 5.4 (Cut-closed family of graphons). A family $\mathcal{F} \subseteq \mathcal{W}$ is called *cut-closed* if, whenever $W_m \in \mathcal{F}$ for every m and

$$\lim_{m \rightarrow \infty} \delta_{\square}(W_m, W) = 0,$$

then $W \in \mathcal{F}$.

Lemma 5.5. *Let $\mathcal{F} \subseteq \mathcal{W}$ be cut-closed, and let $\tilde{\mathcal{F}}$ be its image in $\tilde{\mathcal{W}}$. Then the following statements hold.*

- (1) *If $W \in \mathcal{F}$, $V \in \mathcal{W}$, and $\delta_{\square}(W, V) = 0$, then $V \in \mathcal{F}$.*
- (2) *For every $n \in \mathbb{N}$ and every $G \in \mathbf{G}_n$,*

$$W_G \in \mathcal{F} \iff \tilde{W}_G \in \tilde{\mathcal{F}}.$$

- (3) *The set $\tilde{\mathcal{F}}$ is closed in the metric $d_{\square}$.*

Proof. For the first statement, take the constant sequence $W_m = W$. Then

$$\lim_{m \rightarrow \infty} \delta_{\square}(W_m, V) = 0,$$

so cut-closedness gives $V \in \mathcal{F}$.

The forward implication in the second statement follows from the definition of $\tilde{\mathcal{F}}$. Conversely, suppose that $\tilde{W}_G \in \tilde{\mathcal{F}}$. Then some $V \in \mathcal{F}$ satisfies $\delta_{\square}(W_G, V) = 0$. The first statement gives $W_G \in \mathcal{F}$.

For the third statement, let $\tilde{W}_m \in \tilde{\mathcal{F}}$ and suppose that

$$\lim_{m \rightarrow \infty} d_{\square}(\tilde{W}_m, \tilde{W}) = 0.$$

Choose a representative W of $\tilde{W}$. For each m , choose a representative $V_m \in \mathcal{F}$ of $\tilde{W}_m$. By the definition of $d_{\square}$,

$$\delta_{\square}(V_m, W) = d_{\square}(\tilde{W}_m, \tilde{W}).$$

Hence

$$\lim_{m \rightarrow \infty} \delta_{\square}(V_m, W) = 0.$$

Cut-closedness gives $W \in \mathcal{F}$, and therefore $\tilde{W} \in \tilde{\mathcal{F}}$. □

Notation 5.6 (Uniform measure on labeled graphs). For each $n \in \mathbb{N}$, let $\mathbb{P}_n^{\text{unif}}$ be the uniform probability measure on $\mathbf{G}_n$. Since $\mathbf{G}_n$ contains $2^{\binom{n}{2}}$ graphs, we have

$$(5.7) \quad \mathbb{P}_n^{\text{unif}}(A) = 2^{-\binom{n}{2}} |A|, \quad A \subseteq \mathbf{G}_n.$$

We consider $\mathbf{G}_n$ as a probability space equipped with $\mathbb{P}_n^{\text{unif}}$, and let R_n be the canonical random graph on this space; thus

$$R_n(G) = G \quad \text{for every } G \in \mathbf{G}_n.$$

Equivalently, each possible edge is present independently with probability $\frac{1}{2}$. Hence

$$R_n \sim G\left(n, \frac{1}{2}\right).$$

For every $\tilde{\mathcal{A}} \subseteq \tilde{\mathcal{W}}$, the event $\tilde{W}_{R_n} \in \tilde{\mathcal{A}}$ consists exactly of the graphs whose empirical graphon class belongs to $\tilde{\mathcal{A}}$. Hence

$$(5.8) \quad \mathbb{P}_n^{\text{unif}}(\tilde{W}_{R_n} \in \tilde{\mathcal{A}}) = \mathbb{P}_n^{\text{unif}}(\{G \in \mathbf{G}_n : \tilde{W}_G \in \tilde{\mathcal{A}}\}).$$

Theorem 5.7. *For a graphon $W \in \mathcal{W}$, define*

$$(5.9) \quad I_{\frac{1}{2}}(W) := \frac{1}{2} \int_{[0,1]^2} [W \ln(2W) + (1-W) \ln(2(1-W))] \, dx \, dy,$$

where $0 \ln 0 = 0$.

The functional $I_{\frac{1}{2}}$ is constant on zero-cut-distance classes. Thus, if $\tilde{W} = \pi(W)$, then

$$I_{\frac{1}{2}}(\tilde{W}) := I_{\frac{1}{2}}(W)$$

is well-defined. Moreover,

$$(5.10) \quad I_{\frac{1}{2}}(\tilde{W}) = \frac{\ln 2}{2} - \text{Ent}(\tilde{W}).$$

If $\tilde{\mathcal{F}} \subseteq \tilde{\mathcal{W}}$ is closed, then

$$(5.11) \quad \limsup_{n \rightarrow \infty} \frac{1}{n^2} \ln \mathbb{P}_n^{\text{unif}} \left(\left\{ G \in \mathbf{G}_n : \tilde{W}_G \in \tilde{\mathcal{F}} \right\} \right) \leq - \inf_{\tilde{W} \in \tilde{\mathcal{F}}} I_{\frac{1}{2}}(\tilde{W}).$$

Here $\inf \emptyset = +\infty$, $\sup \emptyset = -\infty$, and $\ln 0 = -\infty$.

Proof. Recall that the binary entropy function is defined by

$$h(u) = -u \ln u - (1-u) \ln(1-u), \quad u \in [0, 1],$$

with the convention $0 \ln 0 = 0$. For $0 < u < 1$, we have

$$\begin{aligned} u \ln(2u) + (1-u) \ln(2(1-u)) &= u(\ln 2 + \ln u) + (1-u)(\ln 2 + \ln(1-u)) \\ &= (u + (1-u)) \ln 2 + u \ln u + (1-u) \ln(1-u) \\ &= \ln 2 - h(u). \end{aligned}$$

The same identity holds for $u = 0$ and $u = 1$ under the convention $0 \ln 0 = 0$. Hence it holds for every $u \in [0, 1]$. Integrating this identity gives

$$I_{\frac{1}{2}}(W) = \frac{\ln 2}{2} - \text{Ent}(W).$$

By Lemma 3.17, entropy is unchanged under zero cut distance. Hence $I_{\frac{1}{2}}$ is also unchanged under zero cut distance, so $I_{\frac{1}{2}}(\tilde{W})$ is well-defined and (5.10) follows.

Under the measure $\mathbb{P}_n^{\text{unif}}$ from Notation 5.6, the random graph R_n has distribution $G(n, \frac{1}{2})$. We may therefore apply [10, Thm. 2.3] with $p = \frac{1}{2}$. The graphon associated with a finite graph in [10, Eq. (4)] agrees almost everywhere with the empirical graphon from Definition 3.2. By Remark 3.9, the reduced graphon space used here is the same reduced cut-metric space as the one used in [10]. Finally, for $p = \frac{1}{2}$, the rate function in [10, Eq. (7)] is exactly (5.9). The closed-set upper bound in [10, Thm. 2.3], together with (5.8), therefore gives (5.11). $\square$

Lemma 5.8. *Let $\tilde{\mathcal{F}} \subseteq \tilde{\mathcal{W}}$ be closed. For every $n \in \mathbb{N}$,*

$$(5.12) \quad \mathbb{P}_n^{\text{unif}} \left(\left\{ G \in \mathbf{G}_n : \tilde{W}_G \in \tilde{\mathcal{F}} \right\} \right) = 2^{-\binom{n}{2}} \left| \left\{ G \in \mathbf{G}_n : \tilde{W}_G \in \tilde{\mathcal{F}} \right\} \right|.$$

Consequently,

$$(5.13) \quad \limsup_{n \rightarrow \infty} \frac{1}{n^2} \ln \left| \left\{ G \in \mathbf{G}_n : \tilde{W}_G \in \tilde{\mathcal{F}} \right\} \right| \leq \sup_{\tilde{W} \in \tilde{\mathcal{F}}} \text{Ent}(\tilde{W}).$$

If $\tilde{\mathcal{F}} \neq \emptyset$, then for every $\gamma > 0$ there is $N = N(\tilde{\mathcal{F}}, \gamma)$ such that, for every $n \geq N$,

$$(5.14) \quad \left| \left\{ G \in \mathbf{G}_n : \widetilde{W}_G \in \tilde{\mathcal{F}} \right\} \right| \leq \exp \left(\left[\sup_{\widetilde{W} \in \tilde{\mathcal{F}}} \text{Ent}(\widetilde{W}) + \gamma \right] n^2 \right).$$

Proof. Equation (5.12) follows immediately from the definition (5.7).

Set

$$N_n(\tilde{\mathcal{F}}) := \left| \left\{ G \in \mathbf{G}_n : \widetilde{W}_G \in \tilde{\mathcal{F}} \right\} \right|.$$

Using (5.12), with $\ln 0 = -\infty$ when $N_n(\tilde{\mathcal{F}}) = 0$, we obtain

$$\frac{1}{n^2} \ln N_n(\tilde{\mathcal{F}}) = \frac{\binom{n}{2}}{n^2} \ln 2 + \frac{1}{n^2} \ln \mathbb{P}_n^{\text{unif}} \left(\left\{ G \in \mathbf{G}_n : \widetilde{W}_G \in \tilde{\mathcal{F}} \right\} \right).$$

Taking the upper limit and applying (5.11) gives

$$\limsup_{n \rightarrow \infty} \frac{1}{n^2} \ln N_n(\tilde{\mathcal{F}}) \leq \frac{\ln 2}{2} - \inf_{\widetilde{W} \in \tilde{\mathcal{F}}} I_{\frac{1}{2}}(\widetilde{W}).$$

By (5.10), the right-hand side equals

$$\sup_{\widetilde{W} \in \tilde{\mathcal{F}}} \text{Ent}(\widetilde{W}).$$

This proves (5.13).

Finally, suppose that $\tilde{\mathcal{F}} \neq \emptyset$ and fix $\gamma > 0$. By (5.13), for all sufficiently large n ,

$$\frac{1}{n^2} \ln N_n(\tilde{\mathcal{F}}) \leq \sup_{\widetilde{W} \in \tilde{\mathcal{F}}} \text{Ent}(\widetilde{W}) + \gamma.$$

Exponentiating proves (5.14). $\square$

Lemma 5.9. *Let $\mathcal{F} \subseteq \mathcal{W}$ be a nonempty cut-closed family. For every $\gamma > 0$, there exists $N = N(\mathcal{F}, \gamma)$ such that, for every $n \geq N$,*

$$(5.15) \quad \left| \left\{ G \in \mathbf{G}_n : W_G \in \mathcal{F} \right\} \right| \leq \exp \left(\left[\sup_{W \in \mathcal{F}} \text{Ent}(W) + \gamma \right] n^2 \right).$$

Equivalently, there is a nonnegative sequence $\alpha_n(\mathcal{F})$ satisfying

$$\lim_{n \rightarrow \infty} \alpha_n(\mathcal{F}) = 0$$

such that

$$(5.16) \quad \left| \left\{ G \in \mathbf{G}_n : W_G \in \mathcal{F} \right\} \right| \leq \exp \left(\left[\sup_{W \in \mathcal{F}} \text{Ent}(W) + \alpha_n(\mathcal{F}) \right] n^2 \right)$$

for every sufficiently large n .

Proof. Let $\tilde{\mathcal{F}}$ be the image of $\mathcal{F}$ in $\widetilde{\mathcal{W}}$. Lemma 5.5 shows that $\tilde{\mathcal{F}}$ is closed and that

$$W_G \in \mathcal{F} \iff \widetilde{W}_G \in \tilde{\mathcal{F}}.$$

Therefore

$$\left| \left\{ G \in \mathbf{G}_n : W_G \in \mathcal{F} \right\} \right| = \left| \left\{ G \in \mathbf{G}_n : \widetilde{W}_G \in \tilde{\mathcal{F}} \right\} \right|.$$

Entropy is constant on every zero-cut-distance class by Lemma 3.17, so

$$\sup_{\widetilde{W} \in \tilde{\mathcal{F}}} \text{Ent}(\widetilde{W}) = \sup_{W \in \mathcal{F}} \text{Ent}(W).$$

Now apply (5.14) to $\tilde{\mathcal{F}}$. This proves (5.15).

For the second form, set

$$S_{\mathcal{F}} := \sup_{W \in \mathcal{F}} \text{Ent}(W)$$

and

$$N_n(\mathcal{F}) := |\{G \in \mathbf{G}_n : W_G \in \mathcal{F}\}|.$$

If $N_n(\mathcal{F}) = 0$, set $b_n = 0$. If $N_n(\mathcal{F}) > 0$, set

$$b_n := \max \left\{ 0, \frac{1}{n^2} \ln N_n(\mathcal{F}) - S_{\mathcal{F}} \right\}.$$

We now show that $\lim_{n \rightarrow \infty} b_n = 0$. Fix $\gamma > 0$. By (5.15), for all sufficiently large n ,

$$N_n(\mathcal{F}) \leq \exp((S_{\mathcal{F}} + \gamma)n^2).$$

If $N_n(\mathcal{F}) = 0$, then $b_n = 0$. If $N_n(\mathcal{F}) > 0$, the last inequality gives

$$\frac{1}{n^2} \ln N_n(\mathcal{F}) - S_{\mathcal{F}} \leq \gamma,$$

and therefore

$$0 \leq b_n \leq \gamma.$$

Since $\gamma > 0$ was arbitrary, it follows that

$$\lim_{n \rightarrow \infty} b_n = 0.$$

Set

$$\alpha_n(\mathcal{F}) := b_n + \frac{1}{n}.$$

Then

$$\lim_{n \rightarrow \infty} \alpha_n(\mathcal{F}) = 0.$$

If $N_n(\mathcal{F}) > 0$, the definition of b_n gives

$$\frac{1}{n^2} \ln N_n(\mathcal{F}) \leq S_{\mathcal{F}} + b_n \leq S_{\mathcal{F}} + \alpha_n(\mathcal{F}).$$

Exponentiating gives (5.16). If $N_n(\mathcal{F}) = 0$, then (5.16) is trivial. Thus (5.16) holds for every sufficiently large n . $\square$

Lemma 5.10. *Fix $r \geq 3$ and set $q = r - 1$. The following subsets of $\mathcal{W}$ are cut-closed.*

(1) *For every $\delta \in [0, 1]$,*

$$\mathcal{F}_{\delta}^{(r)} := \{W \in \mathcal{W} : t(K_r, W) \leq \delta\}.$$

(2) *For every $\varepsilon > 0$,*

$$\mathcal{D}_{\varepsilon, q} := \{W \in \mathcal{W} : \delta_{\square}(W, B_q^{\star}) \geq \varepsilon\}.$$

(3) *For every $\delta \in [0, 1]$ and $\varepsilon > 0$,*

$$\mathcal{F}_{\delta, \varepsilon}^{(r, q)} := \mathcal{F}_{\delta}^{(r)} \cap \mathcal{D}_{\varepsilon, q}.$$

Proof. (1) Fix $\delta \in [0, 1]$. We show that

$$\mathcal{F}_{\delta}^{(r)} = \{W \in \mathcal{W} : t(K_r, W) \leq \delta\}$$

is cut-closed.

Let $(W_m)_{m \geq 1}$ be a sequence of graphons in $\mathcal{F}_{\delta}^{(r)}$, and suppose that

$$\lim_{m \rightarrow \infty} \delta_{\square}(W_m, W) = 0$$

for some graphon $W \in \mathcal{W}$.

Since each W_m belongs to $\mathcal{F}_{\delta}^{(r)}$, we have

$$t(K_r, W_m) \leq \delta \quad \text{for every } m.$$

By Lemma 3.14, the map

$$W \mapsto t(K_r, W)$$

is continuous in cut distance. Hence

$$\lim_{m \rightarrow \infty} t(K_r, W_m) = t(K_r, W).$$

Taking the limit in

$$t(K_r, W_m) \leq \delta$$

gives

$$t(K_r, W) \leq \delta.$$

Therefore

$$W \in \mathcal{F}_\delta^{(r)}.$$

So $\mathcal{F}_\delta^{(r)}$ is cut-closed.

(2) Fix $\varepsilon > 0$. We prove that

$$\mathcal{D}_{\varepsilon, q} = \{W \in \mathcal{W} : \delta_\square(W, B_q^*) \geq \varepsilon\}$$

is cut-closed.

If $\mathcal{H}$ is a nonempty family of graphons, then the function

$$W \mapsto \delta_\square(W, \mathcal{H})$$

is continuous in cut distance. Indeed, let $U, V \in \mathcal{W}$. For every $H \in \mathcal{H}$, the triangle inequality gives

$$\delta_\square(U, H) \leq \delta_\square(U, V) + \delta_\square(V, H).$$

Taking the infimum over all $H \in \mathcal{H}$ gives

$$\delta_\square(U, \mathcal{H}) \leq \delta_\square(U, V) + \delta_\square(V, \mathcal{H}).$$

So

$$\delta_\square(U, \mathcal{H}) - \delta_\square(V, \mathcal{H}) \leq \delta_\square(U, V).$$

Now if we switch the roles of U and V , we get

$$\delta_\square(V, \mathcal{H}) - \delta_\square(U, \mathcal{H}) \leq \delta_\square(U, V).$$

Therefore

$$|\delta_\square(U, \mathcal{H}) - \delta_\square(V, \mathcal{H})| \leq \delta_\square(U, V).$$

Thus $W \mapsto \delta_\square(W, \mathcal{H})$ is continuous.

Now apply this with

$$\mathcal{H} = B_q^*.$$

This family is nonempty, because $W_q^* \in B_q^*$ by Notation 4.3. Hence

$$W \mapsto \delta_\square(W, B_q^*)$$

is continuous in cut distance.

Now let $(W_m)_{m \geq 1}$ be a sequence of graphons in $\mathcal{D}_{\varepsilon, q}$, and suppose that

$$\lim_{m \rightarrow \infty} \delta_\square(W_m, W) = 0$$

for some graphon $W \in \mathcal{W}$.

Since each W_m belongs to $\mathcal{D}_{\varepsilon, q}$, we have

$$\delta_\square(W_m, B_q^*) \geq \varepsilon \quad \text{for every } m.$$

By the continuity established above, we have

$$\lim_{m \rightarrow \infty} \delta_\square(W_m, B_q^*) = \delta_\square(W, B_q^*).$$

Taking the limit in

$$\delta_\square(W_m, B_q^*) \geq \varepsilon$$

gives

$$\delta_\square(W, B_q^*) \geq \varepsilon.$$

Therefore

$$W \in \mathcal{D}_{\varepsilon, q}.$$

So $\mathcal{D}_{\varepsilon, q}$ is cut-closed.

(3) Fix $\delta \in [0, 1]$ and $\varepsilon > 0$. We show that

$$\mathcal{F}_{\delta, \varepsilon}^{(r, q)} = \mathcal{F}_{\delta}^{(r)} \cap \mathcal{D}_{\varepsilon, q}$$

is cut-closed.

Let $(W_m)_{m \geq 1}$ be a sequence of graphons in $\mathcal{F}_{\delta, \varepsilon}^{(r, q)}$, and suppose that

$$\lim_{m \rightarrow \infty} \delta_{\square}(W_m, W) = 0$$

for some graphon $W \in \mathcal{W}$.

Since

$$W_m \in \mathcal{F}_{\delta, \varepsilon}^{(r, q)} = \mathcal{F}_{\delta}^{(r)} \cap \mathcal{D}_{\varepsilon, q},$$

we have

$$W_m \in \mathcal{F}_{\delta}^{(r)} \quad \text{and} \quad W_m \in \mathcal{D}_{\varepsilon, q}$$

for every m .

By item (1), $\mathcal{F}_{\delta}^{(r)}$ is cut-closed. Hence

$$W \in \mathcal{F}_{\delta}^{(r)}.$$

By item (2), $\mathcal{D}_{\varepsilon, q}$ is cut-closed. Hence

$$W \in \mathcal{D}_{\varepsilon, q}.$$

Therefore

$$W \in \mathcal{F}_{\delta}^{(r)} \cap \mathcal{D}_{\varepsilon, q} = \mathcal{F}_{\delta, \varepsilon}^{(r, q)}.$$

So $\mathcal{F}_{\delta, \varepsilon}^{(r, q)}$ is cut-closed. □

5.4. The reduced free-energy limit.

Lemma 5.11 (Balanced multipartite lower bound). *Let $n \in \mathbb{N}$, and use*

$$n = qm + a, \quad m, a \in \mathbb{Z}_{\geq 0}, \quad 0 \leq a < q.$$

Then, with s_q as in Notation 5.1,

$$(5.17) \quad \tilde{Z}_{n, r} \geq \exp \left(s_q n^2 - \frac{\ln 2}{2} a \left(1 - \frac{a}{q} \right) \right) \geq \exp \left(s_q n^2 - \frac{q \ln 2}{8} \right).$$

Proof. Write $[n]$ as the disjoint union of q possibly empty sets

$$[n] = V_1 \sqcup \cdots \sqcup V_q$$

whose sizes differ by at most one. Thus a sets have size $m + 1$, and the remaining $q - a$ sets have size m . Put

$$n_i := |V_i|, \quad i = 1, \dots, q.$$

Let M be the number of unordered pairs of vertices lying in different parts. Then

$$M = \sum_{1 \leq i < j \leq q} n_i n_j.$$

Since

$$\left(\sum_{i=1}^q n_i \right)^2 = \sum_{i=1}^q n_i^2 + 2 \sum_{1 \leq i < j \leq q} n_i n_j,$$

we have

$$(5.18) \quad M = \frac{1}{2} \left(n^2 - \sum_{i=1}^q n_i^2 \right).$$

We now compute the sum of the squared part sizes. By the choice of the partition,

$$\begin{aligned} \sum_{i=1}^q n_i^2 &= (q-a)m^2 + a(m+1)^2 \\ &= qm^2 + 2am + a. \end{aligned}$$

On the other hand, since $n = qm + a$,

$$\frac{n^2}{q} = qm^2 + 2am + \frac{a^2}{q}.$$

Therefore

$$(5.19) \quad \sum_{i=1}^q n_i^2 = \frac{n^2}{q} + a \left(1 - \frac{a}{q} \right).$$

Substituting (5.19) into (5.18) gives

$$(5.20) \quad M = \frac{1}{2} \left(1 - \frac{1}{q} \right) n^2 - \frac{1}{2} a \left(1 - \frac{a}{q} \right).$$

Now choose an arbitrary subset of these M possible between-part edges, and put no edges inside the parts. This gives exactly 2^M labeled graphs on $[n]$.

Every graph constructed in this way is q -partite. Hence every clique in such a graph uses at most one vertex from each part, so every clique has size at most q . Since $q = r-1$, none of these graphs contains a copy of K_r . Therefore each of these graphs has reduced mass

$$\exp(-r! w K_r(G)) = 1.$$

Thus all these graphs contribute 1 each to $\tilde{Z}_{n,r}$, and so

$$\tilde{Z}_{n,r} \geq 2^M.$$

Multiplying (5.20) by $\ln 2$ and then using (5.1), we obtain

$$\begin{aligned} (\ln 2)M &= \frac{\ln 2}{2} \left(1 - \frac{1}{q} \right) n^2 - \frac{\ln 2}{2} a \left(1 - \frac{a}{q} \right) \\ &= s_q n^2 - \frac{\ln 2}{2} a \left(1 - \frac{a}{q} \right). \end{aligned}$$

Consequently,

$$2^M = \exp \left(s_q n^2 - \frac{\ln 2}{2} a \left(1 - \frac{a}{q} \right) \right).$$

Together with $\tilde{Z}_{n,r} \geq 2^M$, this proves the first inequality in (5.17).

Finally, the function

$$x \left(1 - \frac{x}{q} \right)$$

is maximized on $[0, q]$ at $x = q/2$, where its value is $q/4$. Since $0 \leq a < q$, we have

$$0 \leq a \left(1 - \frac{a}{q} \right) \leq \frac{q}{4}.$$

Hence

$$-\frac{\ln 2}{2} a \left(1 - \frac{a}{q} \right) \geq -\frac{q \ln 2}{8}.$$

This gives the second inequality in (5.17). $\square$

We now prove Theorem 2.7(i). We first fix an auxiliary clique coefficient L , then let $n \rightarrow \infty$, and only afterward let $L \rightarrow \infty$.

Proof of Theorem 2.7(i). We prove the upper and lower bounds separately.

Upper bound. Fix $L > 0$ and define

$$\tilde{Z}_{n,r}^{(L)} := \sum_{G \in \mathbf{G}_n} \exp(-Ln^2 t(K_r, W_G)).$$

This is an auxiliary partition function in which the clique penalty L remains fixed as n grows.

For every graphon W ,

$$0 \leq t(K_r, W) \leq 1.$$

Moreover, $t(K_r, W)$ is continuous in cut distance by Lemma 3.14. Hence the functional

$$W \mapsto -L t(K_r, W)$$

is bounded and continuous. We may therefore apply [9, Thm. 3.1], which gives

$$(5.21) \quad \lim_{n \rightarrow \infty} \frac{1}{n^2} \ln \tilde{Z}_{n,r}^{(L)} = \sup_{W \in \mathcal{W}} \{\text{Ent}(W) - L t(K_r, W)\}.$$

We next determine what happens to the right-hand side when L becomes large. In the notation of [9, Thm. 7.1], take

$$H = K_r, \quad \beta_1 = 0, \quad \beta_2 = -L.$$

The rate functional I used there satisfies

$$I(W) = \frac{1}{2} \int_{[0,1]^2} [W \ln W + (1 - W) \ln(1 - W)] dx dy = -\text{Ent}(W),$$

with the usual convention at 0 and 1. Thus, at $\beta_2 = -L$, their variational value is exactly

$$\sup_{W \in \mathcal{W}} \{\text{Ent}(W) - L t(K_r, W)\}.$$

Moreover,

$$\chi(K_r) = r \quad \text{and} \quad p = \frac{e^{2\beta_1}}{1 + e^{2\beta_1}} = \frac{1}{2}.$$

Theorem 7.1 of that paper therefore gives the limiting value

$$\frac{r-2}{2(r-1)} \ln 2 = \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right) = s_q,$$

where we used $q = r - 1$. It follows that

$$(5.22) \quad \lim_{L \rightarrow \infty} \sup_{W \in \mathcal{W}} \{\text{Ent}(W) - L t(K_r, W)\} = s_q.$$

We now compare the auxiliary model with the model in the theorem. Write

$$L_n := wn^{r-2}.$$

By (5.2), the reduced mass of a graph G is

$$\exp(-r! w K_r(G)) = \exp(-n^2 L_n t(K_r, W_G)).$$

Since $r \geq 3$ and $w > 0$, we have

$$\lim_{n \rightarrow \infty} L_n = +\infty.$$

Thus, for every fixed $L > 0$, we have $L_n \geq L$ for all sufficiently large n . Since $t(K_r, W_G) \geq 0$, this gives

$$\exp(-n^2 L_n t(K_r, W_G)) \leq \exp(-n^2 L t(K_r, W_G)).$$

Summing this inequality over all $G \in \mathbf{G}_n$ yields

$$\tilde{Z}_{n,r} \leq \tilde{Z}_{n,r}^{(L)}$$

for all sufficiently large n . Taking logarithms, dividing by n^2 , and using (5.21), we obtain

$$\begin{aligned} \limsup_{n \rightarrow \infty} F_{n,r}(w) &\leq \lim_{n \rightarrow \infty} \frac{1}{n^2} \ln \tilde{Z}_{n,r}^{(L)} \\ &= \sup_{W \in \mathcal{W}} \{ \text{Ent}(W) - L t(K_r, W) \}. \end{aligned}$$

This holds for every fixed $L > 0$. We may therefore let $L \rightarrow \infty$. Equation (5.22) then gives

$$\limsup_{n \rightarrow \infty} F_{n,r}(w) \leq s_q.$$

Lower bound. A graph whose edges lie only between the parts of a q -partite partition contains no copy of K_r , because $q = r - 1$. Such a graph therefore has reduced mass 1. Lemma 5.11 counts these graphs and gives

$$F_{n,r}(w) \geq s_q - \frac{q \ln 2}{8n^2}.$$

Taking the lower limit gives

$$\liminf_{n \rightarrow \infty} F_{n,r}(w) \geq s_q.$$

The upper and lower bounds now agree. Therefore

$$\lim_{n \rightarrow \infty} F_{n,r}(w) = s_q = \frac{\ln 2}{2} \left(1 - \frac{1}{q} \right).$$

□

5.5. Exponential concentration in cut distance. We now prove Theorem 2.7(ii). Fix $\varepsilon > 0$. The event we need to estimate is

$$\delta_{\square}(W_{\mathbf{G}_n}, B_q^*) \geq \varepsilon.$$

After choosing a small number $\delta > 0$, we separate this event according to the empirical clique density $t(K_r, W_{\mathbf{G}_n})$. The part with

$$t(K_r, W_{\mathbf{G}_n}) \geq \delta$$

has very small total reduced mass by Lemma 5.3. The remaining part has

$$t(K_r, W_{\mathbf{G}_n}) < \delta \quad \text{and} \quad \delta_{\square}(W_{\mathbf{G}_n}, B_q^*) \geq \varepsilon.$$

For this part, Lemma 4.12 gives an entropy loss, and the counting bound in Lemma 5.9 turns that entropy loss into an exponential probability bound.

As explained in Remark 4.7, [9, Thm. 3.2] gives concentration for each fixed clique coefficient L , but it does not provide concentration constants uniform in L . Since the present model follows the diagonal sequence $L_n = wn^{r-2}$, for which $\lim_{n \rightarrow \infty} L_n = +\infty$, the proof below establishes the required uniform diagonal estimate directly.

Proof of Theorem 2.7(ii). Fix $\varepsilon > 0$, and let $\mathbf{G}_n \sim \mathbb{P}_{n,r}$. Choose

$$\delta_0 = \delta_{r,\varepsilon} > 0 \quad \text{and} \quad c' = c'_{r,\varepsilon} > 0$$

from Lemma 4.12, and set

$$\delta := \min\{\delta_0, 1\}.$$

The conclusion of Lemma 4.12 still holds under the stronger condition $t(K_r, W) \leq \delta$.

Define

$$\mathcal{A}_{n,\varepsilon} := \{G \in \mathbf{G}_n : \delta_{\square}(W_G, B_q^*) \geq \varepsilon\}.$$

By (2.3),

$$\mathbb{P}(\mathbf{G}_n \in \mathcal{A}_{n,\varepsilon}) = \frac{R_{n,\varepsilon}}{\tilde{Z}_{n,r}},$$

where

$$R_{n,\varepsilon} := \sum_{G \in \mathcal{A}_{n,\varepsilon}} \exp(-r! w K_r(G)).$$

Using (5.2), set

$$R_{n,\varepsilon} = R_{n,\varepsilon}^{\text{large}} + R_{n,\varepsilon}^{\text{small}},$$

where

$$R_{n,\varepsilon}^{\text{large}} := \sum_{\substack{G \in \mathcal{A}_{n,\varepsilon} \\ t(K_r, W_G) \geq \delta}} \exp(-w t(K_r, W_G) n^r)$$

and

$$R_{n,\varepsilon}^{\text{small}} := \sum_{\substack{G \in \mathcal{A}_{n,\varepsilon} \\ t(K_r, W_G) < \delta}} \exp(-w t(K_r, W_G) n^r).$$

For every $M > 0$, Lemma 5.3 gives, for all sufficiently large n ,

$$(5.23) \quad R_{n,\varepsilon}^{\text{large}} \leq \exp(-Mn^2).$$

Every summand in $R_{n,\varepsilon}^{\text{small}}$ is at most 1, and every graph in this sum satisfies

$$W_G \in \mathcal{F}_{\delta,\varepsilon}^{(r,q)}.$$

Hence

$$R_{n,\varepsilon}^{\text{small}} \leq \left| \left\{ G \in \mathbf{G}_n : W_G \in \mathcal{F}_{\delta,\varepsilon}^{(r,q)} \right\} \right|.$$

If $\mathcal{F}_{\delta,\varepsilon}^{(r,q)}$ is empty, this upper bound is 0. If it is nonempty, then it is cut-closed by Lemma 5.10, and Lemma 4.12 gives

$$\sup_{W \in \mathcal{F}_{\delta,\varepsilon}^{(r,q)}} \text{Ent}(W) \leq s_q - c'.$$

Lemma 5.9, with counting error $c'/4$, now gives, for every sufficiently large n ,

$$(5.24) \quad R_{n,\varepsilon}^{\text{small}} \leq \exp\left(\left(s_q - \frac{3c'}{4}\right)n^2\right).$$

Set

$$C_q := \frac{q \ln 2}{8}.$$

Lemma 5.11 gives

$$\tilde{Z}_{n,r} \geq \exp(s_q n^2 - C_q).$$

Choose $M = 1 + c'$. Combining this estimate with (5.23) and (5.24) gives

$$\mathbb{P}(\mathbf{G}_n \in \mathcal{A}_{n,\varepsilon}) \leq \exp(-(M + s_q)n^2 + C_q) + \exp\left(-\frac{3c'}{4}n^2 + C_q\right).$$

For every sufficiently large n , both terms on the right are at most $\exp(-c'n^2/2)$. Increasing the threshold for n once more gives

$$2 \exp(-c'n^2/2) \leq \exp(-c'n^2/4).$$

Therefore, for every sufficiently large n ,

$$\mathbb{P}(\delta_{\square}(W_{\mathbf{G}_n}, B_q^*) \geq \varepsilon) \leq \exp(-c'n^2/4).$$

The theorem follows with $c = c'/4$. The constant c depends only on ε and r . The threshold for n may also depend on w , through Lemma 5.3. $\square$

5.6. Probability bounds needed for the rounding argument.

Lemma 5.12. *The following probability bounds hold:*

- (i) **Markov's inequality.** *If $Y \geq 0$ almost surely and $t > 0$, then*

$$\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t}.$$

- (ii) **Hoeffding's inequality.** *Let $m \geq 1$, and let $Y_1, \dots, Y_m$ be independent random variables. Suppose that, for each $i \in [m]$, there are real numbers $a_i \leq b_i$ such that*

$$a_i \leq Y_i \leq b_i$$

almost surely. Set

$$V := \sum_{i=1}^m (b_i - a_i)^2.$$

If $V > 0$, then, for every $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^m Y_i - \sum_{i=1}^m \mathbb{E}[Y_i]\right| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{V}\right).$$

If $V = 0$, then

$$\sum_{i=1}^m Y_i = \sum_{i=1}^m \mathbb{E}[Y_i]$$

almost surely. Consequently, for every $t > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^m Y_i - \sum_{i=1}^m \mathbb{E}[Y_i]\right| \geq t\right) = 0.$$

In particular, if $0 \leq Y_i \leq 1$ almost surely for every $i \in [m]$, and if

$$S := \sum_{i=1}^m Y_i,$$

then, for every $t \geq 0$,

$$\mathbb{P}(|S - \mathbb{E}[S]| \geq t) \leq 2 \exp\left(-\frac{2t^2}{m}\right).$$

- (iii) **McDiarmid's inequality.** *Let $m \geq 1$, let $S_1, \dots, S_m$ be finite sets, and let $X_1, \dots, X_m$ be independent random variables such that*

$$X_i \in S_i$$

almost surely for every $i \in [m]$. Let

$$f : S_1 \times \dots \times S_m \rightarrow \mathbb{R}$$

be a function. Assume that there is a number $L > 0$ such that changing only one coordinate can change the value of f by at most L . More precisely,

$$|f(x_1, \dots, x_i, \dots, x_m) - f(x_1, \dots, x'_i, \dots, x_m)| \leq L$$

for every $i \in [m]$, every $x_i, x'_i \in S_i$, and every fixed choice of the other coordinates. Then, for every $t \geq 0$,

$$\mathbb{P}(|f(X_1, \dots, X_m) - \mathbb{E}[f(X_1, \dots, X_m)]| \geq t) \leq 2 \exp\left(-\frac{2t^2}{mL^2}\right).$$

Proof. Item (i) is Markov's inequality; see [23, Thm. 3.1], with the notation there renamed as $X = Y$ and $a = t$.

For item (ii), first assume that $V > 0$.

If $t > 0$, Hoeffding's inequality [23, Thm. 4.14] gives

$$\mathbb{P} \left(\left| \sum_{i=1}^m Y_i - \sum_{i=1}^m \mathbb{E}[Y_i] \right| \geq t \right) \leq 2 \exp \left(-\frac{2t^2}{V} \right).$$

If $t = 0$, the same bound is also true, because the probability on the left is at most 1, while the right-hand side is 2.

Now assume that

$$V = 0.$$

Then

$$\sum_{i=1}^m (b_i - a_i)^2 = 0.$$

Each term in this sum is nonnegative, so

$$b_i - a_i = 0 \quad \text{for every } i \in [m].$$

Thus $a_i = b_i$ for every i . Since

$$a_i \leq Y_i \leq b_i$$

almost surely, each Y_i is almost surely constant. Hence

$$Y_i = \mathbb{E}[Y_i] \quad \text{almost surely for every } i \in [m].$$

Adding over $i = 1, \dots, m$ gives

$$\sum_{i=1}^m Y_i = \sum_{i=1}^m \mathbb{E}[Y_i]$$

almost surely. Therefore, for every $t > 0$,

$$\mathbb{P} \left(\left| \sum_{i=1}^m Y_i - \sum_{i=1}^m \mathbb{E}[Y_i] \right| \geq t \right) = 0.$$

For the last part of item (ii), assume that

$$0 \leq Y_i \leq 1 \quad \text{almost surely for every } i \in [m],$$

and set

$$S := \sum_{i=1}^m Y_i.$$

Taking

$$a_i = 0 \quad \text{and} \quad b_i = 1$$

for every i gives

$$V = \sum_{i=1}^m (b_i - a_i)^2 = \sum_{i=1}^m 1 = m.$$

The bound already proved therefore gives, for every $t \geq 0$,

$$\mathbb{P} (|S - \mathbb{E}[S]| \geq t) \leq 2 \exp \left(-\frac{2t^2}{m} \right).$$

For item (iii), the bounded-difference condition is exactly the hypothesis of McDiarmid's inequality in [23, Sec. 13.5.1, Thm. 13.7]. Applying that theorem with $n = m$, $c = L$, and $\lambda = t$ gives the stated bound for $t > 0$. The theorem applies here because the random variables take values in finite sets. For $t = 0$, the bound is immediate, since the right-hand side is 2. $\square$

5.7. Transforming cut-distance closeness to a vertex partition. The next lemma turns cut-distance closeness to B_q^* into a partition of the finite vertex set. It controls the part sizes, the edge density between every pair of parts, and the total number of edges inside the parts.

Lemma 5.13. *Let $q \geq 2$ and $0 < \varepsilon \leq 1$, and set*

$$\rho := \frac{\varepsilon}{100q^2}.$$

There exists $N_{\text{det}} = N_{\text{det}}(\varepsilon, q)$ such that the following holds for every $n \geq N_{\text{det}}$. If G is a graph on $[n]$ and

$$\delta_{\square}(W_G, B_q^*) < \rho,$$

then there is a partition

$$[n] = V_1 \sqcup \cdots \sqcup V_q$$

with

$$||V_{\alpha}| - n/q| \leq \varepsilon n \quad \text{for every } \alpha \in [q],$$

each V_{α} has size at least 2,

$$\left| \frac{e_G(V_{\alpha}, V_{\beta})}{|V_{\alpha}||V_{\beta}|} - \frac{1}{2} \right| \leq \varepsilon \quad \text{for all distinct } \alpha, \beta \in [q],$$

$$\sum_{\alpha=1}^q e_G(V_{\alpha}) \leq \frac{\varepsilon n^2}{10q},$$

and

$$\frac{\sum_{\alpha=1}^q e_G(V_{\alpha})}{\sum_{\alpha=1}^q \binom{|V_{\alpha}|}{2}} \leq \varepsilon.$$

Moreover, if G' is obtained from G by deleting all edges whose endpoints lie in the same part, then G' is q -partite and

$$d_{\text{edit}}(G, G') \leq \frac{\varepsilon}{10q}.$$

Proof. We use a random coloring argument and color the vertices independently using probabilities obtained from the graphon relabeling. We then show that the desired properties hold with positive probability and fix one coloring for which they hold.

Let $(a_{ij})_{i,j=1}^n$ be the adjacency matrix of G . Thus

$$a_{ij} = a_{ji}, \quad a_{ii} = 0.$$

Let W_q^* and its canonical blocks $A_1, \dots, A_q$ be as in Notation 4.3. By that notation,

$$\delta_{\square}(W_G, B_q^*) = \delta_{\square}(W_G, W_q^*) < \rho.$$

Since the infimum in the definition of cut distance is strictly smaller than ρ , there is a measure-preserving bijection $\psi : [0, 1] \rightarrow [0, 1]$ such that

$$\|W_G - (W_q^*)^{\psi}\|_{\square} < \rho.$$

Set

$$\phi := \psi^{-1}.$$

Since the cut norm is invariant under measure-preserving relabelings, we get

$$\begin{aligned} \|W_G^{\phi} - W_q^*\|_{\square} &= \|(W_G - (W_q^*)^{\psi})^{\phi}\|_{\square} \\ (5.25) \quad &= \|W_G - (W_q^*)^{\psi}\|_{\square} \\ &< \rho. \end{aligned}$$

Let $I_1, \dots, I_n$ be the intervals from Definition 3.2. These intervals form a measurable partition of $[0, 1]$.

Define

$$J_i := \phi^{-1}(I_i), \quad i \in [n].$$

Then $J_1, \dots, J_n$ form a measurable partition of $[0, 1]$, and

$$\lambda(J_i) = \frac{1}{n} \quad \text{for every } i \in [n].$$

For $i \in [n]$ and $\alpha \in [q]$, set

$$p_{i,\alpha} := n\lambda(J_i \cap A_\alpha).$$

For each fixed i , these numbers can be used as probabilities:

$$0 \leq p_{i,\alpha} \leq 1 \quad \text{and} \quad \sum_{\alpha=1}^q p_{i,\alpha} = 1.$$

Also, for each fixed α ,

$$\sum_{i=1}^n p_{i,\alpha} = n\lambda(A_\alpha) = \frac{n}{q}.$$

We now define an auxiliary random coloring of the vertices. Let

$$\Sigma := (\xi_1, \dots, \xi_n)$$

be an $[q]^n$ -valued random vector whose coordinates are independent and satisfy

$$\mathbb{P}_{\text{col}}(\xi_i = \alpha) = p_{i,\alpha}, \quad i \in [n], \alpha \in [q].$$

Equivalently, for every $C = (C_1, \dots, C_n) \in [q]^n$,

$$\mathbb{P}_{\text{col}}(\Sigma = C) = \prod_{i=1}^n p_{i,C_i}.$$

Here $\mathbb{P}_{\text{col}}$ and $\mathbb{E}_{\text{col}}$ denote probability and expectation with respect to this product probability measure on $[q]^n$.

For a coloring $C = (C_1, \dots, C_n) \in [q]^n$, define its color classes by

$$V_\alpha(C) := \{i \in [n] : C_i = \alpha\}, \quad \alpha \in [q].$$

The parts are almost balanced. For every $\alpha \in [q]$ and $i \in [n]$, set

$$Z_i^{(\alpha)} := \mathbf{1}_{\{\xi_i = \alpha\}}.$$

For each fixed α , the random variables $Z_1^{(\alpha)}, \dots, Z_n^{(\alpha)}$ are independent and satisfy

$$0 \leq Z_i^{(\alpha)} \leq 1 \quad \text{almost surely.}$$

Moreover,

$$|V_\alpha(\Sigma)| = \sum_{i=1}^n Z_i^{(\alpha)}$$

and

$$\mathbb{E}_{\text{col}}[|V_\alpha(\Sigma)|] = \sum_{i=1}^n \mathbb{E}_{\text{col}}[Z_i^{(\alpha)}] = \sum_{i=1}^n p_{i,\alpha} = \frac{n}{q}.$$

Applying Lemma 5.12(ii) to $Z_1^{(\alpha)}, \dots, Z_n^{(\alpha)}$, with $m = n$ and

$$t = \frac{\varepsilon n}{10q},$$

gives

$$\mathbb{P}_{\text{col}} \left(\left| |V_\alpha(\Sigma)| - \frac{n}{q} \right| \geq \frac{\varepsilon n}{10q} \right) \leq 2 \exp \left(-\frac{\varepsilon^2}{50q^2} n \right).$$

Define

$$\Omega_{\text{bal}} := \left\{ C \in [q]^n : \left| |V_\alpha(C)| - \frac{n}{q} \right| \leq \frac{\varepsilon n}{10q} \text{ for every } \alpha \in [q] \right\}.$$

By the union bound,

$$(5.26) \quad \mathbb{P}_{\text{col}}(\Omega_{\text{bal}}^c) \leq 2q \exp\left(-\frac{\varepsilon^2}{50q^2}n\right).$$

If $C \in \Omega_{\text{bal}}$, then for every $\alpha \in [q]$,

$$\left| |V_\alpha(C)| - \frac{n}{q} \right| \leq \frac{\varepsilon n}{10q} \leq \varepsilon n.$$

Moreover, since $\varepsilon \leq 1$,

$$|V_\alpha(C)| \geq \frac{n}{q} - \frac{\varepsilon n}{10q} \geq \frac{9n}{10q}.$$

Thus, if $n \geq 2q$, then $|V_\alpha(C)| \geq 9/5$. Since $|V_\alpha(C)|$ is an integer, every part has at least two vertices.

After relabeling by ϕ , the empirical graphon W_G^ϕ is equal to a_{ij} on $J_i \times J_j$, except possibly on a null set. Hence, for all $\alpha, \beta \in [q]$,

$$(5.27) \quad \begin{aligned} \int_{A_\alpha \times A_\beta} W_G^\phi(x, y) dx dy &= \sum_{i,j=1}^n a_{ij} \lambda(J_i \cap A_\alpha) \lambda(J_j \cap A_\beta) \\ &= \frac{1}{n^2} \sum_{i,j=1}^n a_{ij} p_{i,\alpha} p_{j,\beta}. \end{aligned}$$

Edges inside the parts. Fix $\alpha \in [q]$. Then

$$2e_G(V_\alpha(\Sigma)) = \sum_{i,j=1}^n a_{ij} \mathbf{1}_{\{\xi_i=\alpha\}} \mathbf{1}_{\{\xi_j=\alpha\}}.$$

The diagonal terms vanish because $a_{ii} = 0$. For $i \neq j$, independence gives

$$\mathbb{E}_{\text{col}} \left[\mathbf{1}_{\{\xi_i=\alpha\}} \mathbf{1}_{\{\xi_j=\alpha\}} \right] = p_{i,\alpha} p_{j,\alpha}.$$

Using (5.27), we obtain

$$\frac{2}{n^2} \mathbb{E}_{\text{col}}[e_G(V_\alpha(\Sigma))] = \int_{A_\alpha \times A_\alpha} W_G^\phi.$$

The graphon W_q^* is 0 on $A_\alpha \times A_\alpha$. Therefore

$$0 \leq \int_{A_\alpha \times A_\alpha} W_G^\phi = \int_{A_\alpha \times A_\alpha} (W_G^\phi - W_q^*) \leq \|W_G^\phi - W_q^*\|_\square < \rho.$$

Hence

$$\mathbb{E}_{\text{col}}[e_G(V_\alpha(\Sigma))] \leq \frac{\rho n^2}{2}.$$

Summing over all $\alpha \in [q]$, we get

$$\mathbb{E}_{\text{col}} \left[\sum_{\alpha=1}^q e_G(V_\alpha(\Sigma)) \right] \leq \frac{q\rho n^2}{2}.$$

For $C \in [q]^n$, set

$$Y(C) := \sum_{\alpha=1}^q e_G(V_\alpha(C)).$$

Define

$$\Omega_{\text{in}} := \left\{ C \in [q]^n : Y(C) \leq \frac{\varepsilon n^2}{10q} \right\}.$$

By Lemma 5.12(i),

$$\begin{aligned} \mathbb{P}_{\text{col}}(\Omega_{\text{in}}^c) &\leq \frac{\mathbb{E}_{\text{col}}[Y(\Sigma)]}{\varepsilon n^2/(10q)} \\ &\leq \frac{q\rho n^2/2}{\varepsilon n^2/(10q)} \\ (5.28) \quad &= \frac{5q^2\rho}{\varepsilon} \\ &= \frac{1}{20}. \end{aligned}$$

Here the last equality uses $\rho = \varepsilon/(100q^2)$.

Lemma 3.12(ii) gives

$$\sum_{\alpha=1}^q \binom{|V_\alpha(C)|}{2} = \frac{1}{2} \left(\sum_{\alpha=1}^q |V_\alpha(C)|^2 - n \right) \geq \frac{1}{2} \left(\frac{n^2}{q} - n \right).$$

If $n \geq 2q$, then

$$\sum_{\alpha=1}^q \binom{|V_\alpha(C)|}{2} \geq \frac{n^2}{4q}.$$

Thus every $C \in \Omega_{\text{in}}$ satisfies

$$\frac{\sum_{\alpha=1}^q e_G(V_\alpha(C))}{\sum_{\alpha=1}^q \binom{|V_\alpha(C)|}{2}} \leq \frac{\varepsilon n^2/(10q)}{n^2/(4q)} = \frac{2\varepsilon}{5} < \varepsilon.$$

Edges between distinct parts. For $1 \leq \alpha < \beta \leq q$ and $C \in [q]^n$, define

$$F_{\alpha,\beta}(C) := e_G(V_\alpha(C), V_\beta(C)).$$

Fix $1 \leq \alpha < \beta \leq q$. Since $V_\alpha(\Sigma)$ and $V_\beta(\Sigma)$ are disjoint,

$$F_{\alpha,\beta}(\Sigma) = \sum_{i,j=1}^n a_{ij} \mathbf{1}_{\{\xi_i=\alpha\}} \mathbf{1}_{\{\xi_j=\beta\}}.$$

Each edge between the two color classes is counted exactly once. The diagonal terms vanish because $a_{ii} = 0$. For $i \neq j$, independence gives

$$\mathbb{E}_{\text{col}} \left[\mathbf{1}_{\{\xi_i=\alpha\}} \mathbf{1}_{\{\xi_j=\beta\}} \right] = p_{i,\alpha} p_{j,\beta}.$$

Therefore

$$\mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)] = \sum_{i,j=1}^n a_{ij} p_{i,\alpha} p_{j,\beta}.$$

Using (5.27), we obtain

$$\frac{1}{n^2} \mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)] = \int_{A_\alpha \times A_\beta} W_G^\phi(x, y) dx dy.$$

On $A_\alpha \times A_\beta$, the graphon $W_q^\star$ equals $1/2$. Since

$$\lambda^2(A_\alpha \times A_\beta) = \frac{1}{q^2},$$

we have

$$\int_{A_\alpha \times A_\beta} W_q^\star(x, y) dx dy = \frac{1}{2q^2}.$$

By Definition 3.5, applied with $S = A_\alpha$ and $T = A_\beta$, and by (5.25),

$$\begin{aligned} & \left| \int_{A_\alpha \times A_\beta} W_G^\phi - \int_{A_\alpha \times A_\beta} W_q^\star \right| \\ &= \left| \int_{A_\alpha \times A_\beta} (W_G^\phi - W_q^\star) \right| \\ &\leq \|W_G^\phi - W_q^\star\|_\square \\ &< \rho. \end{aligned}$$

Consequently,

$$(5.29) \quad \left| \mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)] - \frac{n^2}{2q^2} \right| \leq \rho n^2.$$

Thus $n^2/(2q^2)$ is the deterministic target value; in general it is not the exact value of $\mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)]$.

Changing one coordinate of C can change the counted status only of edges incident to the corresponding vertex. There are at most $n - 1$ such edges, and each contributes at most 1 to the change. Hence $F_{\alpha,\beta}$ changes by at most $n - 1$, and therefore by at most n . By Lemma 5.12(iii), with

$$t = \frac{\varepsilon n^2}{40q^2},$$

we obtain

$$\mathbb{P}_{\text{col}} \left(|F_{\alpha,\beta}(\Sigma) - \mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)]| > \frac{\varepsilon n^2}{40q^2} \right) \leq 2 \exp \left(-\frac{\varepsilon^2}{800q^4} n \right).$$

Define

$$\Omega_{\text{cross}} := \left\{ C \in [q]^n : \begin{array}{l} |F_{\alpha,\beta}(C) - \mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)]| \leq \frac{\varepsilon n^2}{40q^2} \\ \text{for every } 1 \leq \alpha < \beta \leq q \end{array} \right\}.$$

By the union bound,

$$(5.30) \quad \mathbb{P}_{\text{col}}(\Omega_{\text{cross}}^c) \leq q(q-1) \exp \left(-\frac{\varepsilon^2}{800q^4} n \right).$$

If $C \in \Omega_{\text{cross}}$, then for every $\alpha < \beta$, the definition of Ω_{cross} , (5.29), and the triangle inequality give

$$\begin{aligned} & \left| e_G(V_\alpha(C), V_\beta(C)) - \frac{n^2}{2q^2} \right| \\ &\leq |F_{\alpha,\beta}(C) - \mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)]| + \left| \mathbb{E}_{\text{col}}[F_{\alpha,\beta}(\Sigma)] - \frac{n^2}{2q^2} \right| \\ (5.31) \quad &\leq \frac{\varepsilon n^2}{40q^2} + \rho n^2 \\ &= \frac{7\varepsilon n^2}{200q^2} \\ &\leq \frac{\varepsilon n^2}{20q^2}. \end{aligned}$$

Assume also that $C \in \Omega_{\text{bal}}$. Let

$$|V_\alpha(C)| = \frac{n}{q} + \Delta_\alpha, \quad |V_\beta(C)| = \frac{n}{q} + \Delta_\beta.$$

Then

$$|\Delta_\alpha|, |\Delta_\beta| \leq \frac{\varepsilon n}{10q}.$$

Hence

$$\begin{aligned} \left| |V_\alpha(C)||V_\beta(C)| - \frac{n^2}{q^2} \right| &\leq \frac{n}{q}(|\Delta_\alpha| + |\Delta_\beta|) + |\Delta_\alpha \Delta_\beta| \\ &\leq \frac{\varepsilon n^2}{5q^2} + \frac{\varepsilon^2 n^2}{100q^2} \\ &\leq \frac{21\varepsilon n^2}{100q^2} \\ &\leq \frac{\varepsilon n^2}{4q^2}. \end{aligned}$$

Therefore

$$(5.32) \quad \left| \frac{1}{2} |V_\alpha(C)||V_\beta(C)| - \frac{n^2}{2q^2} \right| \leq \frac{\varepsilon n^2}{8q^2}.$$

Combining (5.31) and (5.32), and using the triangle inequality, gives

$$\begin{aligned} &\left| e_G(V_\alpha(C), V_\beta(C)) - \frac{1}{2} |V_\alpha(C)||V_\beta(C)| \right| \\ &\leq \left| e_G(V_\alpha(C), V_\beta(C)) - \frac{n^2}{2q^2} \right| + \left| \frac{n^2}{2q^2} - \frac{1}{2} |V_\alpha(C)||V_\beta(C)| \right| \\ &\leq \frac{\varepsilon n^2}{20q^2} + \frac{\varepsilon n^2}{8q^2} = \frac{7\varepsilon n^2}{40q^2}. \end{aligned}$$

The balance event also gives

$$|V_\alpha(C)||V_\beta(C)| \geq \frac{81}{100} \frac{n^2}{q^2} \geq \frac{4}{5} \frac{n^2}{q^2}.$$

Thus

$$\begin{aligned} \left| \frac{e_G(V_\alpha(C), V_\beta(C))}{|V_\alpha(C)||V_\beta(C)|} - \frac{1}{2} \right| &\leq \frac{7\varepsilon n^2/(40q^2)}{(4/5)n^2/q^2} \\ &= \frac{7\varepsilon}{32} \\ &< \varepsilon. \end{aligned}$$

By symmetry, the same estimate holds for every ordered pair of distinct parts.

Fixing a coloring with the desired properties. Choose $N_{\det}(\varepsilon, q)$ so large that, for every $n \geq N_{\det}$,

$$(5.33) \quad \begin{aligned} n &\geq 2q, \\ 2q \exp\left(-\frac{\varepsilon^2}{50q^2}n\right) &\leq \frac{1}{40}, \\ q(q-1) \exp\left(-\frac{\varepsilon^2}{800q^4}n\right) &\leq \frac{1}{40}. \end{aligned}$$

It follows from (5.26), (5.28), (5.30), and (5.33) that

$$\mathbb{P}_{\text{col}}(\Omega_{\text{bal}}^c) \leq \frac{1}{40}, \quad \mathbb{P}_{\text{col}}(\Omega_{\text{in}}^c) \leq \frac{1}{20}, \quad \mathbb{P}_{\text{col}}(\Omega_{\text{cross}}^c) \leq \frac{1}{40}.$$

Therefore

$$\begin{aligned}\mathbb{P}_{\text{col}}(\Omega_{\text{bal}} \cap \Omega_{\text{in}} \cap \Omega_{\text{cross}}) &\geq 1 - \frac{1}{40} - \frac{1}{20} - \frac{1}{40} \\ &= \frac{9}{10} > 0.\end{aligned}$$

So there exists a coloring

$$C \in \Omega_{\text{bal}} \cap \Omega_{\text{in}} \cap \Omega_{\text{cross}}.$$

Set

$$V_\alpha := V_\alpha(C), \quad \alpha \in [q].$$

The estimates above show that this partition has the required part sizes, the required between-part edge densities, and the required bound on the number of edges inside the parts.

Finally, let G' be obtained from G by deleting all edges inside the parts $V_1, \dots, V_q$. Then G' is q -partite, and

$$|E(G) \triangle E(G')| = \sum_{\alpha=1}^q e_G(V_\alpha) \leq \frac{\varepsilon n^2}{10q}.$$

By the definition of d_{edit} in Section 1,

$$d_{\text{edit}}(G, G') = \frac{|E(G) \triangle E(G')|}{n^2} \leq \frac{\varepsilon}{10q}.$$

□

We now use the cut-distance concentration theorem to prove the finite-graph structure theorem.

Proof of Theorem 2.7(iii). Fix $\varepsilon > 0$. Since Lemma 5.13 is stated for parameters at most 1, set

$$\eta := \min\{\varepsilon, 1\}, \quad \rho := \frac{\eta}{100q^2}.$$

Let $\mathbf{G}_n \sim \mathbb{P}_{n,r}$. Applying Theorem 2.7(ii) with the parameter ρ , we get constants

$$c_0 = c_0(\rho, r) > 0 \quad \text{and} \quad N_0 = N_0(\rho, r, w)$$

such that, for every $n \geq N_0$,

$$\mathbb{P}(\delta_{\square}(W_{\mathbf{G}_n}, B_q^*) \geq \rho) \leq \exp(-c_0 n^2).$$

Let

$$N_{\text{det}} = N_{\text{det}}(\eta, q)$$

be the threshold from Lemma 5.13, and define

$$n_0 := \max\{N_0, N_{\text{det}}\}.$$

Then, for every $n \geq n_0$, with probability at least $1 - \exp(-c_0 n^2)$, we have

$$\delta_{\square}(W_{\mathbf{G}_n}, B_q^*) < \rho.$$

On this event, Lemma 5.13, applied with η in place of ε , gives a partition

$$[n] = V_1 \sqcup \dots \sqcup V_q$$

satisfying all the conclusions of that lemma with parameter η . Since

$$\eta \leq \varepsilon,$$

the same partition satisfies the required bounds in Theorem 2.7(iii) with parameter ε .

In particular, if G'_n is obtained from G_n by deleting all edges whose endpoints lie in the same part, then G'_n is q -partite and

$$d_{\text{edit}}(G_n, G'_n) \leq \frac{\eta}{10q} \leq \frac{\varepsilon}{10q} \leq \frac{\varepsilon}{2},$$

because $q \geq 2$.

Thus the theorem holds with

$$c := c_0.$$

Since $\rho = \eta/(100q^2)$, and since $q = r - 1$, the constant c depends only on ε and r . The threshold n_0 may also depend on w . $\square$

6. EXAMPLES AND CONSEQUENCES

We first discuss the boundary case $r = 2$ and the first case covered by Theorem 2.7, namely $r = 3$. We then show that the induced subgraph on any fixed number of labeled vertices converges in distribution to a stochastic block model, derive the limiting clique densities, and give an exponential bound for colorability with fewer than $r - 1$ colors. Finally, we extend the model to several fixed clique penalties.

6.1. The case $r = 2$. The notions defined in Section 2 are straightforwardly extended to the case $r = 2$, and we get

$$N_2(G) = n^2 - 2|E(G)|,$$

and

$$\tilde{Z}_{n,2} = (1 + e^{-2w})^{\binom{n}{2}}.$$

The possible edges are therefore independent, and each is present with probability

$$\frac{e^{-2w}}{1 + e^{-2w}} = \frac{1}{1 + e^{2w}}.$$

Thus $\mathbb{P}_{n,2}$ is exactly the $G(n, 1/(1 + e^{2w}))$ distribution. In particular, the $(r - 1)$ -partite behavior in Theorem 2.7 begins at $r = 3$.

6.2. The case $r = 3$. Here $q = r - 1 = 2$. By Corollary 2.6, the reduced mass of a graph G is

$$e^{-6wK_3(G)}.$$

Thus triangles are penalized. By Theorem 2.7(ii), the empirical graphon converges in probability, in cut distance, to the class represented by W_2^* . The canonical representative W_2^* has two equal blocks and takes value 0 within each block and $1/2$ between the blocks.

6.3. A stochastic block model limit. For $m \in \mathbb{N}$ and probability measures α and β on $\mathbf{G}_m$, their *total variation distance* is

$$d_{\text{TV}}(\alpha, \beta) := \frac{1}{2} \sum_{J \in \mathbf{G}_m} |\alpha(J) - \beta(J)|.$$

Definition 6.1. For $m \in \mathbb{N}$, $J \in \mathbf{G}_m$, and a graphon W , define the *induced density* of J in W by

$$t_{\text{ind}}(J, W) := \int_{[0,1]^m} \prod_{\{i,j\} \in E(J)} W(x_i, x_j) \prod_{\substack{1 \leq i < j \leq m \\ \{i,j\} \notin E(J)}} (1 - W(x_i, x_j)) dx_1 \cdots dx_m.$$

This is the induced-density formula in [22, Sec. 7.2, Eq. (7.3)]. Under the graphon sampling construction from Section 3, it equals the probability of obtaining the labeled graph J ; see [22, Sec. 10.1, immediately before Eq. (10.3)].

Corollary 6.2 (Stochastic block model limit for fixed induced subgraphs). *Fix an integer $r \geq 3$, set $q = r - 1$, and let $w > 0$ and $m \in \mathbb{N}$. For every $n \geq m$, let $\mathbf{G}_n \sim \mathbb{P}_{n,r}$, and let $\mathbf{H}_{n,m}$ be the subgraph of $\mathbf{G}_n$ induced by $[m]$.*

Define a random graph $\mathbf{S}_m^{(q)}$ on $[m]$ as follows. Choose $C_1, \dots, C_m$ independently and uniformly from $[q]$. Given these labels, include the possible edges independently, with probability

$$\mathbb{P}_{\text{SBM}}(\{i, j\} \in E(\mathbf{S}_m^{(q)}) \mid C_1, \dots, C_m) = \begin{cases} 0, & C_i = C_j, \\ \frac{1}{2}, & C_i \neq C_j, \end{cases} \quad 1 \leq i < j \leq m.$$

Here $\mathbb{P}_{\text{SBM}}(\mathbf{S}_m^{(q)})$ denotes probability of getting $\mathbf{S}_m^{(q)}$ under this construction. Define probability measures $\nu_{n,m}$ and $\nu_m^{(q)}$ on $\mathbf{G}_m$ by

$$\nu_{n,m}(J) := \mathbb{P}(\mathbf{H}_{n,m} = J), \quad \nu_m^{(q)}(J) := \mathbb{P}_{\text{SBM}}(\mathbf{S}_m^{(q)} = J) \quad \text{for } J \in \mathbf{G}_m.$$

Then

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(\nu_{n,m}, \nu_m^{(q)}) = 0.$$

The limiting measure $\nu_m^{(q)}$ depends on $q = r - 1$ but not on w .

Proof. Fix $J \in \mathbf{G}_m$. For the induced density from Definition 6.1, inclusion-exclusion gives [22, Sec. 7.2, Eq. (7.4)]

$$t_{\text{ind}}(J, W) = \sum_{\substack{F \in \mathbf{G}_m \\ E(J) \subseteq E(F)}} (-1)^{|E(F) \setminus E(J)|} t(F, W).$$

By [7, Thm. 3.7(a)], each homomorphism density in this sum is continuous in cut distance. Therefore $t_{\text{ind}}(J, \cdot)$ is also continuous in cut distance.

Theorem 2.7(ii) and the final identity in Notation 4.3 give, for every $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|t_{\text{ind}}(J, W_{\mathbf{G}_n}) - t_{\text{ind}}(J, W_q^*)| \geq \varepsilon) = 0.$$

Since the absolute difference is bounded by 1, for every $\varepsilon > 0$, its expectation is at most ε plus the probability that it is at least ε . Hence,

$$(6.1) \quad \lim_{n \rightarrow \infty} \mathbb{E}[|t_{\text{ind}}(J, W_{\mathbf{G}_n}) - t_{\text{ind}}(J, W_q^*)|] = 0.$$

Define a probability measure $\bar{\nu}_{n,m}$ on $\mathbf{G}_m$ by

$$\bar{\nu}_{n,m}(J) := \mathbb{E}[t_{\text{ind}}(J, W_{\mathbf{G}_n})].$$

By (2.3), the probability assigned to a graph G depends only on $K_r(G)$, which is invariant under vertex relabeling. Hence the distribution of $\mathbf{G}_n$ is invariant under vertex relabeling. Consequently, if $\iota : [m] \rightarrow [n]$ is an independent uniformly chosen injection, then $\mathbf{H}_{n,m}$ has the same distribution as the graph on $[m]$ in which $\{i, j\}$ is an edge exactly when $\{\iota(i), \iota(j)\} \in E(\mathbf{G}_n)$.

On the other hand, sampling from W_G corresponds to choosing vertices of G independently, so the same vertex may be chosen more than once. The probability of choosing the same vertex more than once is at most $\binom{m}{2}/n$. Therefore, by [22, Eq. (10.1)],

$$d_{\text{TV}}(\nu_{n,m}, \bar{\nu}_{n,m}) \leq \frac{\binom{m}{2}}{n}.$$

By Notation 4.3, graphon sampling from W_q^* is exactly the construction of $\mathbf{S}_m^{(q)}$, therefore

$$t_{\text{ind}}(J, W_q^*) = \nu_m^{(q)}(J).$$

It follows that

$$\begin{aligned} d_{\text{TV}}(\nu_{n,m}, \nu_m^{(q)}) &\leq \frac{\binom{m}{2}}{n} + d_{\text{TV}}(\bar{\nu}_{n,m}, \nu_m^{(q)}) \\ &\leq \frac{\binom{m}{2}}{n} + \frac{1}{2} \sum_{J \in \mathbf{G}_m} \mathbb{E} [|t_{\text{ind}}(J, W_{\mathbf{G}_n}) - t_{\text{ind}}(J, W_q^*)|]. \end{aligned}$$

The sum is finite, and each term tends to zero by (6.1). This proves the result. $\square$

Remark 6.3 (The limiting model is not Erdős-Rényi). Let $q \geq 2$ and $m \geq 2$, and consider the model $\mathbf{S}_m^{(q)}$ defined above. For $1 \leq i < j \leq m$, let

$$X_{ij} := \mathbf{1}_{\{\{i,j\} \in E(\mathbf{S}_m^{(q)})\}}.$$

For each edge,

$$\mathbb{P}_{\text{SBM}}(X_{ij} = 1) = \mathbb{P}_{\text{SBM}}(C_i \neq C_j) \frac{1}{2} = \frac{q-1}{q} \cdot \frac{1}{2} = \frac{q-1}{2q}.$$

Any two distinct edge indicators are independent. This is immediate for vertex-disjoint edges because they use different labels and different edge choices. If the edges share a vertex, say $\{i, j\}$ and $\{i, k\}$, then

$$\mathbb{P}_{\text{SBM}}(X_{ij} = X_{ik} = 1) = \left(\frac{q-1}{q}\right)^2 \frac{1}{4} = \left(\frac{q-1}{2q}\right)^2,$$

which is the product of their individual probabilities.

However, if $m \geq 3$, the three indicators associated with the triangle on $\{1, 2, 3\}$ are not independent as a triple, since

$$\mathbb{P}_{\text{SBM}}(X_{12} = X_{13} = X_{23} = 1) = \frac{(q-1)(q-2)}{8q^2},$$

whereas independence would give

$$\left(\frac{q-1}{2q}\right)^3.$$

Indeed,

$$\left(\frac{q-1}{2q}\right)^3 - \frac{(q-1)(q-2)}{8q^2} = \frac{q-1}{8q^3} > 0.$$

Thus any two edge indicators are independent, but the three triangle indicators are not independent as a triple. In particular, $\mathbf{S}_m^{(q)}$ is not an Erdős-Rényi graph when $m \geq 3$.

6.4. Limiting clique densities. For any fixed integer $r \geq 3$ we now determine the limiting density, with respect to $\mathbb{P}_{n,r}$, of K_s for all $s \geq 2$.

Corollary 6.4. *Fix an integer $r \geq 3$ and let $w > 0$. Let $\mathbf{G}_n \sim \mathbb{P}_{n,r}$, and set $q = r - 1$. For every fixed integer $s \geq 2$,*

$$(6.2) \quad t(K_s, W_q^*) = \begin{cases} \frac{q(q-1) \cdots (q-s+1)}{q^s 2^{\binom{s}{2}}}, & 2 \leq s \leq q, \\ 0, & s > q. \end{cases}$$

Moreover, for every $\eta > 0$, there are constants $c = c(\eta, r, s) > 0$ and $N = N(\eta, r, s, w) \geq s$ such that, for every $n \geq N$,

$$(6.3) \quad \mathbb{P}_{n,r} \left(\left\{ G \in \mathbf{G}_n : \left| \frac{s! K_s(G)}{n^s} - t(K_s, W_q^*) \right| \geq \eta \right\} \right) \leq e^{-cn^2}.$$

Consequently, for every $\eta > 0$,

$$(6.4) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{K_s(\mathbf{G}_n)}{\binom{n}{s}} - t(K_s, W_q^*) \right| \geq \eta \right) = 0.$$

Thus every fixed clique K_s with $2 \leq s \leq r-1$ has a positive limiting density, while every fixed clique K_s with $s \geq r$ has limiting density zero.

Proof. Suppose first that $2 \leq s \leq q$. By Definition 3.3,

$$t(K_s, W_q^*) = \int_{[0,1]^s} \prod_{1 \leq i < j \leq s} W_q^*(x_i, x_j) dx_1 \cdots dx_s.$$

Equivalently, choose $x_1, \dots, x_s$ independently and uniformly at random from $[0, 1]$. By Notation 4.3, the index of the canonical block containing each x_i is uniform on $[q]$, and these block indices are independent. Since W_q^* is zero within each block, the integrand is nonzero only when $x_1, \dots, x_s$ lie in distinct blocks. The set of such points has measure

$$\frac{q(q-1) \cdots (q-s+1)}{q^s}.$$

On this set, every factor in the integrand is equal to $1/2$. Since K_s has $\binom{s}{2}$ edges, it follows that

$$t(K_s, W_q^*) = \frac{q(q-1) \cdots (q-s+1)}{q^s 2^{\binom{s}{2}}}.$$

If $s > q$, then among any s points $x_1, \dots, x_s$, at least two lie in the same canonical block. The corresponding factor in the defining integral is zero. Hence

$$t(K_s, W_q^*) = 0.$$

This proves (6.2).

By Lemma 3.18, with s in place of r , every graph $G \in \mathbf{G}_n$ satisfies

$$t(K_s, W_G) = \frac{s! K_s(G)}{n^s}.$$

Lemma 3.14 and Notation 4.3 therefore give

$$\left| \frac{s! K_s(G)}{n^s} - t(K_s, W_q^*) \right| \leq \binom{s}{2} \delta_{\square}(W_G, B_q^*).$$

Thus, if the left-hand side is at least η , then

$$\delta_{\square}(W_G, B_q^*) \geq \frac{\eta}{\binom{s}{2}}.$$

Applying Theorem 2.7(ii) with $\eta/\binom{s}{2}$ in place of ε proves (6.3).

Finally, for $n \geq s$, set

$$X_n := \frac{s! K_s(\mathbf{G}_n)}{n^s}, \quad a_n := \frac{n(n-1) \cdots (n-s+1)}{n^s}.$$

Then

$$\frac{K_s(\mathbf{G}_n)}{\binom{n}{s}} = \frac{X_n}{a_n}.$$

Equation (6.3) says that X_n converges in probability to $t(K_s, W_q^*)$. Since $a_n > 0$ and

$$\lim_{n \rightarrow \infty} a_n = 1,$$

the quotient X_n/a_n converges in probability to the same limit. This is exactly (6.4). $\square$

Example 6.5. For example, let $r = 4$, so that $q = 3$, and let $\mathbf{G}_n \sim \mathbb{P}_{n,4}$. Then

$$t(K_3, W_3^\star) = \frac{3 \cdot 2 \cdot 1}{3^3 2^3} = \frac{1}{36}, \quad t(K_4, W_3^\star) = 0.$$

Hence, for every $\eta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{K_3(\mathbf{G}_n)}{\binom{n}{3}} - \frac{1}{36} \right| \geq \eta \right) = 0, \quad \lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{K_4(\mathbf{G}_n)}{\binom{n}{4}} \geq \eta \right) = 0.$$

Thus a positive proportion of triples still form triangles, even though the proportion of four-element sets that form a copy of K_4 tends to zero.

6.5. Colorability with fewer than $r - 1$ colors.

Corollary 6.6. Fix an integer $r \geq 3$ and let $w > 0$. Set $q = r - 1$, and let $\mathbf{G}_n \sim \mathbb{P}_{n,r}$. Recall that $\chi(\mathbf{G}_n)$ denotes the chromatic number of $\mathbf{G}_n$. For every $n \in \mathbb{N}$ and every integer k with $1 \leq k < q$,

$$(6.5) \quad \mathbb{P}(\chi(\mathbf{G}_n) \leq k) \leq \exp \left(-\frac{(q-k) \ln 2}{2kq} n^2 + n \ln k + \frac{q \ln 2}{8} \right).$$

Consequently, there are constants $c = c(r, k) > 0$ and $N = N(r, k) \in \mathbb{N}$ such that

$$\mathbb{P}(\chi(\mathbf{G}_n) \leq k) \leq e^{-cn^2} \quad \text{for every } n \geq N.$$

In particular, when $r \geq 4$, the probability that $\mathbf{G}_n$ is $(r-2)$ -partite tends to zero exponentially in n^2 .

Proof. Fix a map from $[n]$ to a set of k colors, and let $m_1, \dots, m_k$ be the sizes of its color classes. A graph for which this map is a proper coloring can have edges only between different classes. The number of pairs in different classes is

$$\frac{1}{2} \left(n^2 - \sum_{i=1}^k m_i^2 \right) \leq \frac{1}{2} \left(1 - \frac{1}{k} \right) n^2,$$

where the inequality follows from Lemma 3.12(ii). Hence at most

$$2^{(1-1/k)n^2/2}$$

graphs have this fixed proper coloring. Since there are k^n such maps,

$$(6.6) \quad |\{G \in \mathbf{G}_n : \chi(G) \leq k\}| \leq k^n 2^{(1-1/k)n^2/2}.$$

A k -colorable graph has no clique with more than k vertices. Since $k < q < r$, every graph counted in (6.6) is K_r -free and has reduced mass 1. Hence (6.6) and (5.17) give

$$\mathbb{P}(\chi(\mathbf{G}_n) \leq k) \leq \exp \left(\frac{\ln 2}{2} \left(1 - \frac{1}{k} \right) n^2 + n \ln k - s_q n^2 + \frac{q \ln 2}{8} \right).$$

Using (5.1) gives (6.5). Since $q - k > 0$, the coefficient of n^2 is negative, which proves the final estimate. $\square$

6.6. Several fixed clique penalties. We now consider a set-up with soft constraints for several clique sizes. Fix an integer $\ell \geq 3$. For each $k \in \{3, \dots, \ell\}$ let a weight $w_k \geq 0$ be fixed. Assume that at least one of these w_k is positive and set

$$r := \min\{k \in \{3, \dots, \ell\} : w_k > 0\}, \quad q := r - 1.$$

Thus $w_k = 0$ for $k < r$. For all $k \in \{3, \dots, \ell\}$ and every $G \in \mathbf{G}_n$, define

$$N_k(G) := n^k - k! K_k(G).$$

Hence $N_k(G)$ is the number of k -tuples of vertices of G that do not form a k -clique in G . We now define the “*unnormalized mass*” of $G \in \mathbf{G}_n$ as the number

$$\exp \left(\sum_{k=3}^{\ell} w_k N_k(G) \right) = \exp \left(\sum_{k=r}^{\ell} w_k n^k \right) \exp \left(- \sum_{k=r}^{\ell} k! w_k K_k(G) \right).$$

The first factor is independent of G and cancels after normalization. We therefore define the “*reduced mass*” of $G \in \mathbf{G}_n$ by

$$M_n^{\text{mc}}(G) := \exp \left(- \sum_{k=r}^{\ell} k! w_k K_k(G) \right),$$

and then the corresponding reduced partition function, probability measure, and reduced free energy are defined by

$$\begin{aligned} \tilde{Z}_n^{\text{mc}} &:= \sum_{G \in \mathbf{G}_n} M_n^{\text{mc}}(G), \\ (6.7) \quad \mathbb{P}_n^{\text{mc}}(A) &:= \frac{1}{\tilde{Z}_n^{\text{mc}}} \sum_{G \in A} M_n^{\text{mc}}(G), \quad \text{for } A \subseteq \mathbf{G}_n, \quad \text{and} \\ F_n^{\text{mc}} &:= \frac{1}{n^2} \ln \tilde{Z}_n^{\text{mc}}. \end{aligned}$$

Theorem 6.7 (The smallest positively weighted clique determines the limit structure). *Let $\mathbf{G}_n^{\text{mc}} \sim \mathbb{P}_n^{\text{mc}}$, that is, $\mathbf{G}_n$ is a random graph from $\mathbf{G}_n$ under the probability distribution $\mathbb{P}_n^{\text{mc}}$. Then the following statements hold:*

(i) $\lim_{n \rightarrow \infty} F_n^{\text{mc}} = \frac{\ln 2}{2} \left(1 - \frac{1}{r-1} \right).$

(ii) For every $\varepsilon > 0$, there are constants

$$c = c(\varepsilon, r) > 0 \quad \text{and} \quad N = N(\varepsilon, r, w_r)$$

such that, for every $n \geq N$,

$$\mathbb{P}(\delta_{\square}(W_{\mathbf{G}_n^{\text{mc}}}, B_q^*) \geq \varepsilon) \leq e^{-cn^2}.$$

(iii) For every $\varepsilon > 0$, there are constants

$$c = c(\varepsilon, r) > 0 \quad \text{and} \quad N = N(\varepsilon, r, w_r)$$

such that, for every $n \geq N$, with probability at least $1 - e^{-cn^2}$, there is a partition

$$[n] = V_1 \sqcup \cdots \sqcup V_q$$

with the following properties:

(a) For every $i \in [q]$,

$$||V_i| - n/q| \leq \varepsilon n.$$

(b) For all distinct $i, j \in [q]$,

$$\left| \frac{e_{\mathbf{G}_n^{\text{mc}}}(V_i, V_j)}{|V_i||V_j|} - \frac{1}{2} \right| \leq \varepsilon.$$

(c) The total number of edges inside parts is

$$\sum_{i=1}^q e_{\mathbf{G}_n^{\text{mc}}}(V_i) \leq \frac{\varepsilon n^2}{10q}.$$

(d) The total density of edges inside the parts is at most ε :

$$\frac{\sum_{i=1}^q e_{\mathbf{G}_n^{\text{mc}}}(V_i)}{\sum_{i=1}^q \binom{|V_i|}{2}} \leq \varepsilon.$$

If $\mathbf{G}'_n$ is obtained by deleting all edges whose endpoints lie in the same part, then $\mathbf{G}'_n$ is q -partite and

$$d_{\text{edit}}(\mathbf{G}_n^{\text{mc}}, \mathbf{G}'_n) \leq \frac{\varepsilon}{10q} \leq \frac{\varepsilon}{2}.$$

In particular, the reduced free-energy limit and the cut-distance limit depend on the weight vector $(w_3, \dots, w_\ell)$ only through r . Once r and w_r are fixed, the constants in parts (ii) and (iii) can be chosen independently of ℓ and of the weights $w_{r+1}, \dots, w_\ell$.

Proof. Let $\tilde{Z}_{n,r}$, $F_{n,r}(w_r)$, and $\mathbb{P}_{n,r}$ denote the reduced partition function, reduced free energy, and probability measure of the single- K_r model with weight w_r , and set

$$C_q := \frac{q \ln 2}{8}.$$

Since the higher weights are nonnegative, every $G \in \mathbf{G}_n$ satisfies

$$(6.8) \quad 0 < M_n^{\text{mc}}(G) \leq \exp(-r! w_r K_r(G)) \leq 1.$$

Every graph in the balanced q -partite family used in the proof of Lemma 5.11 is K_r -free and hence K_k -free for every $k \geq r$. Its multi-clique reduced mass is therefore 1. Let

$$s_q := \frac{\ln 2}{2} \left(1 - \frac{1}{q}\right).$$

The same lower-bound construction, together with (6.8), gives

$$(6.9) \quad \exp(s_q n^2 - C_q) \leq \tilde{Z}_n^{\text{mc}} \leq \tilde{Z}_{n,r}.$$

This comparison explains the role of the smallest positively weighted clique; the lower bound is unaffected by any clique penalty, while the upper bound retains only the K_r penalty.

For part (i), taking logarithms in (6.9) and dividing by n^2 gives

$$s_q - \frac{C_q}{n^2} \leq F_n^{\text{mc}} \leq F_{n,r}(w_r).$$

The conclusion follows from Theorem 2.7(i).

For parts (ii) and (iii), set

$$R_n := \frac{\tilde{Z}_{n,r}}{\tilde{Z}_n^{\text{mc}}}.$$

By (6.8), every event $\mathcal{A} \subseteq \mathbf{G}_n$ satisfies

$$(6.10) \quad \mathbb{P}_n^{\text{mc}}(\mathcal{A}) \leq R_n \mathbb{P}_{n,r}(\mathcal{A}).$$

Moreover, (6.9) implies

$$0 \leq \frac{1}{n^2} \ln R_n \leq F_{n,r}(w_r) - s_q + \frac{C_q}{n^2},$$

and hence, by Theorem 2.7(i),

$$(6.11) \quad \lim_{n \rightarrow \infty} \frac{1}{n^2} \ln R_n = 0.$$

The upper bound depends only on n , r , and w_r , and it tends to 0. Since the same bound holds simultaneously for every $\ell \geq r$ and every choice of $w_{r+1}, \dots, w_\ell \geq 0$, the convergence is uniform in those parameters.

Consequently, suppose that a sequence of events $(\mathcal{A}_n)$ satisfies

$$\mathbb{P}_{n,r}(\mathcal{A}_n) \leq e^{-an^2}$$

for some $a > 0$ and all sufficiently large n . After increasing the threshold if necessary, (6.11) gives $R_n \leq e^{an^2/2}$, and (6.10) gives

$$\mathbb{P}_n^{\text{mc}}(\mathcal{A}_n) \leq e^{-an^2/2}.$$

For part (ii), apply this observation to

$$\mathcal{A}_n = \{G \in \mathbf{G}_n : \delta_{\square}(W_G, B_q^*) \geq \varepsilon\}$$

and use Theorem 2.7(ii).

For part (iii), let $\mathcal{B}_{n,\varepsilon}$ be the set of graphs that admit no partition satisfying the five properties stated there. Theorem 2.7(iii) gives an exponential bound for $\mathbb{P}_{n,r}(\mathcal{B}_{n,\varepsilon})$, so the same observation gives the required bound under $\mathbb{P}_n^{\text{mc}}$. The edit-distance conclusion follows from any partition satisfying those five properties, exactly as in Theorem 2.7(iii).

For each of parts (ii) and (iii), one may choose $c = c(\varepsilon, r) > 0$ and $N = N(\varepsilon, r, w_r)$. Neither choice depends on ℓ or on the higher weights $w_{r+1}, \dots, w_\ell$. $\square$

REFERENCES

- [1] Juan Alvarado, Yuyi Wang, Jan Ramon, Limits of multi-relational graphs, *Machine Learning*, Vol. 112 (2023) 177–216.
- [2] Balogh, J., Bollobás, B., and Simonovits, M., The typical structure of graphs without given excluded subgraphs, *Random Structures and Algorithms*, Vol. 34 (2009) 305–318.
- [3] Balogh, J., Bollobás, B., and Simonovits, M., The fine structure of octahedron-free graphs, *Journal of Combinatorial Theory, Series B*, Vol. 101 (2011) 67–84.
- [4] J. Balogh and J. Butterfield, Excluding induced subgraphs: Critical graphs, *Random Structures and Algorithms*, Vol. 38 (2011) 100–120.
- [5] József Balogh, Robert Morris, Wojciech Samotij and Lutz Warnke, The typical structure of sparse K_{r+1} -free graphs, *Transactions of the American Mathematical Society*, Vol. 368 (2016), 6439–6485.
- [6] József Balogh, Neal Bushaw, Maurício Collares, Hong Liu, Robert Morris, and Maryam Sharifzadeh, The typical structure of graphs with no large cliques, *Combinatorica*, Vol. 37 (2017) 617–632.
- [7] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztegombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Adv. Math.* **219** (2008), no. 6, 1801–1851, DOI: 10.1016/j.aim.2008.07.008.
- [8] Guy Van den Broeck, Kristian Kersting, Sriraam Natarajan, David Poole (editors), *An Introduction to Lifted Probabilistic Inference*, MIT Press (2021).
- [9] S. Chatterjee and P. Diaconis, Estimating and understanding exponential random graph models, *Ann. Statist.* **41** (2013), no. 5, 2428–2461, DOI: 10.1214/13-AOS1155.
- [10] S. Chatterjee and S. R. S. Varadhan, The large deviation principle for the Erdős–Rényi random graph, *European J. Combin.* **32** (2011), no. 7, 1000–1017, DOI: 10.1016/j.ejc.2011.03.014.
- [11] Luc De Raedt, Kristian Kersting, Sriraam Natarajan, David Poole, *Statistical Relational Artificial Intelligence: Logic, Probability, and Computation*, Morgan & Claypool Publishers (2016).
- [12] P. Erdős, D. J. Kleitman, and B. L. Rothschild, Asymptotic enumeration of K_n -free graphs, in *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II*, Atti dei Convegni Lincei, no. 17, Accademia Nazionale dei Lincei, Rome, 1976, 19–27.
- [13] Lise Getoor, Ben Taskar (editors), *Introduction to Statistical Relational Learning*, MIT Press (2007).
- [14] J. Hladký, P. Hu, and D. Piguet, Tilings in graphons, *European J. Combin.* **93** (2021), 103284, DOI: 10.1016/j.ejc.2020.103284.
- [15] Hundack, C., Prömel, H. J., and Steger, A., Extremal graph problems for graphs with a color-critical vertex, *Combinatorics, Probability and Computing*, Vol. 2 (1993) 465–477.
- [16] S. Janson, Graphons, cut norm and distance, couplings and rearrangements, *New York J. Math. Monographs* **4** (2013), 1–76.
- [17] J. Kim, D. Kühn, D. Osthus, T. Townsend, Forbidding induced even cycles in a graph: typical structure and counting, *Journal of Combinatorial Theory, Ser. B*, Vol. 131 (2018) 170–219.
- [18] D. J. Kleitman, B. L. Rothschild, Asymptotic enumeration of partial orders on a finite set, *Transactions of the American Mathematical Society*, Vol. 205 (1975) 205–220.
- [19] Ph. G. Kolaitis, H. J. Prömel, and B. L. Rothschild, $K_{\ell+1}$ -free graphs: asymptotic structure and a 0–1 law, *Trans. Amer. Math. Soc.* **303** (1987), no. 2, 637–671, DOI: 10.1090/S0002-9947-1987-0902790-6.
- [20] Koponen, V., Random graphs with bounded maximum degree: asymptotic structure and a logical limit law, *Discrete Mathematics and Theoretical Computer Science*, Vol. 14 (2012) 229–254.

- [21] V. Koponen, Domain size asymptotics for Markov logic networks, arXiv:2509.04192v2 [cs.AI], 2026, DOI: 10.48550/arXiv.2509.04192.
- [22] L. Lovász, *Large Networks and Graph Limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, Providence, RI, 2012, DOI: 10.1090/coll/060.
- [23] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, 2nd ed., Cambridge University Press, Cambridge, 2017, DOI: 10.1017/9781316651124.
- [24] Hans Jürgen Prömel, Angelika Steger, Excluding induced subgraphs: quadrilaterals, *Random Structures and Algorithms*, Vol. 2 (1991) 55–71.
- [25] Bruce Reed, The Global Structure of a Typical Graph Without H as an Induced Subgraph when H is a Cycle, <https://arxiv.org/abs/2506.03544>.
- [26] M. Richardson and P. Domingos, Markov logic networks, *Machine Learning* **62** (2006), no. 1–2, 107–136, DOI: 10.1007/s10994-006-5833-1.
- [27] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, An introduction to exponential random graph (p^*) models for social networks, *Social Networks* **29** (2007), no. 2, 173–191, DOI: 10.1016/j.socnet.2006.08.002.
- [28] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., International Series in Pure and Applied Mathematics, McGraw-Hill, Inc., New York, 1976.
- [29] Caroline Terry, Structure and enumeration theorems for hereditary properties in finite relational languages, *Annals of Pure and Applied Logic*, Vol. 169 (2018) 413–449.
- [30] Y. Tousinejad and V. Koponen, Random coloured digraphs defined by a Markov logic network, arXiv:2606.23715 [math.LO], 2026, DOI: 10.48550/arXiv.2606.23715.
- [31] N. C. Wormald, Models of random regular graphs, in J. Lamb and D. Preece (editors), *Surveys in Combinatorics, 1999*, Vol. 267 of London Mathematical Society Lecture Note Series, 239–298, Cambridge University Press (1999).

DEPARTMENT OF MATHEMATICS, UPPSALA UNIVERSITY, SWEDEN.

Email address: `yasmin.tousinejad@math.uu.se`

DEPARTMENT OF MATHEMATICS, UPPSALA UNIVERSITY, SWEDEN.

Email address: `vera.koponen@math.uu.se`