Combinatorics, Probability and Computing (2025), **34**, pp. 680–753
doi:10.1017/S0963548325000008

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

A large deviation principle for block models

Christian Borgs[1], Jennifer Chayes[2], Julia Gaudio[3], Samantha Petti[4] and Subhabrata Sen[5]

[1]Berkeley AI Research Group, Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA, [2]Department of EECS, Math, Stat, School of Information, University of California, Berkeley, CA, USA, [3]Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA, [4]Department of Math, Tufts University, Medford, MA, USA, and [5]Department of Statistics, Harvard University, Cambridge, MA, USA

Corresponding author: Subhabrata Sen; Email: subhabratasen@fas.harvard.edu

(Received 18 March 2021; revised 20 July 2024; accepted 20 July 2024; first published online 30 June 2025)

Abstract

We initiate a study of large deviations for block model random graphs in the dense regime. Following [14], we establish an LDP for dense block models, viewed as random graphons. As an application of our result, we study upper tail large deviations for homomorphism densities of regular graphs. We identify the existence of a 'symmetric' phase, where the graph, conditioned on the rare event, looks like a block model with the same block sizes as the generating graphon. In specific examples, we also identify the existence of a 'symmetry breaking' regime, where the conditional structure is not a block model with compatible dimensions. This identifies a 'reentrant phase transition' phenomenon for this problem – analogous to one established for Erdős–Rényi random graphs [13, 14]. Finally, extending the analysis of [34], we identify the precise boundary between the symmetry and symmetry breaking regimes for homomorphism densities of regular graphs and the operator norm on Erdős–Rényi bipartite graphs.

Keywords: Large deviation; stochastic block model; symmetry/symmetry breaking; bipartite Erdős–Rényi graph

2020 MSC Codes: Primary: 60F10; Secondary: 05C80, 60C05

1. Introduction

The study of large deviation problems on random graphs has a long and rich history in Probability and Combinatorics. Research in this area is motivated by the following fundamental question: *What is the structure of a random graph, conditioned on a rare event?*

In a seminal paper, Chatterjee and Varadhan [14] formalised this question by combining the theory of graph limits [32] with classical Large Deviations theory [19], and established a Large Deviation Principle (LDP) for the Erdős–Rényi binomial random graph $G(n, p)$. This is the simplest random graph model, constructed by adding edges independently among n vertices with probability p. As an application of this LDP, Chatterjee and Varadhan [14] examined upper tail large deviations for regular subgraph counts. The homomorphism density $t(H, G)$ of a graph H on v vertices measures the probability that H appears on v randomly chosen vertices of a graph G (see Definition 5). Let H be a d-regular graph, and for notational convenience, define the event $\mathcal{E}_\delta = \{t(H, G) > (1 + \delta)\mathbb{E}[t(H, G)]\}$. Chatterjee and Varadhan [14] established the existence of $0 < \delta_{\min}(H) < \delta_{\max}(H)$ such that if $\delta < \delta_{\min}(H)$ or $\delta > \delta_{\max}(H)$, conditioned on $\mathcal{E}_\delta$, $G(n, p)$ 'looks like' an Erdős–Rényi random graph, albeit with a higher edge density. They call this the 'replica symmetric' phase. On the contrary, [14] also establishes that for p sufficiently small, there exists

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$\delta \in [\delta_{\min}(H), \delta_{\max}(H)]$ such that conditioned on $\mathcal{E}_\delta$, the graph is not distributed as an Erdős–Rényi random graph – this regime was termed as the 'replica symmetry breaking' regime. Using the framework of [14], Lubetzky and Zhao [34] characterised the precise boundary between the symmetry and the symmetry breaking regimes, in terms of δ and p. We defer an in-depth survey of large deviations on random graphs to Section 1.7.

Random graphs are simple stochastic models for large networks observed in a myriad of scientific applications, and in this context, it is often natural to study graph models with *inhomogeneities* or *constraints*. Large deviation phenomena are of natural interest in this general setting, although progress in this direction requires several new ideas. The study of large deviations for constrained random graphs has been initiated in the recent literature – [18] studies large deviations for the uniform random graph with a given number of edges, while in [20], in joint work with Souvik Dhara, S.S. studied large deviations for random graphs with given degrees. Finally, [7] focuses on large deviations for random regular graphs in the sparse regime. In contrast, large deviations for inhomogeneous random graphs is relatively unexplored (see [7] for some preliminary results on sparse graphs). This paper seeks to fill this gap by initiating the study of large deviations for block model random graphs.

We study large deviations for block model graphs drawn from a 'base graphon' with k blocks that is specified by a set of values $\{p_{ij}\}_{1\le i\le j\le k}$, $p_{ij} \in [0, 1]$. A graph on kn vertices is sampled from this model as follows: (i) collect the vertices into k groups of size n, indexed by $1, \dots, k$, and (ii) connect two vertices from groups i and j with probability p_{ij}. (See Section 1.1 for a formal definition.) Our contributions in this article can be summarised as follows:

1. We adopt the framework of [14], and establish an LDP for block model random graphs, viewed as random graphons. The induced law of the random graph satisfies an LDP with speed n^2 – the rate function in this case is the lower semicontinuous envelope of an appropriate relative entropy functional (see formal statement as Theorem 11). Perhaps surprisingly, although the block model is quite similar to the Erdős–Rényi random graph, our derivation of the LDP requires going substantially beyond the ideas introduced in [14]. In particular, the derivation of the LDP in [14] relies heavily on the fact that an Erdős–Rényi random graph remains invariant in law under permutations of the vertices, a fact that is no longer true for general block models. To overcome this barrier, we rely on a two-step approach. (a) Using Szemerédi's Regularity Lemma, we construct a Szemerédi net of block graphons and *cover* an event by a finite union of open balls centred on the elements of this net. Thus it suffices to characterise the limiting probability of each open ball. (b) To this end, we employ a 'method of types'-style argument, similar to the classical proof of Sanov's Theorem. A similar two-step strategy was employed earlier in [20] while deriving an LDP for random graphs with given degrees. A crucial technical difference between the two settings is that in [20], the graphon being sampled from was bounded away from zero and one, whereas our results include block models that take value zero or one. As an immediate application of this general result, we obtain an LDP for the Erdős–Rényi bipartite graph.
2. Our general LDP, in turn, directly implies an LDP for graph parameters continuous with respect to the cut topology (see Theorem 16), e.g., homomorphism density, largest eigenvalue, etc. For such graph parameters, the rate function is expressed as a variational problem on the space of graphons.
3. Next, we turn our attention to the variational problem for upper tail large deviations of regular subgraphs. In Theorem 19, we establish that close to the expected value, this problem exhibits a symmetric phase – where the variational problem admits solutions which exhibit the same block structure as the base graphon. We also demonstrate that for large target values of the homomorphism density, the variational principle admits symmetric solutions.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

4. In some specific block graphons, we exhibit the existence of a non-symmetric phase – where there does not exist a symmetric minimiser (see Section 1.5 for the specific examples). This establishes an analogue of the reentrant phase transition[1] phenomenon, noted earlier for the upper tail problem on Erdős–Rényi random graphs [13, 14].
5. Finally, we turn to the bipartite Erdős–Rényi random graph in Section 1.6 and study the variational problems corresponding to the upper tails of regular subgraphs and largest eigenvalue. We extend the analysis of Lubetzky and Zhao [34] and determine the precise transition boundary between the symmetric and the symmetry breaking regimes.

We present a brief review of the relevant facts from graph limit theory [8, 9, 33] and detail our main results in the rest of this section.

1.1 Graph limit theory: a brief review

In this section, we collect some facts from the theory of dense graph limits [8, 9, 33] which will be relevant for the subsequent discussion. We refer the interested reader to [32] for an in-depth survey of this area. Define the function $\vartheta_n : [0,1] \to [n]$ as

$$\vartheta_n(x) = \begin{cases} 1 & 0 \le x \le \frac{1}{n}, \\ i & \frac{i-1}{n} < x \le \frac{i}{n}, \ \ 1 < i \le n. \end{cases} \tag{1}$$

Definition 1 (Graphon). *Let $\mathcal{W}$ be the space of all measurable functions $f : [0,1]^2 \mapsto [0,1]$ such that $f(x,y) = f(y,x)$ for all $(x,y) \in [0,1]^2$. We call $f \in \mathcal{W}$ a graphon.*

Definition 2 (Empirical Graphon). *Let G be a simple graph on $[n] = \{1, \dots, n\}$. The empirical graphon $f^G : [0,1]^2 \to [0,1]$ is defined as follows*

$$f^G(x,y) = \begin{cases} 1 & \text{if} \quad (\vartheta_n(x), \vartheta_n(y)) \text{ is an edge in } G, \\ 0 & \text{otherwise.} \end{cases}$$

Next, we recall the notions of the cut distance and cut metric.

Definition 3 (Cut Distance). *The cut distance between two graphons $f, g \in \mathcal{W}$ is defined as*

$$d_\square(f,g) = \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} \big(f(x,y) - g(x,y)\big)\, dxdy \right|,$$

where S, T are measurable subsets of $[0,1]$.

Definition 4 (Cut Metric). *For $f, g \in \mathcal{W}$, the cut metric is defined as*

$$\delta_\square(f,g) = \inf_{\phi \in \mathcal{M}} d_\square(f, g^\phi) = \inf_{\phi,\psi \in \mathcal{M}} d_\square(f^\psi, g^\phi) = \inf_{\psi \in \mathcal{M}} d_\square(f^\psi, g),$$

where $\mathcal{M}$ denotes the set of bijective, Lebesgue measure-preserving maps $\phi : [0,1] \mapsto [0,1]$.

We will establish our large deviation principle in the natural quotient space associated with $\delta_\square$. For $f, g \in \mathcal{W}$, write $f \sim g$ if $\delta_\square(f,g) = 0$ and $\tilde{f}$ for the equivalence class of all $f' \sim f$. For convenience of notation, we write $\tilde{f^G}$ rather than $\tilde{f^G}$ for the equivalence class containing the step function f^G.

[1] In statistical physics and chemistry, a reentrant phase transition describes a phenomenon where while walking on a straight line in parameter space, one leaves one phase, enters a new one, and then reenters into the first phase; we prefer to use this standard notion to the term 'double' phase transition used in [13, 14].

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Set $\widetilde{\mathcal{W}} = \mathcal{W}/\sim$. For $f, g \in \mathcal{W}$, we set

$$\delta_\square(\tilde{f}, g) = \delta_\square(f, g) \qquad \text{and} \qquad \delta_\square(\tilde{f}, \tilde{g}) = \delta_\square(f, g).$$

The above are well-defined, as for all f_1, f_2 such that $f_1 \sim f_2$, it holds that $\delta_\square(f_1, g) = \delta_\square(f_2, g)$. In [33], Lovász and Szegedy prove one of the central results in graph limit theory – the metric space $(\widetilde{\mathcal{W}}, \delta_\square)$ is compact. In particular, it is separable, which implies that the Borel σ-algebra over $\widetilde{\mathcal{W}}$ is generated by the open balls. We use $\tilde{\mathcal{A}}$ to denote this σ-algebra.

Throughout the paper we use $V(G)$ and $E(G)$ to denote the sets of vertices and edges of the graph G respectively.

Definition 5 (Homomorphism Density). *Let $H = (V(H), E(H))$ be a simple graph, where the vertices are labelled as $[v] = \{1, \ldots, v\}$, where $v = |V(H)|$. Define the homomorphism density of H in $f \in \mathcal{W}$ as*

$$t(H, f) = \int_{[0,1]^v} \prod_{(i,j)\in E(H)} f(x_i, x_j) dx_1 \ldots dx_v.$$

Since $t(H, f) = t(H, g)$ whenever $f \sim g$, $t(H, \cdot)$ is well-defined on $\widetilde{\mathcal{W}}$. With a slight abuse of notation, we use the same symbol for the function $t(H, \cdot): \widetilde{\mathcal{W}} \to [0, 1]: \tilde{f} \mapsto t(H, \tilde{f})$. As shown in [8, 33], this function is continuous for any finite graph H.

In this article, we study large deviations for block model random graphs. To this end, we denote as $\mathcal{B}^\gamma$ the set of block graphons where the width of the blocks are given by the values in the vector γ, which we assume to be rational. Let $\Delta_m = \{\gamma \in [0, 1]^m \cap \mathbb{Q}^m : \sum_{i=1}^m \gamma_i = 1\}$ denote the $(m-1)$-dimensional simplex restricted to rational points.

Definition 6. *Given $\gamma \in \Delta_m$, we define $I_1 = [0, \gamma_1]$ and*

$$I_j = \left(\sum_{k=0}^{j-1} \gamma_k, \sum_{k=0}^{j} \gamma_k \right] \qquad 1 < j \le m.$$

From these intervals, define the interval membership function

$$\vartheta_\gamma(x) = \sum_{j=1}^{m} j \mathbb{1}\{x \in I_j\}. \tag{2}$$

When γ is clear from context, we write $\vartheta(x)$. Let $\mathcal{B}^\gamma$ be the set of graphons $f \in \mathcal{W}$ of the form

$$f(x, y) = p_{\vartheta(x), \vartheta(y)},$$

where $p_{ij} = p_{ji} \in [0, 1]$. We call such a graphon an m-block graphon. When γ is clear from context, we write $f \in \mathcal{B}^\gamma$ as $f = (p_{ij})_{i,j\in[m]}$. When $\gamma = (1/m, \ldots 1/m)$, we say $f \in \mathcal{B}^\gamma$ is a uniform size (or simply uniform) m-block graphon. When $m = 2$, we write $\mathcal{B}^{(\gamma, 1-\gamma)}$ for $\gamma \in (0, 1) \cap \mathbb{Q}$ to denote the set of graphons with blocks of size γ and $1 - \gamma$.

Let

$$\mathcal{B}^{\gamma,*} = \left\{ f \in \mathcal{B}^\gamma : f \notin \mathcal{B}^\eta \text{ for all } \eta \in \Delta_{m-1} \right\}.$$

In other words, $\mathcal{B}^{\gamma,*}$ is the subset of graphons in $\mathcal{B}^\gamma$ that cannot be described by a smaller number of blocks. Let

$$\tilde{\mathcal{B}}^\gamma = \{\tilde{f} \in \widetilde{\mathcal{W}} : \delta_\square(\tilde{f}, g) = 0 \text{ for some } g \in \mathcal{B}^\gamma\}.$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Finally we define the sampling distribution for dense block model random graphs. We recall that $\tilde{\mathcal{A}}$ denotes the Borel σ-algebra over the metric space $(\widetilde{\mathcal{W}}, \delta_\square)$.

Definition 7 (Sampling from a block model). *Let $W_0 = (p_{ij})_{i,j\in[k]}$ be a uniform k-block graphon. Let $\mathbb{P}_{kn,W_0}$ denote the probability distribution over $\mathcal{W}$ obtained by sampling from W_0 as follows. Construct a simple graph G on kn vertices with unique labels in $[kn]$. Independently, add an edge between vertex i and vertex $j \neq i$ with probability $W_0(i/kn, j/kn) = p_{\lceil i/n\rceil,\lceil j/n\rceil}$. Return the empirical graphon f^G. Let $\tilde{\mathbb{P}}_{kn,W_0}$ denote the probability distribution induced on $\widetilde{\mathcal{W}}$ by the measure $\mathbb{P}_{kn,W_0}$, i.e., $\tilde{\mathbb{P}}_{kn,W_0}(\tilde{A}) = \mathbb{P}_{kn,W_0}(\tilde{f}^G \in \tilde{A})$ for all $A \in \tilde{\mathcal{A}}$.*

Remark 8. Note that any graphon with rational-length blocks is a uniform k-block graphon for some k, and thus the above scheme can be used to sample from such graphons.

1.2 A large deviation principle for block models

First we define the relative entropy function, both pointwise and for entire graphons. These definitions will be used to define the rate function for the LDP. Throughout we use the conventions $0 \log 0 = 0$ and $0 \log (0/0) = 0$.

Definition 9 (Relative entropy). *Define $I_{W_0} : \mathcal{W} \to \mathbb{R} \cup \{\infty\}$ as*

$$I_{W_0}(f) = \frac{1}{2} \int_{[0,1]^2} h_{W_0(x,y)}\big(f(x,y)\big)\, dxdy,$$

where $h_p(u)$ is the usual relative entropy,

$$h_p(u) = u \log \frac{u}{p} + (1-u) \log \frac{1-u}{1-p}.$$

While the previous definition applies to all graphons, we now specialise to block graphons. Given W_0, let $\Omega = \{(x,y) : W_0(x,y) \in (0,1)\}$. Define

$$\mathcal{W}_\Omega = \{f \in \mathcal{W} : \lambda\left(\{(x,y) \in \Omega^c : f(x,y) \neq W_0(x,y)\}\right) = 0\} \tag{3}$$

and

$$\widetilde{\mathcal{W}}_\Omega = \{\tilde{f} \in \widetilde{\mathcal{W}} : \delta_\square(f,g) = 0 \text{ for some } g \in \mathcal{W}_\Omega\}, \tag{4}$$

where $\lambda(\,\cdot\,)$ is the Lebesgue measure on $[0,1]^2$. In other words, $\mathcal{W}_\Omega$ is the set of graphons that agree with W_0 wherever W_0 takes value 0 or 1, except possibly on a measure-zero set. Note that $\mathbb{P}_{kn,W_0}$ and $\tilde{\mathbb{P}}_{kn,W_0}$ are supported on $\mathcal{W}_\Omega$ and $\widetilde{\mathcal{W}}_\Omega$ respectively. Lemma 31 states that $\widetilde{\mathcal{W}}_\Omega$ is closed (and hence compact), and Proposition 34 states that I_{W_0} is bounded on $\mathcal{W}_\Omega$, and infinite on $\mathcal{W} \setminus \mathcal{W}_\Omega$.

Note that Erdős–Rényi random graphs correspond to the constant base graphon $W_0 = p$ – this model satisfies an LDP with speed n^2, and rate function I_p [14]. However, in the general case, the function $I_{W_0}(\,\cdot\,)$ is not well-defined on the quotient space $\widetilde{\mathcal{W}}$, and thus cannot be the rate function for our LDP. We introduce our candidate rate function J_{W_0} on $\widetilde{\mathcal{W}}$ as follows. To this end, we will use the symbols B and S to denote the closed balls in $\mathcal{W}$ and $\widetilde{\mathcal{W}}$:

$$B(\tilde{f}, \varepsilon) = \{g \in \mathcal{W} : \delta_\square(\tilde{f}, g) \leq \varepsilon\}$$

$$S(\tilde{f}, \varepsilon) = \{\tilde{g} \in \widetilde{\mathcal{W}} : \delta_\square(\tilde{f}, \tilde{g}) \leq \varepsilon\}.$$

Definition 10 (Rate function). *The rate function is defined as*

$$J_{W_0}(\tilde{f}) = \begin{cases} \sup_{\eta>0} \inf_{h \in B(\tilde{f},\eta)} I_{W_0}(h) & \tilde{f} \in \widetilde{\mathcal{W}}_\Omega \\ \infty & \tilde{f} \notin \widetilde{\mathcal{W}}_\Omega. \end{cases}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

In Section 2.1 we prove that J_{W_0} is lower semi-continuous on $\widetilde{\mathcal{W}}$ (Lemma 32), and that it is bounded by some constant $C(W_0) < \infty$ on $\widetilde{\mathcal{W}}_\Omega$ (Proposition 34). Follow-up work by Grebik and Pikhurko [24] simplified the expression for the rate function, and established that one can instead work with

$$J_{W_0}(\tilde{f}) = \inf_{h:\delta_\square(h,\tilde{f})=0} I_{W_0}(h).$$

It would be interesting to see if this alternate expression can simplify our subsequent analysis. Additional follow-up work by Markering [35] showed that the same rate function applies when $\log(W_0), \log(\mathbf{1} - W_0) \in L^1([0,1]^2)$. Note that a block graphon with blocks containing zeros or ones does not satisfy such an integrability condition.

Theorem 11. *Let W_0 be a uniform k-block graphon. The sequence $\tilde{\mathbb{P}}_{kn,W_0}$ obeys a large deviation principle in the space $\widetilde{\mathcal{W}}$ (equipped with the cut metric $\delta_\square$) with speed $(kn)^2$ and rate function J_{W_0}. Explicitly,*

1. *For any open set $\tilde{U} \subseteq \widetilde{\mathcal{W}}$, $\liminf_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}(\tilde{U}) \geq -\inf_{\tilde{h}\in\tilde{U}} J_{W_0}(\tilde{h})$,*
2. *For any closed set $\tilde{F} \subseteq \widetilde{\mathcal{W}}$, $\limsup_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}(\tilde{F}) \leq -\inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h})$,*

(where we define the inf *over the empty set to be ∞.)*

Remark 12. Note that any graphon with rational-length blocks is a uniform k-block graphon for some k. Therefore, our result also describes large deviations events for any base graphon with rational-length blocks.

The proof of the LDP requires several new ideas, beyond those introduced in [14]. To explain the main additional difficulties, note that for the Erdős–Rényi random graph, W_0 is the constant graphon taking a value p, and thus the cut-distance $\delta_\square(W_0, f)$ to an arbitrary graphon $f \in \mathcal{W}$ is equal to the distance $d_\square(W_0, f)$. Somewhat related, the relative entropy I_{W_0} is a well-defined rate function on equivalence classes $\tilde{f} = \{g : \delta_\square(f,g) = 0\}$. Neither of these holds if W_0 is a block model with more than one block. To some extent, similar issues were faced in [20] in the context of large deviations for dense random graphs with given degrees. Our proof follows their general proof outline. However, the graphons W_0 considered in [20] are bounded away from zero and one, thus making the distinction between $\tilde{\mathcal{W}}$ and $\tilde{\mathcal{W}}_\Omega$ unnecessary. In contrast, the base graphon W_0 in our setting can have zero or one blocks – this creates many new analytic and probabilistic hurdles, and makes our analysis substantially more challenging.

1.3 LDP for graph parameters and the associated variational problem

In this section, we turn our attention to upper tail large deviations for continuous graph parameters.

Definition 13. *A graph parameter is a function $\tau : \widetilde{\mathcal{W}} \to \mathbb{R}$. We extend such a function τ to $\mathcal{W}$ by setting $\tau(f) = \tau(\tilde{f})$, where as before, $\tilde{f}$ is the equivalence class containing f. We further write $\tau(G) = \tau(f^G)$ for any graph G. We set $t^\tau_{\max}(\widetilde{\mathcal{W}}) = \max_{\tilde{f}\in\widetilde{\mathcal{W}}} \tau(\tilde{f})$ and $t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega) = \max_{\tilde{f}\in\widetilde{\mathcal{W}}_\Omega} \tau(\tilde{f})$. When τ and $\widetilde{\mathcal{W}}$ or $\widetilde{\mathcal{W}}_\Omega$ are clear from context, we simply write $t_{\max}$. We similarly define $t^\tau_{\min}(\widetilde{\mathcal{W}})$, $t^\tau_{\min}(\widetilde{\mathcal{W}}_\Omega)$, and $t_{\min}$ as the corresponding minimal graph parameter values.*

A graph parameter τ is continuous if it is continuous with respect to $\delta_\square$.

Remark 14. Note that the metric space $(\widetilde{\mathcal{W}}, \delta_\square)$ is compact, and thus every continuous graph parameter is, in fact, uniformly continuous with respect to $\delta_\square$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Note that by the compactness of $\widetilde{\mathcal{W}}$ and $\widetilde{\mathcal{W}}_\Omega$, the maxima in the above expressions are actually maxima and not suprema. Also, note that $t^\tau_{\max}(\widetilde{\mathcal{W}}) = \max_{f\in\mathcal{W}} \tau(f)$.

Definition 15. *Let τ be a continuous graph parameter. For $W_0 \in \mathcal{W}$ and $t \le t^\tau_{\max}(\widetilde{\mathcal{W}})$ we set*

$$\phi_\tau(W_0, t) = \min\{J_{W_0}(\tilde{f}) : \tilde{f} \in \widetilde{\mathcal{W}}, \tau(\tilde{f}) \ge t\}. \tag{5}$$

For $t > t^\tau_{\max}(\widetilde{\mathcal{W}})$, we set $\phi_\tau(W_0, t) = \infty$.

Note that continuity of τ and compactness of $(\widetilde{\mathcal{W}}, \delta_\square)$ imply that $\{\tilde{f} \in \widetilde{\mathcal{W}} : \tau(\tilde{f}) \ge t\}$ is compact. Since the lower semi-continuous function J_{W_0} (Lemma 32) attains its minimum on any compact set, it follows that $\phi_\tau(W_0, t)$ is well defined. Note also that $\{\tilde{f} \in \widetilde{\mathcal{W}} : \tau(\tilde{f}) \ge t\}$ has non-empty intersection with $\widetilde{\mathcal{W}}_\Omega$ if $t \le t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$, in which case $\phi_\tau(W_0, t)$ is bounded above by a constant depending on W_0, and that $\{\tilde{f} \in \widetilde{\mathcal{W}} : \tau(\tilde{f}) \ge t\} \cap \widetilde{\mathcal{W}}_\Omega = \emptyset$ and $\phi_\tau(W_0, t) = \infty$ if $t > t_{\max}$. So in particular, $\phi_\tau(W_0, t)$ is discontinuous at $t = t_{\max}$. In addition, $\phi_\tau(W_0, t) = 0$ if $t \le \tau(W_0)$, and $\phi_\tau(W_0, t) > 0$ on $(\tau(W_0), t_{\max}]$. To see this, observe that if $t \le \tau(W_0)$, then $\{\tilde{f} \in \widetilde{\mathcal{W}} : \tau(\tilde{f}) \ge t\}$ contains the equivalence class $\widetilde{W}_0$, and thus $\phi_\tau(W_0, t) = 0$. On the other hand, $J_{W_0}(\tilde{f}) = 0$ if and only if $\delta_\square(\tilde{f}, W_0) = 0$ (Proposition 40), and thus $\phi_\tau(W_0, t) > 0$ for $t \in (\tau(W_0), t_{\max}]$.

Our next result establishes ϕ_τ as the rate function for the upper tail large deviation of the graph parameter τ. Moreover, this result proves that conditioned on the rare event, the random graph *concentrates* on the minimisers of (5). This result is a direct adaptation of [34, Theorem 2.7] to general k-block graphons W_0.

Theorem 16. *Let W_0 be a uniform k-block graphon. Let τ be a continuous graph parameter, $t \le t^\tau_{\max}(\widetilde{\mathcal{W}})$, and let G_{kn} be the graph on kn vertices sampled from W_0 according to the probability distribution $\mathbb{P}_{kn,W_0}$. Recall $\phi_\tau(W_0, t)$ from (5), and assume that $\phi_\tau(W_0, \cdot)$ is continuous at t. Then*

$$\lim_{n\to\infty} \frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}\left(\tau(G_{kn}) \ge t\right) = -\phi_\tau(W_0, t).$$

Set $\tilde{F}^\star$ to be the set of minimisers of (5). Then $\tilde{F}^\star$ is a non-empty compact subset of $\widetilde{\mathcal{W}}$. If $t > t_{\max}$, then $\mathbb{P}_{kn,W_0}(\tau(G_{kn}) \ge t) = 0$, and if $t < t_{\max}$, then for n sufficiently large and each $\varepsilon > 0$, there exists $C = C(\tau, \varepsilon, W_0, t) > 0$ such that

$$\mathbb{P}_{kn,W_0}\left(\delta_\square(G_{kn}, \tilde{F}^\star) < \varepsilon \,\middle|\, \tau(G_{kn}) \ge t\right) \ge 1 - \exp(-Cn^2).$$

In particular, if $\tilde{F}^\star = \{\tilde{f}^\star\}$ for some $\tilde{f}^\star \in \widetilde{\mathcal{W}}$, then as $n \to \infty$, the conditional distribution of $\tilde{f}^{G_{kn}}$ given the event $\tau(G_{kn}) \ge t$ converges to the point mass at $\tilde{f}^\star$.

Remark 17. Note that, in general, Theorem 16 holds only at the continuity points of $\phi_\tau(W_0, t)$. Remark 60 explains that ϕ_τ has at most countably many points of discontinuity when τ is a continuous graph parameter. Moreover, we establish (see Lemma 58) that ϕ_τ is continuous on $\mathbb{R} \setminus \{t_{\max}\}$ if τ satisfies the 'sufficient increase property' (Definition 57). In turn, the proof of Lemma 55 establishes that homomorphism densities $t(H, \cdot)$ have the sufficient increase property for all finite graphs H and all step functions W_0, and the proof of Lemma 56 establishes that the operator norm has the sufficient increase property for a specific family of graphons W_0, namely those which generate bipartite Erdős–Rényi graphs.

So in particular, we know that for these graph parameters, the conclusions of Theorem 16 hold for all t, with the possible exception of $t = t_{\max}$. As we will see in Theorem 21, for the graph parameter $\tau = t(H, \cdot)$, where H is a finite d-regular graph, they also hold at $t = t_{\max}$, in spite of the fact that $\phi_\tau(W_0, t)$ is not continuous at this point.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Theorem 16 establishes that typical behaviour under the upper tail large deviation event is governed by the solutions of the variational problem (5). This directly motivates our subsequent investigations into the properties of this problem.

Definition 18 (Symmetric Regime). *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^{\gamma,*}$, and let τ be a continuous graph parameter. We say that $t \le t^{\tau}_{\max}(\widetilde{\mathcal{W}}_\Omega)$ is in the symmetric regime for W_0 and τ if all minimisers $\tilde{g}$ of*

$$\min_{\tilde{f} \in \widetilde{\mathcal{W}}} \{J_{W_0}(\tilde{f}) \colon \tau(\tilde{f}) \ge t\}, \tag{6}$$

satisfy $\tilde{g} \in \tilde{\mathcal{B}}^{\gamma}$. We call the symmetric solution unique if a unique element of $\widetilde{\mathcal{W}}_\Omega$ minimises (6).

Theorem 16 implies that in the symmetric regime, the conditional distribution of the random graph concentrates on a set of graphons with block structure agreeing with W_0. In addition, if there is a unique symmetric solution, the graph concentrates on the point mass corresponding to this solution. Our subsequent results explore the existence of a symmetric regime for specific graph parameters and establish uniqueness for a class of bipartite graphons W_0.

Next we specialise to the graph parameter defined by d-regular subgraph densities, i.e., to the graph parameter $\tau : \tilde{f} \mapsto t(H, \tilde{f})$ for a d regular graph H. In Section 1.4, we first show that for δ sufficiently small, $t = (1+\delta)t(H, W_0)$ is in the symmetric regime of W_0 and this graph parameter. Then we show that when t is sufficiently close to the maximum homomorphism density, t is also in the symmetric regime. In Section 1.5, we study examples of two-block graphons W_0 that have a non-symmetric regime – this exhibits that in these examples, these two symmetric regimes are separated by a non-symmetric regime, establishing a 'reentrant' phase transition phenomenon for large deviations in stochastic block models, analogous to the one established in [13, 14] for large deviations in Erdős–Rényi random graphs.

1.4 The existence of a symmetric regime for d-regular graphs

The next theorem establishes the existence of a symmetric regime for δ sufficiently small.

Theorem 19. *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^{\gamma,*}$. Let H be a d-regular graph, and let $\tau = t(H, \cdot)$. If $t(H, W_0) < t^{\tau}_{\max}(\widetilde{\mathcal{W}}_\Omega)$, then there exists $\delta > 0$ sufficiently small such that for all $t \in [t(H, W_0), (1+\delta)t(H, W_0))$, t is in the symmetric regime for W_0 and $t(H, \cdot)$. If h is a minimiser, then*

$$J_{W_0}(\tilde{h}) = I_{W_0}(h) = \min\{I_{W_0}(g) : g \in \mathcal{B}^{\gamma}, t(H, g) \ge (1+\delta)t(H, W_0)\}.$$

Next, we explore the variational problem near the maximum homomorphism density, and establish the existence of a symmetric regime in this setting.

Theorem 20. *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^{\gamma,*}$. Let H be a d-regular graph, and let $\tau = t(H, \cdot)$. If $t(H, W_0) < t^{\tau}_{\max}(\widetilde{\mathcal{W}}_\Omega)$, then exists $\eta > 0$ such that for all $t \in ((1-\eta)t_{\max}, t_{\max}]$, t is in the symmetric regime for W_0 and $t(H, \cdot)$. In addition, if h is a minimiser, then*

$$J_{W_0}(\tilde{h}) = I_{W_0}(h) = \min\{I_{W_0}(g) : g \in \mathcal{B}^{\gamma}, t(H, g) \ge (1+\delta)t(H, W_0)\}.$$

Theorems 19 and 20 establish the existence of a symmetric regime for the homomorphism density of regular graphs. This is challenging due to the form of the rate function $J_{W_0}(\cdot)$, and is one of the main technical contributions of this paper. To this end, our first contribution is to establish that

$$\min\{J_{W_0}(\tilde{f}) : \tilde{f} \in \mathcal{W}, \tau(\tilde{f}) \ge t\} = \inf\{I_{W_0}(f) : f \in \mathcal{W}, \tau(f) \ge t\} \tag{7}$$

under mild assumptions on the graph parameter τ, which are satisfied for homomorphism densities and the operator norm (Lemma 64). This insight facilitates our subsequent analysis, and

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

allows us to work with the relatively entropy functional I_{W_0}, instead of the complicated rate function J_{W_0}.

Even with this simplification, our proof is quite involved. To exhibit the existence of a symmetric phase, we will establish that for certain ranges of t (depending on W_0), any minimiser of (6) is in $\tilde{\mathcal{B}}^\gamma$. To this end, we will establish that if $\tilde{f}$ is a minimiser of (6), there exists a sequence of block constant graphons $\{f_n : n \geq 1\} \subseteq \mathcal{B}^\gamma$ such that $\delta_\square(\tilde{f}, f_n) \to 0$. This will imply that $\tilde{f} \in \tilde{\mathcal{B}}^\gamma$, as $\tilde{\mathcal{B}}^\gamma$ is closed in $(\widetilde{\mathcal{W}}, \delta_\square)$. We refer the reader to Section 4 for details on the construction of this sequence $\{f_n : n \geq 1\}$.

Combined with Theorem 16, these two theorems characterise the 'typical' structure of the graph, conditioned on an upper tail large deviation event for the graph parameter $\tau = t(H, \cdot)$ in the vicinity of the endpoints of $[t(H, W_0), t_{\max}]$. However, as stated, Theorem 16 applies only for $t \in [t(H, W_0), t_{\max})$. It is natural to wonder what happens when $t = t_{\max}$. This is the content of the next theorem.

To state it, we recall the notation $W_0 = (p_{ij})_{i,j\in[m]}$ for a graphon with blocks $I_i \times I_j$, $i, j \in [m]$, and define a block $I_a \times I_b$ to be relevant if $p_{ab} > 0$ and $t(H, W_0)$ strictly decreases if p_{ab} is lowered. Note that by definition, all blocks where $p_{ab} = 0$ are not relevant, while the blocks where $p_{ab} = 1$ may or may not be relevant. Note that the maximum homomorphism density of a fixed subgraph H in a random graph drawn from $\mathbb{P}_{kn,W_0}$ is $\max_{f\in\mathcal{W}_\Omega} t(H, f)$.

Theorem 21. *Let W_0 be a uniform k-block graphon. Let H be a finite d-regular graph, let $\tau = t(H, \cdot)$, and let $f_{\max}$ be the step function which is equal to 1 on all relevant blocks, and equal to W_0 on all irrelevant blocks. Then $\tilde{f}_{\max}$ is the unique minimiser of (6) at $t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$. Moreover, for any $\varepsilon > 0$, there exists a constant $C > 0$ such that*

$$\mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}}, \tilde{f}_{\max}) < \varepsilon \,\middle|\, t(H, G_{kn}) \geq t_{\max}\right) \geq 1 - \exp(-Cn^2),$$

implying that as $n \to \infty$, the conditional distribution of $\tilde{f}^{G_{kn}}$ given the event $\tau(G_{kn}) \geq t_{\max}$ converges to the point mass at $\tilde{f}_{\max}$.

Remark 22. If $\tau(\tilde{f}) = t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$ has a unique solution $\tilde{f}_{\max}$, it is immediately clear that

$$\mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}}, \tilde{f}_{\max}) = 0 \,\middle|\, \tau(\tilde{f}^{G_{kn}}) \geq t_{\max}\right) = 1,$$

implying that the conditional distribution of $\tilde{f}^{G_{kn}}$ given the event $\{\tau(\tilde{f}^{G_{kn}}) \geq t_{\max}\}$ is the point mass at $\tilde{f}_{\max}$. The equation $\tau(\tilde{f}) = t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$ has a unique solution, for example, if $\tau = t(H, \cdot)$, where H is a finite d-regular graph, and all blocks of W_0 that are subsets of Ω are relevant. In this case $f_{\max} = \mathbf{1}_{W_0>0}$. This also holds if W_0 is a bipartite graphon with two blocks and $\tau(\tilde{f}) = \|f\|_{\mathrm{op}}$ (see Theorem 28), again with $f_{\max} = \mathbf{1}_{W_0>0}$. See the Appendix for additional details.

The proof of Theorem 21 is relatively straightforward given the proofs of Theorems 16, 19 and 20, and is deferred to the Appendix.

1.5 A non-symmetric regime in special cases

Next, we establish the existence of a non-symmetric regime in some specific families of two-block graphons. Let

$$f^\gamma_{p,q,r}(x, y) = \begin{cases} p & \text{if } (x, y) \in [0, \gamma]^2 \\ r & \text{if } (x, y) \in (\gamma, 1]^2 \\ q & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

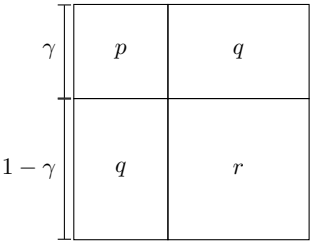


Figure 1. Illustration of the graphon $f^{\gamma}_{p,q,r}$.

We show the existence of a non-symmetric regime for base graphons of the form $f^{\gamma}_{0,p,p}, f^{\gamma}_{1,p,p}, f^{\gamma}_{1,p,0}$ when p is sufficiently small. The first model corresponds to an Erdős–Rényi random graph with a planted independent set, while the second example covers Erdős–Rényi graphs with a planted clique. Finally, the third graphon leads to a bipartite Erdős–Rényi random graph with a planted clique in one of the partitions.

Theorem 23. *Let $\gamma \in (0,1) \cap \mathbb{Q}$ and H be a d-regular graph. Assume that*

1. $0 < t < t(H, f^{\gamma}_{0,0,1})$ *and denote* $W_p = f^{\gamma}_{0,p,p}$, *or*
2. $t(H, f^{\gamma}_{1,0,0}) < t < 1$ *and denote* $W_p = f^{\gamma}_{1,p,p}$, *or*
3. $t(H, f^{\gamma}_{1,0,0}) < t < t(H, f^{\gamma}_{1,1,0})$ *and denote* $W_p = f^{\gamma}_{1,p,0}$.

Separately, under each of these assumptions, there exists $p_0 > 0$ such that if $p < p_0$,

$$\min\{J_{W_p}(\tilde{g}) : t(H,\tilde{g}) \geq t\} < \min\{J_{W_p}(\tilde{g}) : \tilde{g} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)}, t(H,\tilde{g}) \geq t\}.$$

These statements imply that for p small enough, the minimiser of the variational problem (6) is non-symmetric.

In Proposition 43, we show that $\{\tilde{g} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)} : t(H,\tilde{g}) \geq t\}$ is compact, which justifies the minimum on the right hand side in Theorem 23. Note that in the second and third case, we cover all $t \in (t_{\min}, t_{\max})$, where $t_{\min}, t_{\max}$ are the minimal and maximal values of $t(H, W_p)$ ranging over $p \in (0,1)$. On the other hand, in the first case, we do not consider the full range, since we exclude $t \in [t(H, f^{\gamma}_{0,0,1}), t(H, f^{\gamma}_{0,1,1}))$.

To establish this result, we recall that I_{W_0} is significantly more tractable than the rate function J_{W_0}. Our first step (see Lemma 85) is to show that if W_0 is a graphon of the form $f^{\gamma}_{0,p,p}$ or $f^{\gamma}_{1,p,p}$, then

$$\min\{J_{W_0}(\tilde{f}) : \tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)},\ \tau(f) \geq t\} = \min\{I_{W_0}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)},\ \tau(f) \geq t\}. \quad (8)$$

A similar simplification occurs for W_0 of the form $f^{\gamma}_{z_1,p,z_2}$, where $z_1, z_2 \in \{0,1\}$.

Next we show that for graphons W_p of the form $W_p = f^{\gamma}_{0,p,p}$ and $W_p = f^{\gamma}_{1,p,p}$, there exists $p_0 > 0$ such that if $p < p_0$,

$$\inf\{I_{W_p}(f) : t(H,f) \geq t\} < \min\{I_{W_p}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)},\ t(H,f) \geq t\} \quad (9)$$

for some range of t (Lemma 86). We establish this by constructing explicit graphons with lower entropy than that of all graphons in $\mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}$. A similar simplification occurs for $W_p = f^{\gamma}_{1,p,0}$. Together with (7) and (8), (9) implies the desired conclusion.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

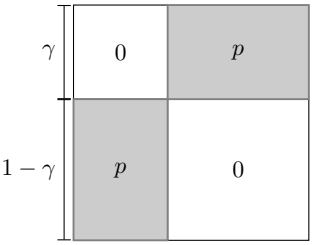


Figure 2. Illustration of the graphon f_p^γ.

1.6 Bipartite Erdős–Rényi graphs: symmetry vs. symmetry breaking

Lubetzky and Zhao [34] characterise the symmetric regimes for d-regular subgraph counts and the largest eigenvalue in the Erdős–Rényi model. We extend these results to bipartite Erdős–Rényi random graphs. Let f_p^γ denote the graphon $f_{0,p,0}^\gamma$, i.e.,

$$f_p^\gamma(x,y) = \begin{cases} 0 & (x,y) \in [0,\gamma]^2 \cup (\gamma,1]^2 \\ p & \text{otherwise.} \end{cases}$$

For W_0 of the form f_p^γ, the following theorem completely characterises the symmetric and non-symmetric regimes for $t(H,\cdot)$, where H is a regular graph.

Theorem 24. *Fix $0 < p < 1$, $\gamma \in (0,1) \cap \mathbb{Q}$, and H a d-regular graph with $d \geq 1$. Let $W_0 = f_p^\gamma$. Let $r \in [p,1]$ and define $t_r^\gamma = t(H, f_r^\gamma)$.*

1. *If $(r^d, h_p(r))$ lies on the convex minorant of $x \mapsto h_p(x^{1/d})$, then t_r^γ is in the symmetric regime for W_0 and $t(H,\cdot)$. Moreover, $\tilde{f}_r^\gamma$ is the unique symmetric solution.*
2. *If $(r^d, h_p(r))$ does not lie on the convex minorant of $x \mapsto h_p(x^{1/d})$, then t_r^γ is not in the symmetric regime of W_0 and $t(H,\cdot)$.*

Remark 25. The symmetric regime for subgraph counts in Erdős–Rényi graphs [34] takes a similar form, with t_r^γ replaced by $t(H,r)$, where r denotes the constant graphon with value r.

Finally, we characterise the symmetric regime for the largest eigenvalue. Similar to Erdős–Rényi graphs, the boundary for the symmetric regime for the largest eigenvalue coincides with that of the density of two-regular graphs.

Definition 26. *For a graphon $f \in \mathcal{W}$, define the Hilbert–Schmidt kernel operator T_f on $L^2([0,1])$ by*

$$(T_f u)(x) = \int_0^1 f(x,y)u(y)dy$$

for any $u \in L^2([0,1])$. The operator norm is given by

$$\|f\|_{op} = \min\{c \geq 0 : \|T_f u\|_2 \leq c\|u\|_2 \text{ for all } u \in L^2([0,1])\}.$$

Lemma 27 ([34], Lemma 3.6). *The function $\|\cdot\|_{op}$ is a continuous extension of the normalised graph spectral norm, i.e., $\lambda_1(G)/n$ for a graph G on n vertices, to $(\widetilde{\mathcal{W}}, \delta)$. Moreover, $\|\cdot\|_{op}$ is a continuous graph parameter.*

Theorem 28. *Fix $0 < p < 1$, $\gamma \in (0,1) \cap \mathbb{Q}$, and let $W_0 = f_p^\gamma$. Let $r \in [p,1]$ and define $t_r^\gamma = \|f_r^\gamma\|_{op}$.*

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

1. *If* $(r^2, h_p(r))$ *lies on the convex minorant of* $x \mapsto h_p(x^{1/2})$*, then* t_r^γ *is in the symmetric regime for* W_0 *and* $\|\cdot\|_{op}$*. Moreover,* $\tilde{f}_r^\gamma$ *is the unique symmetric solution.*
2. *If* $(r^2, h_p(r))$ *does not lie on the convex minorant of* $x \mapsto h_p(x^{1/2})$*, then* t_r^γ *is not in the symmetric regime for* W_0 *and* $\|\cdot\|_{op}$*.*

Remark 29. It is not hard to see that for $\tau = t(H, \cdot)$ and $\tau = \|\cdot\|_{\mathrm{op}}$, the function $r \mapsto \tau(f_r^\gamma)$ is a continuous and non-decreasing function on $[p, 1]$ and that $\tau(f_1^\gamma) = t_{\max}^\tau(\widetilde{\mathcal{W}}_\Omega)$. Thus Theorems 24 and 28 cover the full range $[\tau(W_0), t_{\max}]$.

To establish Theorems 24 and 28, we follow the general approach introduced in [34]. Lemma 85 implies that (8) holds for $\tau(g) = t(H, g)$ where H is a d-regular graph or $\tau(g) = \|g\|_{\mathrm{op}}$, meaning that we can again reason about symmetry through the function I_{W_0} rather than J_{W_0}. For $r \in (0, 1]$, let f_r^γ be the bipartite graphon with value r, and $t_r^\gamma = t(H, f_r^\gamma)$ be the corresponding homomorphism density of a d-regular graph H. We apply a generalised Hölder inequality to show that whenever $f \in \mathcal{W}_\Omega$ satisfies $t(H, f) \geq t(H, f_r^\gamma)$, it holds that $\|f\|_d^d \geq 2\gamma(1-\gamma)r^d$ (Lemma 93). Finally, we show that if $(r^d, h_p(r))$ lies on the convex minorant of $x \mapsto h_p(x^{1/d})$ and $\|f\|_d^d \geq 2\gamma(1-\gamma)r^d$, then $I_{W_0}(f) \geq I_{W_0}(f_r^\gamma)$, with equality occurring if and only if $f = f_r^\gamma$ (Lemma 92). To establish the non-symmetric regime, we show that whenever $(r^d, h_p(r))$ is not on the convex minorant, we can construct a graphon g with $t(H, g) > t(H, f_r^\gamma)$ and $I_{W_0}(g) < I_{W_0}(f_r^\gamma)$ (Lemma 91). This construction is more complicated than the one in [34], due to the bipartite nature of the underlying graph (see Figure 8 for the construction). The proof for the spectral norm $\tau(g) = \|g\|_{\mathrm{op}}$ follows using similar arguments.

1.7 History and related work

The upper tail large deviation problem for subgraphs of $G(n, p)$ has attracted considerable attention in Probability and Combinatorics. By applying the theory of graph limits, Chatterjee and Varadhan found the precise constant in the large deviation probability in the dense case [14]. This approach does not work in the sparse regime where $p \to 0$, as graphon theory only applies to dense graphs.

The challenge of deriving an LDP for sparse graphs has attracted considerable attention in recent years. In the sparse regime, even determining the right order of this probability on the exponential scale proved to be considerably challenging. Following partial advances [27–30, 37], this was finally resolved for $H = K_3$ in [10, 16]. Subsequently, [17] identified the right order of this probability for $H = K_r$, $r \geq 4$, and formulated a conjecture regarding the correct order for general subgraphs. See [36] for a recent counterexample to this general conjecture. Recently, the development of general theory [2, 12, 22] and problem-specific ideas [1, 3, 15, 25] have contributed to rapid progress on large deviations in the sparse setting. These results relate the large deviation probability to an entropic variational problem. In turn, some of these variational problems have also been solved [4–6], leading to deep insights regarding the structure of the random graph, conditioned on the rare event.

We emphasise that these remarkable results are mostly applicable for sparse random graphs or hypergraphs [31], and do not shed any direct insight on the problem considered in this paper. Instead, our work is the first step towards a full generalisation of the work of [14] and [34] to block models. As [14] did for Erdős–Rényi graphs, we establish an LDP for block models and demonstrate the existence of a reentrant phase transition for the upper tail of d-regular subgraph counts. While we exhibit a reentrant phase transition for a limited class of block models, we show the existence of a symmetric regime for arbitrary block models. Our methods are inspired by the work of [34], which completely characterises the symmetric and non-symmetric regimes for

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Erdős–Rényi graphs. Moreover, analogous to [34], we fully characterise the symmetric and non-symmetric regimes for bipartite Erdős–Rényi graphs. As discussed in the introduction, our work fits into the broader theme of large deviations for dense random graphs with inhomogeneities or constraints, and provides the first rigorous analysis of the large deviations problem for dense block models.

Following the posting of this paper to arXiv, there has been some follow-up work. Grebik and Pikhurko [24] simplified our function given by Definition 10 showing that taking the lower-semicontinuous regularisation is not necessary. Building on our work, Grebik and Pikhurko [24] derived an LDP for graphs sampled from step graphons whose blocks are not necessarily of rational length. Markering [35] derived a large deviation principle for inhomogeneous Erdős–Rényi random graphs, showing that the rate function takes a simple form under certain integrability assumptions.

Outline: The rest of the paper is structured as follows. We establish our main LDP results, Theorem 11 and Theorem 16, in Section 2. In Section 3, we derive some analytic properties of ϕ_τ which are crucial in the analysis of the variational problem. Section 4 establishes the existence of a symmetric regime in the upper tail, while Section 5 establishes the existence of a non-symmetric regime in specific examples. Finally, we characterise the symmetric regime in Erdős–Rényi bipartite models in Section 6. We finish with some open problems in Section 7.

2. Large deviation principle

In this section we establish the LDP. Since the space $(\widetilde{\mathcal{W}}, \delta_\square)$ is compact, it will be enough to prove the bounds in Theorem 11 for balls in the metric $\delta_\square$; the precise statement is given in the following lemma.

Lemma 30. *Since the space $(\widetilde{\mathcal{W}}, \delta_\square)$ is compact, the bounds in Theorem 11 are equivalent to*

1. *For all $\varepsilon > 0$ and $\tilde{h} \in \widetilde{\mathcal{W}}$, $\liminf_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}\left(S(\tilde{h},\varepsilon)\right) \geq -J_{W_0}(\tilde{h})$;*
2. *For all $\tilde{g} \in \widetilde{\mathcal{W}}$, $\lim_{\alpha\to 0} \limsup_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}\left(S(\tilde{g},\alpha)\right) \leq -J_{W_0}(\tilde{g})$.*

The proof is standard (see e.g., [19, Theorems 4.1.11, 4.1.18]), and is thus omitted.

In Section 2.1, we begin by establishing several useful facts about the rate function and the space $\mathcal{W}_\Omega$. We establish the LDP lower and upper bounds in Sections 2.2 and 2.3 respectively. Finally, in Section 2.4, we prove Theorem 16, which establishes upper tail large deviations for continuous graph parameters.

2.1 Preliminaries

In this section, we establish a variety of useful analytical properties. Of particular interest are Lemma 31 and Lemma 32, which state that $\widetilde{\mathcal{W}}_\Omega$ is closed and that J_{W_0} is lower semi-continuous on $(\widetilde{\mathcal{W}}, \delta_\square)$.

Lemma 31. *For any $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and $W_0 \in \mathcal{B}^\gamma$, the set $\widetilde{\mathcal{W}}_\Omega$ is closed in $\widetilde{\mathcal{W}}$ with respect to the cut metric topology $(\widetilde{\mathcal{W}}, \delta_\square)$.*

Lemma 32. *For any $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and $W_0 \in \mathcal{B}^\gamma$, the function $J_{W_0}(\,\cdot\,)$ is lower semi-continuous on $(\widetilde{\mathcal{W}}, \delta_\square)$.*

We start by stating some elementary properties of the relative entropy $h_p(\,\cdot\,)$.

Lemma 33. *Let $\beta \in (0, 1/2]$ and let $G_p(a,q) = aq - \log(pe^a + 1 - p)$. Then the following holds*

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

(i) *For all* $p \in [\beta, 1-\beta]$, $\|h_p\|_\infty \le \log(2/\beta)$.

(ii) *The family of functions* $(h_p)_{p\in[\beta,1-\beta]}$ *is equicontinuous on* $[0,1]$.

(iii) *For all* $p \notin \{0,1\}$, $h_p(q) = \sup_{a\in\mathbb{R}} G_p(a,q)$

(iv) *For* $q \notin \{0,1\}$, *the* sup *in (iii) is achieved by* $a = \log\left(\frac{q}{1-q}\frac{1-p}{p}\right)$.

Proof. (i) Follows by observing that $|x\log x + (1-x)\log(1-x)| \le \log 2$ and $|x\log p| + |(1-x)\log p| \le x|\log\beta| + (1-x)|\log\beta| = -\log\beta$.

(ii) Follows from uniform continuity of the function $x \mapsto x\log x + (1-x)\log(1-x)$.

(iii) and (iv) are elementary exercises left to the reader. □

The function I_{W_0} also comes up naturally in [20] – however, in [20] it is assumed that the base graphon W_0 is bounded away from zero and one, and thus the function I_{W_0} is necessarily finite. This is not the case in our context. We use $Im(W_0)$ to denote the image of W_0 in $[0,1]$.

Proposition 34. *Let* $W_0 \in \mathcal{W}$. *If* $f \notin \mathcal{W}_\Omega$, *then* $I_{W_0}(f) = \infty$. *If* $f \in \mathcal{W}_\Omega$ *and* W_0 *obeys the assumption*

$$\beta = \inf\{w > 0 : w \in Im(W_0) \text{ or } 1-w \in Im(W_0)\} > 0, \tag{10}$$

then $I_{W_0}(f) \le \frac{1}{2}\log(2/\beta)$.

Proof. Since $h_0(x) = \infty$ for $x \neq 0$ and $h_1(x) = \infty$ for $x \neq 1$, it follows that $I_{W_0}(f) = \infty$ when $f \notin \mathcal{W}_\Omega$. To bound $I_{W_0}(f)$ for $f \in \mathcal{W}_\Omega$, observe that

$$I_{W_0}(f) = \frac{1}{2}\int_\Omega h_{W_0(x,y)}\big(f(x,y)\big)\,dxdy, \tag{11}$$

since f and W_0 agree on Ω^c. The proof is completed by invoking Lemma 33 (i). □

Proposition 35. *Let* $\varepsilon > 0$ *and assume that* $W_0 \in \mathcal{W}$ *obeys the condition (10). Then there exists* $\eta > 0$ *such that if* $f, g \in \mathcal{W}_\Omega$ *and* $\|f-g\|_\infty \le \eta$, *then* $|I_{W_0}(f) - I_{W_0}(g)| \le \varepsilon$.

Proof. Fix $\varepsilon > 0$. By Lemma 33 (ii) there exists an $\eta > 0$ such that $|h_p(u) - h_p(v)| \le \varepsilon$ whenever $|u-v| \le \eta$ and $p \in [\beta, 1-\beta]$. Inserted into (11), this completes the proof. □

We derive a variational representation for I_{W_0} using convex duality.

Proposition 36. *Let* $W_0 \in \mathcal{W}$ *and let* S *be the set of all symmetric functions in* $L_2([0,1]^2)$. *For* $a \in S$ *and* $f \in \mathcal{W}$, *define*

$$K_{W_0}(f,a) = \int_{[0,1]^2}\left[a(x,y)f(x,y) - \log\left(W_0(x,y)e^{a(x,y)} + 1 - W_0(x,y)\right)\right]dxdy. \tag{12}$$

Then $I_{W_0}(f) = \frac{1}{2}\sup_{a\in S} K_{W_0}(f,a)$.

Proof. First, we consider the case $f \notin \mathcal{W}_\Omega$ (in which case $I_{W_0}(f) = \infty$ by Proposition 34). Then there exists $\Gamma \subseteq [0,1]^2$ with positive measure such that $W_0(x,y) \in \{0,1\}$ and $W_0(x,y) \neq f(x,y)$ for $(x,y) \in \Gamma$. Choosing

$$a_M(x,y) = \begin{cases} 0 & (x,y) \notin \Gamma \\ M & W_0(x,y) = 0 \\ -M & W_0(x,y) = 1. \end{cases}$$

and taking $M \to \infty$, we see that $\sup_{a\in S} K_{W_0}(f,a) = \infty$ in this case.

Next we consider the case $f \in \mathcal{W}_\Omega$. Recalling the definition of G_p from Lemma 33 and noting that the integrand in (12) is zero if $W_0(x,y) = f(x,y) \in \{0,1\}$, we then have $K_{W_0}(f,a) = \int_\Omega G_{W_0(x,y)}(a(x,y), f(x,y))dxdy$. Combined with (11) and Lemma 33 (iii), this

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

shows that

$$I_{W_0}(f) = \frac{1}{2}\int_\Omega h_{W_0(x,y)}(f(x,y))dxdy \geq \frac{1}{2}\sup_{a\in S}\int_\Omega G_{W_0(x,y)}(a(x,y),f(x,y))dxdy = \frac{1}{2}\sup_{a\in S} K_{W_0}(f,a).$$

To prove equality, we may w.l.o.g. assume that the right hand side is finite. We may further restrict the integrals on both sides to the subset $\tilde{\Omega}\subset\Omega$ where $f(x,y)\notin\{0,1\}$, since the contributions of both sides to the complement can easily be seen to be equal. Finally, on $\tilde{\Omega}$, we may use Lemma 33 (iv) to conclude that

$$\int_{\tilde{\Omega}} h_{W_0(x,y)}(f(x,y))dxdy = \int_{\tilde{\Omega}} G_{W_0(x,y)}(a_0(x,y),f(x,y))dxdy$$

where $a_0(x,y) = \log\left(\frac{f(x,y)}{1-f(x,y)}\frac{1-W_0(x,y)}{W_0(x,y)}\right)$. If $a_0\in L^2(\tilde{\Omega})$, the right hand side is bounded by the sup over all $a\in S$, giving the desired upper bound. If it is not, we replace $\tilde{\Omega}$ on both sides by its intersection with the set of points for which $|a_0(x,y)|\leq M$ before bounding the right hand side by a sup over all square integrable a. The proof is concluded by using the monotone convergence theorem. □

The following propositions will be used to establish the lower semi-continuity of I_{W_0}.

Proposition 37. *Let $m\in\mathbb{Z}^+$ and $\gamma\in\Delta_m$. For $W_0\in\mathcal{B}^\gamma$, the set $\mathcal{W}_\Omega$ is a closed subset of $\mathcal{W}$ with respect to the topology $(\mathcal{W}, d_\square)$.*

Proof. Let $\{f_n : n\geq 1\}\subseteq\mathcal{W}_\Omega$ be a convergent sequence of graphons satisfying $d_\square(f_n,f)\to 0$ for some $f\in\mathcal{W}$. Then on each block $I\times J$ of W_0 that takes value 0 or 1, $f_n = W_0$ for all $n\geq 1$. Since $|\int_{I\times J} f_n - f| \leq d_\square(f_n,f)$ and $d_\square(f_n,f)\to 0$, it follows that $f = W_0$ on $I\times J$. Thus $f\in\mathcal{W}_\Omega$. □

Proposition 38. *Let (S,d) be either $(\mathcal{W},d_\square)$ or $(\tilde{\mathcal{W}},\delta_\square)$, and let $F\subseteq S$ be a closed subset of S. Let $f: F\to\mathbb{R}$ be a lower semi-continuous function on (F,d). The extension $f^*: S\to\mathbb{R}\cup\{\infty\}$ where*

$$f^*(x) = \begin{cases} f(x) & x\in F \\ \infty & x\in S\setminus F \end{cases}$$

is lower semi-continuous on (S,d).

Proof. We show that f^* is lower semi-continuous on S by demonstrating that for all $\alpha\in\mathbb{R}$, the set $\{x\in S | f^*(x)>\alpha\}$ is open. Observe

$$\{x\in S | f^*(x)>\alpha\} = \{x\in F | f(x)>\alpha\}\cup(S\setminus F)\,.$$

By lower semi-continuity of f on F, $A = \{x\in F | f(x)>\alpha\}$ is open in F, and so A^c is relatively closed with respect to F and therefore closed in S (since F is closed). It follows that

$$\{x\in S | f^*(x)>\alpha\}^c = F\cap A^c$$

is closed, and so we conclude that $\{x\in S | f^*(x)>\alpha\}$ is open. □

Lemma 39. *Let $m\in\mathbb{Z}^+$ and $\gamma\in\Delta_m$. For $W_0\in\mathcal{B}^\gamma$, the function $I_{W_0}(\,\cdot\,)$ is lower semi-continuous on $(\mathcal{W},d_\square)$.*

Proof. First, note that by Propositions 37 and 38, and the observation that $I_{W_0}(f)=\infty$ for all $f\in\mathcal{W}\setminus\mathcal{W}_\Omega$, it is enough to establish that I_{W_0} is lower semi-continuous on $(\mathcal{W}_\Omega,d_\square)$. Second, by Proposition 36, I_{W_0} can be written as supremum over the functions $K_{W_0}(\cdot,a)$, so it will be enough to show that for all $a\in S$, the function $K_{W_0}(\cdot,a)$ is continuous on $(\mathcal{W}_\Omega,d_\square)$.

Consider two functions $f,g\in\mathcal{W}_\Omega$, and observe that every $a\in S$ can be approximated by step functions in L^2. Given ε and a we can therefore find $k<\infty$ and a k-step function a_k such that

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$\|a - a_k\|_2 \le \frac{\varepsilon}{2}$. As a consequence

$$\left|K_{W_0}(f,a) - K_{W_0}(g,a)\right| = \left|\int a(f-g)\right| \le \|a(f-g)\|_1 \le \|a_k(f-g)\|_1 + \frac{\varepsilon}{2} \le \|a_k\|_\infty k^2 d_\square(f,g) + \frac{\varepsilon}{2}.$$

This is smaller than ε if $d_\square(f,g) < \varepsilon/(\|a_k\|_\infty k^2)$, which proves that $K_{W_0}(\cdot, a)$ is continuous on $(\mathcal{W}_\Omega, d_\square)$, as required. □

The following proposition will be used to prove that for any continuous graph parameter τ and graphon W_0, the function $\phi_\tau(W_0, \cdot)$ is strictly positive on $(\tau(W_0), t_{\max}]$.

Proposition 40. *For $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$, and $\tilde{f} \in \widetilde{\mathcal{W}}$, it holds that*

$$J_{W_0}(\tilde{f}) = 0 \text{ if and only if } \delta_\square(\tilde{f}, W_0) = 0.$$

Proof. Noting that our definition of J_{W_0} agrees with the definition of J_{W_0} given in [20] whenever $f \in \mathcal{W}_\Omega$, and that $\delta_\square(f, W_0) > 0$ if $f \notin \mathcal{W}_\Omega$, the proposition follows from the analogous statement in [20, Lemma 2.2]. □

Our next result establishes that $\widetilde{\mathcal{W}}_\Omega$ is closed in the cut metric. For a partition $\mathcal{P}$ of $[0,1]$, we define $W_\mathcal{P}$ as the step function graphon that is obtained by averaging over all blocks induced by the partition classes. Setting $\Gamma(x) \subseteq [0,1]$ to be the partition class in $\mathcal{P}$ that contains x, we obtain

$$W_\mathcal{P}(x,y) = \frac{1}{|\Gamma(x)| \cdot |\Gamma(y)|} \int_{\Gamma(x)\times\Gamma(y)} W(u,v)dudv.$$

We call $\mathcal{P}$ an equipartition if all classes have the same measure, and use $|\mathcal{P}|$ to denote the number of classes in $\mathcal{P}$. Note that up to sets of measure zero, there is just one equipartition of $[0,1]$ into n intervals; for definiteness and consistency with our previous conventions, we use the partition $([0,1/n], (1/n, 2/n], \ldots, ((n-1)/n, 1])$.

Lemma 41 (Corollary 3.4 of [8]). *Let $f \in \mathcal{W}$ and s be a positive integer. For every equipartition $\mathcal{Q}$ of $[0,1]$, there is an equipartition $\mathcal{P}$ with $s|\mathcal{Q}|$ classes such that $\mathcal{P}$ refines $\mathcal{Q}$ and*

$$d_\square(f, f_\mathcal{P}) \le \sqrt{\frac{20}{\log_2 s}}.$$

The next lemma follows from Lemma 41.

Lemma 42. *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and $W_0 \in \mathcal{B}^\gamma$. Then there exists a sequence of refining partitions $\mathcal{P}_k$ of $[0,1]$ into equal length intervals such that for all $f \in \mathcal{W}_\Omega$ there exists a sequence of step functions $f_k \in \mathcal{W}_\Omega$ with steps in $\mathcal{P}_k$ such that (i) $(f_{k+1})_{\mathcal{P}_k} = f_k$ and (ii) $\delta_\square(f, f_k) \le 1/k$ for all $k \ge 1$.*

Proof. Let s_k be such that $\sqrt{20/\log_2 s_k} \le 1/k$, let q_1 be such that the lengths of the intervals described by γ are integer multiples of $1/q_1$, define q_k inductively by $q_k = s_k q_{k-1}$, and let $\mathcal{P}_k$ be the partition of $[0,1]$ into intervals of lengths $1/q_k$. We will define f_k as $f_k = (g_k)_{\mathcal{P}_k}$ where $g_k \in \mathcal{W}_\Omega$ will be inductively defined in such a way that (a) $(g_k)_{\mathcal{P}_{k-1}} = (g_{k-1})_{\mathcal{P}_{k-1}}$ for all $k \ge 2$, (b) $\delta_\square(f, g_k) = 0$ for all $k \ge 1$ and (c) $d_\square(g_k, (g_k)_{\mathcal{P}_k}) \le 1/k$ for all $k \ge 1$. This clearly implies the statement of the lemma, since $g_k \in \mathcal{W}_\Omega$ implies $(g_k)_{\mathcal{P}_k} \in \mathcal{W}_\Omega$ by the fact that $\mathcal{P}_k$ is a refinement of $\mathcal{P}_1$,

$$(f_{k+1})_{\mathcal{P}_k} = ((g_{k+1})_{\mathcal{P}_{k+1}})_{\mathcal{P}_k} = (g_{k+1})_{\mathcal{P}_k} = (g_k)_{\mathcal{P}_k} = f_k$$

by (a) and the fact that $\mathcal{P}_{k+1}$ is a refinement of $\mathcal{P}_k$, and the two statements (b) and (c) imply (ii).

We start our inductive construction by setting $g_1 = f$. Noting that $d_\square(h, h') \le 1$ for all $h, h' \in \mathcal{W}$, this shows that g_1 satisfies the inductive assumptions.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Let $k \geq 2$ and assume that g_{k-1} satisfies the inductive assumption. By Lemma 41, we can find an equipartition $\mathcal{Q}_k$ of $[0,1]$ into $q_k = s_k|\mathcal{P}_{k-1}|$ classes such that $\mathcal{Q}_k$ refines $\mathcal{P}_{k-1}$ and $d_\square(g_{k-1}, (g_{k-1})_{\mathcal{Q}_k}) \leq 1/k$. We now define a measure preserving bijection $\phi : [0,1] \to [0,1]$ as follows: Let I be an interval in $\mathcal{P}_{k-1}$, and let $Y_1, \ldots Y_{s_k}$ those elements of $\mathcal{Q}_k$ subdividing I. By Theorem A.7 in [26], we can find a measure preserving bijection from I to itself such that the image of $Y_1, \ldots Y_{s_k}$ are the s_k intervals in $\mathcal{P}_k$ that subdivide I. Doing this for all intervals in $\mathcal{Q}_k$ we obtain a measure preserving bijection ϕ such that the images of the partition classes of $\mathcal{Q}_k$ are the partition classes of $\mathcal{P}_k$, and such that ϕ maps each interval in $\mathcal{P}_{k-1}$ onto itself. Applying this bijection to g_{k-1} gives a graphon $g_k \in \mathcal{W}_\Omega$ such that $\delta_\square(g_k, g_{k-1}) = 0$ and $d_\square(g_k, (g_k)_{\mathcal{P}_k}) = d_\square(g_{k-1}, (g_{k-1})_{\mathcal{Q}_k}) \leq 1/k$. By the inductive assumption (b) we have that $\delta_\square(g_k, f) = 0$, and by the fact that ϕ maps each interval in $\mathcal{P}_{k-1}$ onto itself we have that $(g_k)_{\mathcal{P}_{k-1}} = (g_{k-1})_{\mathcal{P}_{k-1}}$. This completes the inductive proof. □

We show that $\widetilde{\mathcal{W}}_\Omega$ is closed, using ideas from the proof that $(\widetilde{\mathcal{W}}, \delta_\square)$ is compact [33, Theorem 5.1].

Proof of Lemma 31. We establish the lemma by showing that $\widetilde{\mathcal{W}}_\Omega$ contains its limit points. Let $(\widetilde{W}_n)_{n\geq 0}$ be a sequence of graphons in $\widetilde{\mathcal{W}}_\Omega$ that converges to $\widetilde{W} \in \widetilde{\mathcal{W}}$. Since $\widetilde{W}_n \in \widetilde{\mathcal{W}}_\Omega$, we may chose a sequence $W_n \in \mathcal{W}_\Omega$ such that $\delta_\square(W_n, \widetilde{W}) \to 0$. We claim that $\widetilde{W} \in \widetilde{\mathcal{W}}_\Omega$.

By Lemma 42, we can find a sequence of refining partitions $\mathcal{P}_k$ of $[0,1]$ into intervals of length $1/|\mathcal{P}_k|$ and sequences $W_{n,k} \in \mathcal{W}_\Omega$ such that

1. $\delta_\square(W_n, W_{n,k}) \leq 1/k$
2. $(W_{n,k+1})_{\mathcal{P}_k} = W_{n,k}$.

Next we claim that it is possible to replace (W_n) with a subsequence such that for all k, $W_{n,k}$ converges almost everywhere to a step function U_k with steps made out of the intervals in $\mathcal{P}_k$. Indeed, select a subsequence of (W_n) such that the value of $W_{n,1}$ converges on the product of all intervals $I, I' \in \mathcal{P}_1$. We obtain $W_{n,1} \to U_1$ almost everywhere for U_1 a step function on s_1 intervals of $[0,1]$. Taking further subsequences for $k = 2, 3 \ldots$, we obtain a subsequence of (W_n) such that $W_{n,k} \to U_k$ almost everywhere for all k. By the Dominated Convergence Theorem, $\|W_{n,k} - U_k\|_1 \to 0$, and so $d_\square(W_{n,k}, U_k) \to 0$. Each U_k is a step function on s_k intervals of $[0,1]$. Note that since $W_{n,k} \in \mathcal{W}_\Omega$ and $d_\square(W_{n,k}, U_k) \to 0$, Proposition 37 implies that $U_k \in \mathcal{W}_\Omega$. For the remainder of this proof, we replace (W_n) with this subsequence; doing so does not change the limit of the corresponding sequence in $\widetilde{\mathcal{W}}$.

Next we claim that the sequence $(U_k)_{k\geq 1}$ has a limit U in $\mathcal{W}_\Omega$. It follows from (ii) that $U_k = (U_\ell)_{\mathcal{P}_k}$ for all $\ell > k$. Let (x, y) be a uniform random point in $[0,1]$. Since $U_k = (U_\ell)_{\mathcal{P}_k}$, the sequence $(U_1(x,y), U_2(x,y), \ldots)$ is a martingale with respect to the canonical filtration. The random variables $U_i(x,y)$ are bounded, and so the Martingale Convergence Theorem [21, Theorem 4.2.11] implies that the sequence $(U_1(x,y), U_2(x,y), \ldots)$ converges with probability one. Thus there exists $U \in \mathcal{W}$ such that $U_k \to U$ almost everywhere. By the Dominated Convergence Theorem, $\|U_k - U\|_1 \to 0$ and therefore $d_\square(U, U_k) \to 0$. Since $U_k \in \mathcal{W}_\Omega$ for all k, and $\mathcal{W}_\Omega$ is closed, it follows that $U \in \mathcal{W}_\Omega$. Moreover $\widetilde{U} \in \widetilde{\mathcal{W}}_\Omega$.

It remains to show that $\delta_\square(W_n, U) \to 0$ (as this implies that $\delta_\square(\widetilde{W}_n, \widetilde{U}) \to 0$, which establishes that the limit of the sequence is in $\widetilde{\mathcal{W}}_\Omega$). Let $\varepsilon > 0$. Choose $k > 3/\varepsilon$ sufficiently large such that $\|U - U_k\|_1 < \varepsilon/3$. For this fixed k, there exists n_0 such that $\|U_k - W_{n,k}\|_1 < \varepsilon/3$ for all $n \geq n_0$. Observe that

$$\begin{aligned}\delta_\square(U, W_n) &\leq d_\square(U, U_k) + d_\square(U_k, W_{n,k}) + \delta_\square(W_{n,k}, W_n)\\ &\leq \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \delta_\square(W_{n,k}, W_n) \leq \varepsilon.\end{aligned}$$

□

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Proof of Lemma 32. We modify the proof of [20, Lemma 2.1] to allow for W_0 with values in $\{0,1\}$. For $\tilde{f} \in \widetilde{\mathcal{W}}$, let

$$H(\tilde{f}) = \inf_{g \in \mathcal{W} : \delta_\square(g,\tilde{f})=0} I_{W_0}(g).$$

If $\tilde{f} \notin \widetilde{\mathcal{W}}_\Omega$, then $g \notin \mathcal{W}_\Omega$ for all g contributing to the infimum and by Proposition 34, $H(\tilde{f}) = \infty$. On the other hand, if $\tilde{f} \in \widetilde{\mathcal{W}}_\Omega$, there exists a $g \in \mathcal{W}_\Omega$ contributing to the infimum, so with the help of Proposition 34 we conclude that $H(\,\cdot\,)$ is bounded on $\widetilde{\mathcal{W}}_\Omega$. Combined with the fact that

$$h \in B(\tilde{f}, \delta) \iff \delta_\square(\tilde{g}, h) = 0 \text{ for some } \tilde{g} \in S(\tilde{f}, \delta),$$

we obtain that for $\tilde{f} \in \widetilde{\mathcal{W}}_\Omega$

$$J_{W_0}(\tilde{f}) = \sup_{\delta>0} \inf_{h \in B(\tilde{f},\delta)} I_{W_0}(h) = \sup_{\delta>0} \inf_{\tilde{g} \in S(\tilde{f},\delta)} H(\tilde{g}) = \sup_{\delta>0} \inf_{\tilde{h} \in S(\tilde{f},\delta) \cap \widetilde{\mathcal{W}}_\Omega} H(\tilde{h}) = \liminf_{\tilde{h} \to \tilde{f}} H(\tilde{h}).$$

Therefore, $J_{W_0}(\,\cdot\,)$ is the pointwise lim inf of a bounded function. This implies that $J_{W_0}(\,\cdot\,)$ is lower semi-continuous on $\widetilde{\mathcal{W}}_\Omega$.

Note that $J_{W_0}(f) = \infty$ for all $f \in \widetilde{\mathcal{W}} \setminus \widetilde{\mathcal{W}}_\Omega$. Therefore the lower semi-continuity of $J_{W_0}(\,\cdot\,)$ on $\widetilde{\mathcal{W}}$ follows by Proposition 38 and Lemma 31. □

We close this preliminary section with a proposition and a lemma which will be used in the proofs of Theorems 19, 20 and 23.

Proposition 43. *Let $m \in \mathbb{Z}^+$ and $\gamma \in \Delta_m$. Let τ be a continuous graph parameter. Then the following holds:*

(i) *The set $\{\tilde{g} \in \tilde{\mathcal{B}}^\gamma : \tau(\tilde{g}) \geq t\}$ is a compact set in $(\widetilde{\mathcal{W}}, \delta_\square)$.*
(ii) *The set $\{g \in \mathcal{B}^\gamma : \tau(g) \geq t\}$ is a closed set in $(\mathcal{W}, d_\square)$.*

Proof. Since $\widetilde{\mathcal{W}}$ is compact, it suffices to show that $\{\tilde{g} \in \tilde{\mathcal{B}}^\gamma : \tau(\tilde{g}) \geq t\}$ is closed. Let $\tilde{f}_n \in \tilde{\mathcal{B}}^\gamma$ be such that $\tau(\tilde{f}_n) \geq t$ and $\tilde{f}_n$ converges to some graphon $\tilde{f}$. Since τ is continuous, $\lim_{n\to\infty} \tau(\tilde{f}_n) = \tau(\tilde{f}) \geq t$.

It remains to show that $\tilde{f} \in \tilde{\mathcal{B}}^\gamma$. Without loss of generality, we may assume $f_n \in \mathcal{B}^\gamma$, and write $f_n = (\alpha_{ij}^n)_{i,j\in[m]}$, where each $\alpha_{ij}^n \in [0,1]$. By the compactness of $[0,1]^{m^2}$, there exists a subsequence such that

$$\alpha_{ij}^{n_k} \to \beta_{ij} \text{ for all } i,j \in [m].$$

Let $g = (\beta_{ij})_{i,j\in[m]} \in \mathcal{B}^\gamma$. Since $f_{n_k} \to g$ pointwise and $d_\square(f_{n_k}, g) \leq \|f_{n_k} - g\|_1$, the Dominated Convergence Theorem implies that $d_\square(f_{n_k}, g) \to 0$. Since $\delta_\square(\tilde{f}_n, \tilde{f}) \to 0$, we have $\delta_\square(\tilde{g}, \tilde{f}) = 0$ and thus $\tilde{f} \in \tilde{\mathcal{B}}^\gamma$.

The proof of the second statement is identical, except that it starts from a sequence $f_n \in \mathcal{B}^\gamma$ such that $\tau(f_n) \geq t$ and f_n converges to some $f \in \mathcal{W}$ in the metric $d_\square$. □

Lemma 44. *Suppose $f \in \mathcal{W}$ is of the form $f = \sum_{i,j\in[k]} \beta_{ij} 1_{Y_i} 1_{Y_j}$ where $\beta_{ij} = \beta_{ji} \in [0,1]$ and $Y_1, \ldots, Y_k$ form a partition of $[0,1]$ into measurable sets. Let $g \in \mathcal{W}$. Then $\delta_\square(f,g) = 0$ if and only if there exists a partition of $[0,1]$ into measurable subsets $Y_1', \ldots Y_k'$ such that $g = \sum_{i,j\in[k]} \beta_{ij} 1_{Y_i'} 1_{Y_j'}$ and $\lambda(Y_i) = \lambda(Y_i')$ almost everywhere.*

Proof. If g can be written as $\sum_{i,j\in[k]} \beta_{ij} 1_{Y_i'} 1_{Y_j'}$ with $\lambda(Y_i) = \lambda(Y_i')$ almost everywhere, then clearly $\delta_\square(f,g) = 0$. To establish the other direction, we will use Theorem 8.6 (vi) from [26]. This will require us to turn f into what is called a twin-free graphon, defined as a graphon W such that

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

there exists no pair $(x, x') \in [0,1]$ such that $W(x,\cdot) = W(x',\cdot)$ almost everywhere. Unfortunately, by their very definition, step functions are not twin-free. To remedy this, we introduce graphons over a general probability space $(\Omega, \mathcal{F}, \mu)$, defined as measurable functions $W{:}\Omega^2 \to [0,1]$ such that $W(x,y) = W(y,x)$ for all $x, y \in \Omega$. We also need to define a cut distance between graphons W_i on (potentially) different probability spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$. It is defined as

$$\delta_\square(W_1, W_2) = \inf_{\mu} \sup_{S,T \subset \Omega_1 \times \Omega_2} \left| \int (W_1(x,x') - W_2(y,y')) d\mu(x,y) d\mu(x',y') \right|$$

where the inf goes over all couplings of μ_1 and μ_2. It is easy to see that for graphons defined on $[0,1]$, this definition agrees with the previous one (see, e.g., Lemma 3.5 in [8]). With these new definitions, we define two graphons W_1, W_2 over two possibly different probability spaces to be equivalent if $\delta_\square(W_1, W_2) = 0$.

With this definition, the graphon f is equivalent to the 'discrete' graphon $f'_{ij} = \beta_{i,j}$ where i and j lie in the probability space $([k], 2^{[k]}, \mu)$ with $\mu(i) = \lambda(Y_i)$ (use the coupling which pairs $i \in [k]$ with the uniform measure on Y_i). It is also easy to turn f' into a twin free graphon as follows: if i and i' are twins, i.e., if the i^{th} and j^{th} row of β are identical, just merge the sets Y_i and Y_j into a new set of measure $\lambda(Y_i) + \lambda(Y_j)$, reducing k by one. Note that this does not change the function f, just the representation of the form $f = \sum_{i,j \in [k]} \beta_{ij} 1_{Y_i} 1_{Y_j}$. Iterating this procedure, we eventually obtain a twin free graphon f' which has cut-distance zero from f, $\delta_\square(f', f) = 0$.

At this point, we use Theorem 8.6 (vi) from [26] which says that $\delta_\square(f', g) = 0$ if and only if there exists a measure preserving map $\phi : [0,1] \to [k]$ such that $g(x,y) = f'_{\phi(x),\phi(y)}$ almost everywhere. Defining $Y'_i = \phi^{-1}(\{i\})$ proves the lemma, in the case where the rows of β are pairwise distinct. Otherwise, the intervals of g may be split in order to match the representation of f. □

2.2 Lower bound

In order to prove Statement (1) of Theorem 11, we closely follow the proof of Theorem 2.3 in [14].

Proof of Theorem 11, Statement (1). We will prove the bound in the form given in Lemma 30. Let $f^{G_{kn}}$ be the empirical graphon of a graph on kn vertices drawn according to $\mathbb{P}_{kn,W_0}$. First, we claim that if

$$\liminf_{n\to\infty} \frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}\left(d_\square(f^{G_{kn}}, g) \le \varepsilon\right) \ge -I_{W_0}(g) \tag{13}$$

holds for all $g \in \mathcal{W}$ and $\varepsilon > 0$, then the theorem follows.

To see this, we first observe that for $\tilde{h} \in \widetilde{\mathcal{W}}$, $0 < \eta \le \varepsilon/2$ and $g \in B(\tilde{h}, \eta)$

$$\tilde{\mathbb{P}}_{kn,W_0}(S(\tilde{h}, \varepsilon)) = \mathbb{P}_{kn,W_0}(\delta_\square(f^{G_{kn}}, \tilde{h}) \le \varepsilon) \ge \mathbb{P}_{kn,W_0}(d_\square(f^{G_{kn}}, g) \le \varepsilon/2)$$

where the identity follows from the definition of $\tilde{\mathbb{P}}_{kn,W_0}$ and $S(\tilde{h}, \varepsilon)$, and the lower bound follows upon noting that $\delta_\square(f^{G_{kn}}, \tilde{h}) \le d_\square(f^{G_{kn}}, g) + \delta_\square(g, \tilde{h})$. Therefore, assuming (13) yields

$$\liminf_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}(S(\tilde{h}, \varepsilon)) \ge -I_{W_0}(g)$$

for all $0 < \eta \le \varepsilon/2$ and all $g \in B(\tilde{h}, \varepsilon/2)$. It follows that

$$\liminf_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}(S(\tilde{h}, \varepsilon)) \ge -\sup_{\eta \in (0,\varepsilon/2]} \inf_{g \in B(\tilde{h},\eta)} I_{W_0}(g) = -\sup_{\eta > 0} \inf_{g \in B(\tilde{h},\eta)} I_{W_0}(g) = -J_{W_0}(\tilde{h})$$

as required.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

We have shown that (13) implies the theorem; we now turn to its proof. By Proposition 34, (13) holds trivially for $g \notin \mathcal{W}_\Omega$. We may assume $g \in \mathcal{W}_\Omega$. Let $\varepsilon > 0$. We define $(g_n)_{n\geq 1}$, a sequence of $kn \times kn$ block graphons that approximate g. Recall (1). For $i, j \in [kn]$, let

$$p_{ij}^{(n)} = (kn)^2 \int \int_{[\frac{i-1}{kn}, \frac{i}{kn}]\times[\frac{j-1}{kn}, \frac{j}{kn}]} g(x,y)dxdy \qquad \text{and} \qquad g_n(x,y) = p^{(n)}_{\vartheta_{kn}(x),\vartheta_{kn}(y)}.$$

Since $\|g_n - g\|_1 \to 0$, in order to prove (13) it suffices to show that

$$\liminf_{n\to\infty} \frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}\left(B_{\varepsilon,n}\right) \geq -I_{W_0}(g),$$

where $B_{\varepsilon,n} = \{f : d_\square(f, g_n) \leq \varepsilon/2\}$. We will apply the following proposition, which is proved as part of Theorem 2.3 in [14]. For completion, we include a proof sketch in the Appendix (Section 8.2).

Proposition 45. *Let f_n be a graphon drawn from the measure $\mathbb{P}_{kn,g_n}$, with g_n as defined above. For any $\varepsilon > 0$, it holds that*

$$\lim_{n\to\infty} \mathbb{P}_{kn,g_n}(d_\square(f_n, g_n) \geq \varepsilon) = 0.$$

We now apply a tilting argument to establish the lower bound, following [14]. Note that since $g \in \mathcal{W}_\Omega$, $g_n \in \mathcal{W}_\Omega$, and so the support of $\mathbb{P}_{kn,g_n}$ is contained in the support of $\mathbb{P}_{kn,W_0}$. Therefore we may write

$$\begin{aligned}\mathbb{P}_{kn,W_0}(B_{\varepsilon,n}) &= \int_{B_{\varepsilon,n}} d\mathbb{P}_{kn,W_0} = \int_{B_{\varepsilon,n}} \exp\left(-\log \frac{d\mathbb{P}_{kn,g_n}}{d\mathbb{P}_{kn,W_0}}\right) d\mathbb{P}_{kn,g_n} \\ &= \mathbb{P}_{kn,g_n}(B_{\varepsilon,n}) \frac{1}{\mathbb{P}_{kn,g_n}(B_{\varepsilon,n})} \int_{B_{\varepsilon,n}} \exp\left(-\log \frac{d\mathbb{P}_{kn,g_n}}{d\mathbb{P}_{kn,W_0}}\right) d\mathbb{P}_{kn,g_n}.\end{aligned}$$

Proposition 45 implies that for n sufficiently large, $\mathbb{P}_{kn,W_0}(B_{\varepsilon,n}) > 0$. Taking the logarithm of both sides and applying Jensen's inequality, we obtain

$$\log \mathbb{P}_{kn,W_0}(B_{\varepsilon,n}) \geq \log \mathbb{P}_{kn,g_n}(B_{\varepsilon,n}) - \frac{1}{\mathbb{P}_{kn,g_n}(B_{\varepsilon,n})} \int_{B_{\varepsilon,n}} \log\left(\frac{d\mathbb{P}_{kn,g_n}}{d\mathbb{P}_{kn,W_0}}\right) d\mathbb{P}_{kn,g_n}.$$

By Proposition 45, it holds that $\mathbb{P}_{kn,g_n}(B_{\varepsilon,n}) \to 1$. Furthermore, letting $p_{\min}$ and $p_{\max}$ be the minimal and maximal values taken on by W_0 in $(0,1)$, we have

$$\begin{aligned}&\frac{1}{(kn)^2} \int_{\mathcal{W}_\Omega \setminus B_{\varepsilon,n}} \log\left(\frac{d\mathbb{P}_{kn,g_n}}{d\mathbb{P}_{kn,W_0}}\right) d\mathbb{P}_{kn,g_n} \\ &\leq \frac{1}{(kn)^2} \int_{\mathcal{W}_\Omega \setminus B_{\varepsilon,n}} \log\left(\left(\frac{1}{\min\{p_{\min}, 1-p_{\max}\}}\right)^{\binom{kn}{2}}\right) d\mathbb{P}_{kn,g_n} \to 0.\end{aligned}$$

Therefore,

$$\liminf_{n\to\infty} \frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}(B_{\varepsilon,n}) \geq -\limsup_{n\to\infty} \frac{1}{(kn)^2} \int \log\left(\frac{d\mathbb{P}_{kn,g_n}}{d\mathbb{P}_{kn,W_0}}\right) d\mathbb{P}_{kn,g_n}.$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Observe that the probability of sampling a graph G on kn vertices is

$$\mathbb{P}_{kn,g_n}(\{G\}) = \prod_{\substack{i,j\in[kn]\\ i<j}} \left(p_{ij}^{(n)}\mathbb{1}_{(i,j)\in E(G)} + (1-p_{ij}^{(n)})\mathbb{1}_{(i,j)\notin E(G)}\right).$$

By construction if $W_0(i/(kn), j/(kn)) = 0$, then $p_{ij}^{(n)} = 0$ and if $W_0(i/(kn), j/(kn)) = 1$, then $p_{ij}^{(n)} = 1$. Let $V(G)$ denote the vertex set of G. Recalling the convention that $0\log(0/0) = 0$, we obtain

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{(kn)^2}\int \log\left(\frac{d\mathbb{P}_{kn,g_n}}{d\mathbb{P}_{kn,W_0}}\right) d\mathbb{P}_{kn,g_n}\\
&= \limsup_{n\to\infty} \frac{1}{(kn)^2} \sum_{G:V(G)=[kn]} \mathbb{P}_{kn,g_n}(\{G\}) \log\left(\frac{\mathbb{P}_{kn,g_n}(\{G\})}{\mathbb{P}_{kn,W_0}(\{G\})}\right)\\
&= \limsup_{n\to\infty} \frac{1}{(kn)^2} \sum_{G:V(G)=[kn]} \prod_{\substack{i,j\in[kn]\\ i<j}} \left(p_{ij}^{(n)}\mathbb{1}_{(i,j)\in E(G)} + (1-p_{ij}^{(n)})\mathbb{1}_{(i,j)\notin E(G)}\right) \times\\
&\qquad \left(\sum_{\substack{i,j\in[kn]\\ i<j}} \log\frac{p_{ij}^{(n)}}{W_0(\frac{i}{kn},\frac{j}{kn})}\mathbb{1}\{(i,j)\in E(G)\} + \log\frac{1-p_{ij}^{(n)}}{1-W_0(\frac{i}{kn},\frac{j}{kn})}\mathbb{1}\{(i,j)\notin E(G)\}\right)\\
&= \limsup_{n\to\infty} \frac{1}{(kn)^2} \sum_{\substack{i,j\in[kn]\\ i<j}} p_{ij}^{(n)} \log\frac{p_{ij}^{(n)}}{W_0(\frac{i}{kn},\frac{j}{kn})} + (1-p_{ij}^{(n)})\log\frac{1-p_{ij}^{(n)}}{1-W_0(\frac{i}{kn},\frac{j}{kn})}\\
&= \limsup_{n\to\infty} I_{W_0}(g_n) = I_{W_0}(g).
\end{aligned}$$

The last equality follows from a straightforward adaptation of an argument from [11, Lemma 5.7]: for any $\varepsilon > 0$, set $A_\varepsilon = \{(x,y) : \|g_n(x,y) - g(x,y)\|_1 > \varepsilon\}$. Observe that $\|g_n - g\|_1 \geq \lambda(A_\varepsilon)\varepsilon$, implying $\lambda(A_\varepsilon) \to 0$ for any $\varepsilon > 0$. In addition, note that $Im(W_0)$ is finite and $h_p(\,\cdot\,)$ is uniformly continuous for any fixed p. Therefore, for every $\delta > 0$, there exists $\varepsilon > 0$ such that $|h_{W_0(x,y)}(g_n(x,y)) - h_{W_0(x,y)}(g(x,y))| < \delta$ whenever $|g_n(x,y) - g(x,y)| < \varepsilon$. Consequently,

$$\begin{aligned}
|I_{W_0}(g_n) - I_{W_0}(g)| &\leq \frac{1}{2}\int_{A_\varepsilon} |h_{W_0(x,y)}(g_n(x,y)) - h_{W_0(x,y)}(g(x,y))|dxdy\\
&\quad + \frac{1}{2}\int_{[0,1]^2\setminus A_\varepsilon} |h_{W_0(x,y)}(g_n(x,y)) - h_{W_0(x,y)}(g(x,y))|dxdy\\
&\leq \frac{C}{2}\lambda(A_\varepsilon) + \frac{1}{2}\delta
\end{aligned}$$

where $C > 0$ is such that $h_{W_0(x,y)}(a) < C$ for all $a \in [0,1]$ and $(x,y) \in [0,1]^2$. This completes the proof.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

2.3 Upper bound

In this section we prove the upper bound of Theorem 11. The proof requires two key lemmas. The first one establishes that as long as we look at balls around block constant graphons, we can restrict to a finite number of measure preserving bijections, and the second one gives an upper bound on the probability that sampling a graphon and applying an invertible transformation yields a graphon in a particular $d_\square$ ball. This is formalised in the following two lemmas. Let $\mathcal{M}_{kn}$ be the set of all permutations of $[0,1]$ corresponding to permuting the vertices of a graph on kn vertices, i.e. if $\sigma \in \mathcal{M}_{kn}$, then for all $i \in [kn]$, the image of the interval $\vartheta_{kn}^{-1}(i)$ under σ is an interval of the form $\vartheta_{kn}^{-1}(j)$ for some $j \in [kn]$.

Lemma 46. *Let $s_h, s_w \in \mathbb{Z}^+$ and $\eta > 0$. Given $\mu \in \Delta_{s_h}$, $\gamma \in \Delta_{s_w}$, $h \in \mathcal{B}^\mu$, and a k-uniform block graphon $W_0 \in \mathcal{B}^\gamma$, there exists a finite set of invertible measure preserving transformations $T \subseteq \mathcal{M}$ with $|T| = N(\eta, s_w, s_h)$ such that the following holds for all $n > 12s_h^2 s_w/(k\eta)$: For all $\sigma \in \mathcal{M}_{kn}$ there exists $\tau \in T$ such that for all $\varepsilon > 0$*

$$\mathbb{P}_{kn,W_0}\left(d_\square(f^{G^\sigma_{kn}}, h) \le \varepsilon\right) \le \mathbb{P}_{kn,W_0}\left(d_\square(f^{G^\tau_{kn}}, h) \le \varepsilon + \eta\right),$$

where $f^{G_{kn}}$ is the empirical graphon obtained by sampling W_0 according to $\mathbb{P}_{kn,W_0}$.

Lemma 47. *Let W_0 be a uniform k-block graphon, $\varepsilon > 0$ and $h \in \mathcal{W}$. Let G_{kn} be the graph drawn from $\mathbb{P}_{kn,W_0}$, and let $f^{G_{kn}}$ denote the corresponding empirical graphon. For all invertible $\tau \in \mathcal{M}$, it holds that*

$$\limsup_{n\to\infty} \frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}\left(d_\square\left(f^{G^\tau_{kn}}, h\right) \le \varepsilon\right) \le -\inf_{f:\delta_\square(f,h)\le\varepsilon} I_{W_0}(f).$$

To prove the second lemma, we use the following LDP upper bound with respect to the weak topology. Recall that the weak topology on $\mathcal{W}$ is the smallest topology under which the maps $f \mapsto \int_{[0,1]^2} f(x,y)g(x,y)dxdy$ are continuous for every $g \in L^2([0,1]^2)$.

Theorem 48. *Let W_0 be a uniform k-block graphon. For every weakly closed set $F \in \mathcal{W}$,*

$$\limsup_{n\to\infty} \frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}(F) \le -\inf_{f\in F} I_{W_0}(f).$$

The proof of Theorem 48 is a straightforward generalisation of Theorem 5.1 of [11]. For completeness we include the proof in the Appendix. We delay the proofs of Lemmas 46 and 47 to the end of the section and proceed to the proof of the upper bound in Theorem 11.

In the proof of the upper bound in Theorem 11, we also use the following version of the Weak Regularity Lemma, which follows directly from Theorem 3.1 in [11].

Lemma 49 (Weak Regularity Lemma). *Given $\varepsilon > 0$ there exists a finite set $H(\varepsilon) \subset \mathcal{W}$ of block-graphons, such that if f is a uniform n-block graphon there exists $\sigma \in \mathcal{M}_n$ and an $h \in H(\varepsilon)$ such that*

$$d_\square(f^\sigma, h) < \varepsilon.$$

We call such a set $H(\varepsilon) \subset \mathcal{W}$ an ε-net.

Proof of Theorem 11, Statement 2. We prove the bound in the form given in Lemma 30. Fix $\tilde{g}$ and $\alpha > 0$. Let $\varepsilon, \eta < \alpha/2$, let $H(\varepsilon/2)$ be an $\varepsilon/2$ net as given in Lemma 49, and let T^* be the union of the sets of invertible transformations T given by Lemma 46 for each $h \in H(\varepsilon/2)$. We index the finite set as $T^* = \{\tau_1, \tau_2, \ldots \tau_{N(\eta,W_0,\varepsilon)}\}$. Then

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$$\begin{aligned}
\tilde{\mathbb{P}}_{kn,W_0}\left(S(\tilde{g},\alpha)\right) &= \mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}},g)\le\alpha\right)\\
&\le \sum_{h\in H(\varepsilon/2)} \mathbb{P}_{kn,W_0}\left(\bigcup_{\sigma\in\mathcal{M}_{kn}}\left\{f^{G_{kn}^\sigma}\in\{f:\delta_\square(f,g)\le\alpha\}\cap\left\{f:d_\square(f,h)\le\frac{\varepsilon}{2}\right\}\right\}\right)\\
&\le \sum_{h\in H(\varepsilon/2)}\sum_{\sigma\in\mathcal{M}_{kn}} \mathbb{P}_{kn,W_0}\left(\left\{f^{G_{kn}^\sigma}\in\{f:\delta_\square(f,g)\le\alpha\}\cap\left\{f:d_\square(f,h)\le\frac{\varepsilon}{2}\right\}\right\}\right)\\
&\le \sum_{\substack{h\in H(\varepsilon/2)\\ \delta_\square(g,h)\le\alpha+\varepsilon/2}}\sum_{\sigma\in\mathcal{M}_{kn}} \mathbb{P}_{kn,W_0}\left(d_\square(f^{G_{kn}^\sigma},h)\le\frac{\varepsilon}{2}\right)\\
&\le (kn)!|H(\varepsilon/2)| \max_{\substack{h\in H(\varepsilon/2)\\ \delta_\square(g,h)\le\alpha+\varepsilon/2\\ i\in[N(\eta,W_0,\varepsilon)]}} \mathbb{P}_{kn,W_0}\left(d_\square(f^{G_{kn}^{\tau_i}},h)\le\frac{\varepsilon}{2}+\eta\right),
\end{aligned}$$

where the last inequality follows from Lemma 46. Lemma 47 implies that

$$\begin{aligned}
\limsup_{n\to\infty}\frac{1}{(kn)^2}\log\mathbb{P}_{kn,W_0}\left(d_\square(f^{G_{kn}^{\tau_i}},h)\le\frac{\varepsilon}{2}+\eta\right) &\le -\inf_{f:\delta_\square(f,h)\le\frac{\varepsilon}{2}+\eta} I_{W_0}(f)\\
&\le -\inf_{f:\delta_\square(f,\tilde{g})\le\alpha+\varepsilon+\eta} I_{W_0}(f),
\end{aligned}$$

when h is such that $\delta_\square(g,h)\le\alpha+\varepsilon/2$ and τ_i is invertible.

Let $M\in\mathbb{Z}^+$, and consider the M sequences $\{y_n^i\}_{n\ge1}$ where $i\in[M]$. A basic analysis argument implies that if $\limsup_{n\to\infty}\frac{1}{(kn)^2}\log y_n^i\le C$ for all $i\in[M]$, then $\limsup_{n\to\infty}\frac{1}{(kn)^2}\log\left(\max_{i\in[M]}y_n^i\right)\le C$. It follows that

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}\left(S(\tilde{g},\alpha)\right)\\
&\le\limsup_{n\to\infty}\frac{1}{(kn)^2}\log\left((kn)!|H(\varepsilon/2)|\max_{\substack{h\in H(\varepsilon/2)\\ \delta_\square(g,h)\le\alpha+\varepsilon/2\\ i\in[N(\eta,W_0,\varepsilon)]}}\mathbb{P}_{kn,W_0}\left(d_\square(f^{G_{kn}^{\tau_i}},h)\le\varepsilon/2+\eta\right)\right)\\
&=\limsup_{n\to\infty}\frac{1}{(kn)^2}\log\left(\max_{\substack{h\in H(\varepsilon/2)\\ \delta_\square(g,h)\le\alpha+\varepsilon/2\\ i\in[N(\eta,W_0,\varepsilon)]}}\mathbb{P}_{kn,W_0}\left(d_\square(f^{G_{kn}^{\tau_i}},h)\le\varepsilon/2+\eta\right)\right)\\
&\le -\inf_{f:\delta_\square(f,\tilde{g})\le\alpha+\varepsilon+\eta} I_{W_0}(f).
\end{aligned}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Since $\varepsilon, \eta < \alpha/2$, we obtain

$$\limsup_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}\left(S(\tilde{g},\alpha)\right) \le -\inf_{f:\delta_\square(f,\tilde{g})\le 2\alpha} I_{W_0}(f).$$

Since $(-\inf_{f:\delta_\square(f,\tilde{g})\le 2\alpha} I_{W_0}(f))$ is a non-increasing function as $\alpha \to 0$,

$$\begin{aligned}\lim_{\alpha\to 0}\limsup_{n\to\infty} \frac{1}{(kn)^2} \log \tilde{\mathbb{P}}_{kn,W_0}\left(S(\tilde{g},\alpha)\right) &\le \inf_{\alpha>0}\left(-\inf_{f:\delta_\square(f,\tilde{g})\le 2\alpha} I_{W_0}(f)\right)\\ &= -\sup_{\alpha>0}\inf_{f:\delta_\square(f,\tilde{g})\le 2\alpha} I_{W_0}(f)\\ &= -J_{W_0}(g),\end{aligned}$$

as required. □

We will use the following definition and proposition in the proof of Lemma 46.

Definition 50. *Let $\phi \in \mathcal{M}$. Let $I_1 \ldots, I_k$ be a partition of $[0,1]$. We say that ϕ respects the interval structure of $I_1, \ldots, I_k$ if for all j and $X \subseteq I_j$, $\phi(X) \subseteq I_j$.*

Proposition 51. *Let h be a graphon that is constant on each block $I_i \times I_j$ for $i,j \in [k]$. If ϕ is invertible and respects the interval structure of $I_1 \ldots I_k$, then for all $g \in \mathcal{W}$*

$$d_\square(g^\phi, h) = d_\square(g,h).$$

Proof. Note that since h is constant on each block $I_i \times I_j$ and ϕ respects the interval structure,

$$h^\phi(x,y) = h(\phi(x),\phi(y)) = h(x,y).$$

It follows that

$$d_\square(g,h) = d_\square(g^\phi, h^\phi) = d_\square(g^\phi, h).$$

□

Proof of Lemma 46. For ease of notation, we denote $f = f^{G_{kn}}$ as the empirical graphon obtained by sampling W_0 according to $\mathbb{P}_{kn,W_0}$. Let $\{H_i : i \in [s_h]\}$ and $\{W_{0,j} : j \in [s_w]\}$ denote the intervals of the block structure governed by μ and γ respectively; formally, $H_1 = [0, \mu_1]$, $W_{0,1} = [0, \gamma_1]$,

$$H_i = \left(\sum_{j=1}^{i-1}\mu_j, \sum_{j=1}^{i}\mu_j\right] \text{ for } 2 \le i \le s_h, \quad \text{and} \quad W_{0,i} = \left(\sum_{j=1}^{i-1}\gamma_j, \sum_{j=1}^{i}\gamma_j\right] \text{ for } 2 \le i \le s_w.$$

$\{H_i\}$ are the intervals corresponding to the blocks of h and $\{W_{0,i}\}$ are the intervals corresponding to the blocks of W_0.

We begin with a proof outline. First we construct a finite set of measure preserving transformations $T \subseteq \mathcal{M}$ such that each $\sigma \in \mathcal{M}_{kn}$ is 'close' to a transformation τ in T. Then, in order to compare f^σ and f^τ, we define α, β invertible transformations that respect the intervals $H_1, \ldots H_{s_h}$. Proposition 51 implies that $d_\square(f^\sigma, h) = d_\square((f^\sigma)^\alpha, h)$ and $d_\square(f^\tau, h) = d_\square((f^\tau)^\beta, h)$. It follows that

$$\mathbb{P}_{kn,W_0}\left(d_\square(f^\sigma, h) \le \varepsilon\right) = \mathbb{P}_{kn,W_0}\left(d_\square(\left(f^\sigma\right)^\alpha, h) \le \varepsilon\right), \tag{14}$$

$$\mathbb{P}_{kn,W_0}\left(d_\square(f^\tau, h) \le \varepsilon+\eta\right) = \mathbb{P}_{kn,W_0}\left(d_\square(\left(f^\tau\right)^\beta, h) \le \varepsilon+\eta\right). \tag{15}$$

Finally, we will describe a coupling of f and g, each with marginal distribution $\mathbb{P}_{kn,W_0}$, that guarantees

$$d_\square\left(\left(f^\sigma\right)^\alpha, \left(g^\tau\right)^\beta\right) \le \eta \tag{16}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

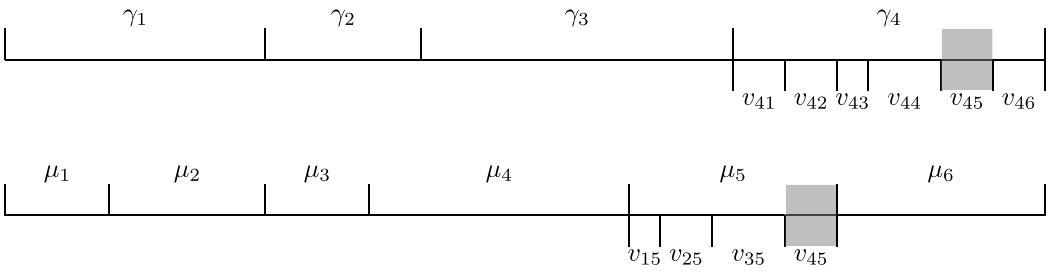


Figure 3. Illustration of the transformation $I_{45} \mapsto \tau(I_{45})$.

with probability one. The triangle inequality implies

$$\mathbb{P}_{kn,W_0}\left(d_\square(\left(f^\sigma\right)^\alpha, h) \le \varepsilon\right) \le \mathbb{P}_{kn,W_0}\left(d_\square(\left(f^\tau\right)^\beta, h) \le \varepsilon + \eta\right). \tag{17}$$

The statement follows directly from (14), (15), and (17).

To complete the proof according to this outline, we must complete the following tasks:

1. Define a finite net of measure preserving transformations $T \subseteq \mathcal{M}$ with $|T| = N(\eta, s_h, s_w)$.
2. For each transformation $\sigma \in \mathcal{M}_{kn}$, define a 'close' transformation $\tau \in T$. Informally, we will say that two transformations are close if they map approximately the same amount of mass from $W_{0,i}$ to H_j for all $i \in [s_w]$ and $j \in [s_h]$.
3. Define α and β, invertible transformations that respect the intervals $H_1 \dots H_{s_h}$.
4. Exhibit a coupling of f and g each sampled according to $\mathbb{P}_{kn,W_0}$ that guarantees (16).

We begin with (a). For convenience we index vectors $v \in \mathbb{R}^{s_w \cdot s_h}$ by pairs $(i,j) \in [s_w] \times [s_h]$. Let

$$V = \left\{ v \in \mathbb{R}^{s_w \cdot s_h}_{\ge 0} \;\middle|\; \sum_{i=1}^{s_w}\sum_{j=1}^{s_h} v_{ij} = 1, \sum_{i=1}^{s_w} v_{ij} = \mu_j, \sum_{j=1}^{s_h} v_{ij} = \gamma_i \right\}.$$

Recall that $\{H_i\}$ are the intervals corresponding to the blocks of h and $\{W_{0,i}\}$ are the intervals corresponding to the blocks of W_0. For each $v \in V$ we associate an invertible measure preserving transformation $\tau_v \in \mathcal{M}$ that maps an interval of length v_{ij} contained in $W_{0,i}$ to an interval that is contained completely in H_j for each (i,j). To this end, let

$$I_{11} = [0, v_{11}] \qquad \text{and} \qquad I_{ij} = \left(\sum_{a=1}^{i-1} \gamma_a + \sum_{b=1}^{j-1} v_{ib}, \sum_{a=1}^{i-1} \gamma_a + \sum_{b=1}^{j} v_{ib} \right], (i,j) \neq (1,1).$$

Note that the intervals I_{ij} are sorted first by the first index, then by the second index. Define $\tau : [0,1] \to [0,1]$ to be the transformation that translates the intervals so they are first sorted by the second index, then by the first index,

$$\tau(I_{11}) = [0, v_{11}] \qquad \text{and} \qquad \tau(I_{ij}) = \left(\sum_{b=1}^{j-1} \mu_b + \sum_{a=1}^{i-1} v_{aj}, \sum_{b=1}^{j-1} \mu_b + \sum_{a=1}^{i} v_{aj} \right], (i,j) \neq (1,1).$$

For an illustration of this transformation, see Figure 3.

Observe that $I_{ij} \subseteq W_{0,i}$ and $\tau(I_{ij}) \subseteq H_j$. Note that V is a compact set, and thus we can construct a finite net $V^* \subseteq V$ such that for all $v \in V$, there exists $v^* \in V^*$ such that $\|v - v^*\|_\infty \le \eta/(8 s_w s_h)$. Let $T = \{\tau_v | v \in V^*\} \cup \{\tau_v^{-1} | v \in V^*\}$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Next we address (b). Let $M = \{\bar{m}_1, \dots \bar{m}_{kn}\}$ be the set of intervals corresponding to vertices in an empirical graphon with kn vertices, where $\bar{m}_1 = \left[0, \frac{1}{kn}\right]$ and

$$\bar{m}_i = \left(\frac{i-1}{kn}, \frac{i}{kn}\right] \text{ for } 2 \le i \le kn.$$

We call the intervals in M 'vertex intervals'.

For each transformation in $\mathcal{M}_{kn}$, we find a transformation in T that moves roughly the same amount of mass between intervals $W_{0,i}$ and H_j for all $i \in [s_w]$ and $j \in [s_h]$. We construct an element in T that is close to σ^{-1} rather than σ to make the construction in the next section more convenient. (Note that both σ and σ^{-1} are in $\mathcal{M}_{kn}$.)

Let N_{ij} be the set of vertex intervals that are mapped from $W_{0,i}$ to H_j under σ^{-1},

$$N_{ij} = \{\bar{m}_\ell \in M | \bar{m}_\ell \subseteq W_{0,i} \text{ and } \sigma(\bar{m}_\ell) \subseteq H_j\},$$

and let $n_{ij} = |N_{ij}|$.

Each vertex interval is contained in some $W_{0,i}$ by construction. Since σ^{-1} maps at most $s_h - 1$ vertex intervals to the boundary between intervals of h,

$$\sum_{i=1}^{s_w} \sum_{j=1}^{s_h} n_{ij} \ge kn - s_h + 1.$$

Define $v \in \mathbb{R}^{s_w \cdot s_h}$, $v_{ij} = n_{ij}/(kn)$. We claim that there exists $v' \in V$ with $v'_{ij} \ge v_{ij}$ for all i, j such that

$$\|v - v'\|_\infty \le \frac{s_h - 1}{kn} < \frac{\eta}{8 s_w s_h}.$$

To see this, let $\tilde{\mu}_j = \sum_{i=1}^{s_w} v_{ij}$ and $\tilde{\gamma}_i = \sum_{j=1}^{s_h} v_{ij}$. Then $\tilde{\mu}_j \le \mu_j$ and $\tilde{\gamma}_i \le \gamma_i$ and

$$\sum_i (\gamma_i - \tilde{\gamma}_i) = \sum_j (\mu_j - \tilde{\mu}_j) = 1 - \sum_{i=1}^{s_w} \sum_{j=1}^{s_h} v_{ij} =: \Delta, \qquad \Delta \le \frac{s_h - 1}{kn}.$$

Taking a coupling ζ_{ij} of the probability distributions $\left(\frac{1}{\Delta}(\gamma_i - \tilde{\gamma}_i)\right)_{i \in [s_w]}$ and $\left(\frac{1}{\Delta}(\mu_j - \tilde{\mu}_j)\right)_{j \in [s_h]}$ and setting $v'_{ij} = v_{ij} + \zeta_{ij}\Delta$ proves the claim. (Observe that $n \ge 8{s_h}^2 s_w/(k\eta)$ by the assumption of the lemma.)

Choosing $v^* \in V^*$ such that $\|v' - v^*\|_\infty \le \eta/(8 s_w s_h)$, we then have

$$\|v - v^*\|_\infty < \frac{\eta}{4 s_w s_h}.$$

We associate σ^{-1} with $\tau_{v^*} \in T$. Let $\tau = \tau_{v^*}^{-1}$, and note $\tau \in T$.

Next we address (c), defining the transformations α and β. We must define α and β in a way that conveniently facilitates a coupling satisfying (16) in step (d). In particular, we will define a coupling so that $\left(f^\sigma\right)^\alpha$ and $\left(g^\tau\right)^\beta$ are identical on many sets of the form $\bar{m} \times \bar{m}'$ where $\bar{m}$ and $\bar{m}'$ are vertex intervals. We can exactly couple the values $f(\sigma(\alpha(x)), \sigma(\alpha(y)))$ and $g(\tau(\beta(x)), \tau(\beta(y)))$ on $\bar{m} \times \bar{m}'$, provided $\sigma(\alpha(x))$ and $\tau(\beta(x))$ are in the same interval $W_{0,i}$ and $\sigma(\alpha(y))$ and $\tau(\beta(y))$ are in the same interval $W_{0,j}$. In this case, on both $\bar{m}$ and $\bar{m}'$ $f(\sigma(\alpha(x)), \sigma(\alpha(y))), g(\tau(\beta(x)), \tau(\beta(y))) \sim Bern(p_{ij})$ where p_{ij} is the value of W_0 on $W_{0,i} \times W_{0,j}$, and thus they can be coupled.

We say that a vertex interval $\bar{m} \in M$ is 'synchronised' if $\bar{m} \subseteq H_j$ for some $j \in [s_h]$, and $\sigma(\alpha(\bar{m}))$ and $\tau(\beta(\bar{m}))$ are vertex intervals that belong to the same interval $W_{0,i}$ for some $i \in [s_w]$. We construct α and β so that at least a $(1 - \eta/2)$ fraction of the vertex intervals $\bar{m} \in M$ are synchronised. In step (d), we will couple the behaviour of the vertices corresponding to $\sigma(\alpha(\bar{m}))$ and $\tau(\beta(\bar{m}))$ for each synchronised vertex interval $\bar{m}$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

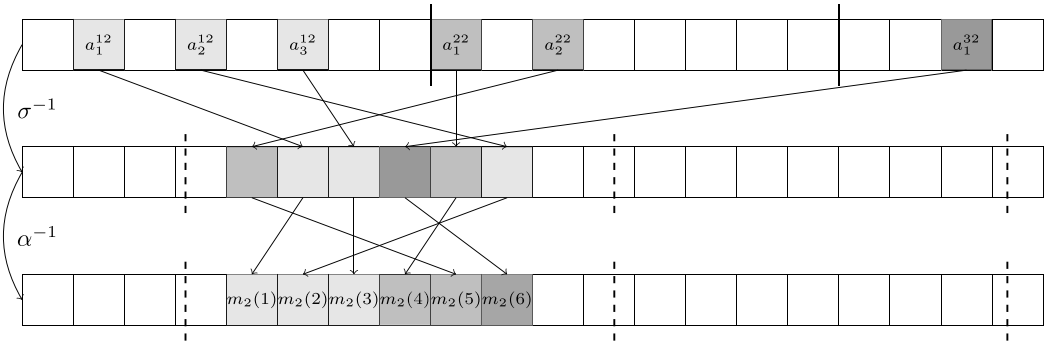


Figure 4. The construction of α^{-1}. The tall solid vertical lines represent the divisions between the intervals $W_{0,1}, \ldots W_{0,3}$, and the tall dashed vertical lines represent the divisions between the intervals $H_1, \ldots H_4$. All arrows indicate that the respective transformations map the specified vertex intervals to vertex intervals by translation.

Let n_{ij}, I_{ij}, v^* be as defined in parts (a) and (b). Let $k_{ij} = \min\{n_{ij}, \lfloor v^*_{ij} kn \rfloor - 1\}$. We will construct α and β so that there are k_{ij} synchronised vertex intervals contained in H_j whose images under $(\sigma \circ \alpha)$ and $(\tau \circ \beta)$ are contained in $W_{0,i}$.

The transformations σ^{-1} and τ^{-1} map approximately the same amount of mass from $W_{0,i}$ to H_j for all $i \in [s_w]$ and $j \in [s_h]$, but the intersection of the image of $W_{0,i}$ and H_j may be very different under the two maps. We design α^{-1} and β^{-1} so that $\alpha^{-1} \circ \sigma^{-1}$ and $\beta^{-1} \circ \tau^{-1}$ both map mass from $W_{0,i}$ to the same subinterval of H_j. Working with the inverse functions allows us to think of α^{-1} and β^{-1} as functions that reorganise the images of $W_{0,i}$ under σ^{-1} and τ^{-1} (respectively) within each interval H_j. We now formally construct α and β by constructing their inverses.

First we construct α^{-1}, as illustrated in Figure 4. There are n_{ij} vertex intervals contained in $W_{0,i}$ that are mapped to vertex intervals in H_j under σ^{-1}. Informally, α^{-1} will rearrange the images of these vertex intervals within H_j by sorting them by their origin interval $W_{0,i}$. Under α^{-1}, the image of vertex intervals originating in $W_{0,1}$ are mapped to the leftmost vertex intervals contained completely in H_j.

Formally, let $a_1^{ij}, a_2^{ij} \ldots a_{k_{ij}}^{ij}$ enumerate k_{ij} of the n_{ij} vertex intervals contained in $W_{0,i}$ that are mapped to vertex intervals in H_j under σ^{-1}. Let $m_j(i)$ be the i^{th} interval of M that is entirely contained in H_j. Define $\alpha \in \mathcal{M}$ so that α^{-1} translates the interval $\sigma^{-1}(a_\ell^{ij})$ to the vertex interval specified as follows:

$$\alpha^{-1}(\sigma^{-1}(a_\ell^{ij})) = m_j\left(\sum_{z=1}^{i-1} k_{zj} + \ell\right),$$

and $H_j \setminus \left(\bigcup_{i=1}^{s_w} \bigcup_{\ell=1}^{k_{ij}} \sigma^{-1}(a_\ell^{ij})\right)$ maps to $H_j \setminus \left(\bigcup_{i=1}^{s_w} \bigcup_{\ell=1}^{k_{ij}} \alpha^{-1}(\sigma^{-1}(a_\ell^{ij}))\right)$ under α^{-1} in any invertible manner. Since $\alpha^{-1}(H_i) = H_i$, α and α^{-1} respect the intervals $H_1, \ldots H_{s_h}$.

Next we construct the map β^{-1}, as illustrated in Figure 5. Recall the definition of I_{ij} described in the construction of $\tau_{v^*} = \tau^{-1}$. Each interval I_{ij} is contained in $W_{0,i}$ and $\tau^{-1}(I_{1j})$, $\tau^{-1}(I_{2j})$, $\ldots \tau^{-1}(I_{s_h j})$ are consecutive intervals (in that order) whose union is H_j. Unlike σ^{-1}, τ^{-1} may not be in $\mathcal{M}_{kn}$, and so the image of vertex intervals under τ^{-1} are not necessarily vertex intervals. Informally, β^{-1} will map the images of vertex intervals under τ^{-1} to vertex intervals in a way that maintains their relative order in H_j.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

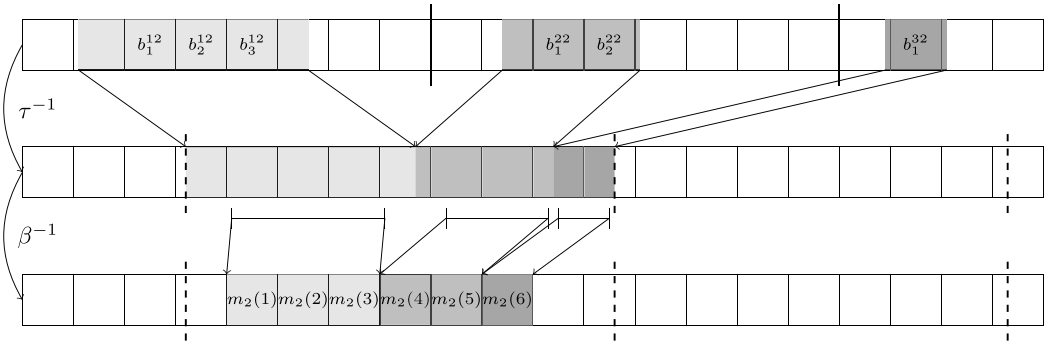


Figure 5. The construction of β^{-1}. The tall solid vertical lines represent the divisions between the intervals $W_{0,1}, \ldots W_{0,3}$, and the tall dashed vertical lines represent the divisions between the intervals $H_1, \ldots H_4$. The arrows corresponding to τ^{-1} illustrate that τ^{-1} map intervals to intervals by translation. The arrows depicting β^{-1} show that β^{-1} maps adjacent intervals of the form $\tau^{-1}(b_\ell^{ij})$ (shown by horizontal line segments) to vertex intervals by translation.

We now formally describe β^{-1}. Since I_{ij} has length v_{ij}^*, there are at least $\lfloor v_{ij}^* kn \rfloor - 1$ vertex intervals contained in $I_{ij} \subseteq W_{0,i}$, all of which are mapped to H_j under τ^{-1}. Let $b_1^{ij}, b_2^{ij} \ldots b_{k_{ij}}^{ij}$ enumerate k_{ij} of these vertex intervals contained in I_{ij}. Define $\beta \in \mathcal{M}$ so that β^{-1} translates the interval (which is not necessarily a vertex interval) $\tau^{-1}(b_\ell^{ij})$ to the vertex interval specified as follows:

$$\beta^{-1}(\tau^{-1}(b_\ell^{ij})) = m_j\left(\sum_{z=1}^{i-1} k_{zj} + \ell\right),$$

and $H_j \setminus \left(\bigcup_{i=1}^{s_w} \bigcup_{\ell=1}^{k_{ij}} \tau^{-1}(b_\ell^{ij})\right)$ maps to $H_j \setminus \left(\bigcup_{i=1}^{s_w} \bigcup_{\ell=1}^{k_{ij}} \beta(\tau^{-1}(b_\ell^{ij}))\right)$ under β^{-1} in any invertible manner. Since $\beta(H_i) = H_i$, β and β^{-1} respect the intervals $H_1, \ldots H_{s_h}$.

Next we construct K, a set of synchronised vertex intervals. Note that for all triples i, j, ℓ with $i \in [s_w]$, $j \in [s_h]$, and $\ell \in [k_{ij}]$, $m_j\left(\sum_{z=1}^{i-1} k_{zj} + \ell\right)$ is a synchronised vertex interval since a_ℓ^{ij} and b_ℓ^{ij} are vertex intervals contained in $W_{0,i}$. Let

$$K = \left\{ m_j\left(\sum_{z=1}^{i-1} k_{zj} + \ell\right) | i \in [s_w], j \in [s_h], \text{ and } \ell \in [k_{ij}] \right\}.$$

Finally, we bound the size of K. Recall that by construction $v_{ij} = n_{ij}/(kn)$ and $\|v - v^*\|_\infty \leq \eta/(4s_w s_h)$. It follows that

$$|n_{ij} - v_{ij}^* kn| \leq \frac{\eta kn}{4s_w s_h}.$$

Since $|\lfloor v_{ij}^* kn \rfloor - 1 - v_{ij}^* kn| \leq 2$, it follows that

$$k_{ij} = \min\{n_{ij}, \lfloor v_{ij}^* kn \rfloor - 1\} \geq n_{ij} - 2 - \frac{\eta kn}{4s_w s_h}.$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

We use this to lower bound the total number of synchronised intervals in K,

$$|K| = \sum_{i=1}^{s_w}\sum_{j=1}^{s_h} k_{ij} \geq \left(\sum_{i=1}^{s_w}\sum_{j=1}^{s_h} n_{ij}\right) - 2s_w s_h - \frac{\eta kn}{4} \geq kn - s_h + 1 - 2s_w s_h - \frac{\eta kn}{4}$$

$$\geq kn\left(1 - \frac{3s_w s_h}{kn} - \frac{\eta}{4}\right) \geq kn\,(1-\eta/2)\,,$$

since $n \geq 12s_h^2 s_w/(k\eta) \geq 12s_h s_w/(k\eta)$ by the assumption of the lemma.

Finally, we address (d). We construct a coupling of f and g so that $(f^\sigma)^\alpha$ and $(g^\tau)^\beta$ agree on sets of the form $\bar{m}\times\bar{m}'$ where $\bar{m},\bar{m}' \in K$ are synchronised intervals. Let v and v' be the vertices in f corresponding to the vertex intervals that are mapped to $\bar{m}$ and $\bar{m}'$ respectively under $\sigma\circ\alpha$. Let u and u' be the vertices in g corresponding to the vertex intervals that are mapped to $\bar{m}$ and $\bar{m}'$ respectively under $\tau\circ\beta$. By construction, v and u correspond to vertex intervals contained in the same interval $W_{0,i}$, and likewise v' and u' correspond to vertex intervals contained in the same interval $W_{0,j}$. Let X_{ab} and Y_{ab} be the indicator random variables for the events that there is an edge between vertices a and b in f and g respectively. Recall that p_{ij} denotes the value W_0 takes on $W_{0,i}\times W_{0,j}$. Since $X_{vv'} \sim Bern(p_{ij})$ and $Y_{uu'} \sim Bern(p_{ij})$, we can couple $X_{vv'}$ and $Y_{uu'}$ exactly, which then guarantees that $(f^\sigma)^\alpha$ and $(g^\tau)^\beta$ agree on the set $\bar{m}\times\bar{m}'$.

Since $(f^\sigma)^\alpha$ and $(g^\tau)^\beta$ agree on the synchronised vertex intervals $\left(\bigcup_{\bar{m}\in K}\bar{m}\right)^2$ and $|\left(\bigcup_{\bar{m}\in K}\bar{m}\right)| \geq 1-\eta/2$, it follows that $d_\square\left((f^\sigma)^\alpha,(g^\tau)^\beta\right)\leq\eta$, as desired. □

Proof of Lemma 47. Fix $\tau\in\mathcal{M}$. Note that τ is invertible and $d_\square(f^\tau,g) = d_\square(f,g^{\tau^{-1}})$. It follows that

$$\begin{aligned}\mathbb{P}_{kn,W_0}\left(d_\square\left(f^{G^\tau_{kn}},h\right)\leq\varepsilon\right) &= \mathbb{P}_{kn,W_0}\left(d_\square\left(f^{G_{kn}},h^{\tau^{-1}}\right)\leq\varepsilon\right)\\ &= \mathbb{P}_{kn,W_0}\left(f^{G_{kn}}\in\left\{g : d_\square\left(g,h^{\tau^{-1}}\right)\leq\varepsilon\right\}\right).\end{aligned}$$

Note that [11, Lemma 5.4] implies that the set $\{g : d_\square(g,h^{\tau^{-1}})\leq\varepsilon\}$ is closed in the weak topology. Applying Theorem 48, we obtain

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{(kn)^2}\log\mathbb{P}_{kn,W_0}\left(d_\square(f^{G^\tau_{kn}},h)\leq\varepsilon\right)\\ &= \limsup_{n\to\infty}\frac{1}{(kn)^2}\log\mathbb{P}_{kn,W_0}\left(f^{G_{kn}}\in\left\{g : d_\square\left(g,h^{\tau^{-1}}\right)\leq\varepsilon\right\}\right)\\ &\leq -\inf_{f\in\{g:d_\square(g,h^{\tau^{-1}})\leq\varepsilon\}} I_{W_0}(f) \leq -\inf_{f:\delta_\square(f,h)\leq\varepsilon} I_{W_0}(f),\end{aligned}$$

where the last line follows from the observation that $\delta_\square(g,h)\leq d_\square(g,h^{\tau^{-1}})$. □

2.4 Proof of Theorem 16

We begin with the following theorem, which is a direct adaptation of [14, Theorem 3.1] to general k-block graphons W_0. As usual, for $\tilde{f}\in\mathcal{W}$ and $\tilde{H}\subseteq\widetilde{\mathcal{W}}$, define $\delta_\square(\tilde{f},\tilde{H}) \triangleq \inf_{\tilde{h}\in\tilde{H}}\delta_\square(\tilde{f},\tilde{h})$.

Theorem 52. *Let W_0 be a uniform k-block graphon. Let $\tilde{F}$ be a closed subset of $\widetilde{\mathcal{W}}$, and let $\tilde{F}^0$ be its interior. Suppose*

$$\inf_{\tilde{h}\in\tilde{F}^0} J_{W_0}(\tilde{h}) = \inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}). \tag{18}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Let $\tilde{F}^\star$ be the subset of $\tilde{F}$ where J_{W_0} is minimised. Then $\tilde{F}^\star$ is non-empty and compact, and

$$\min_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}) = -\lim_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}(\tilde{F}). \tag{19}$$

If $\min_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}) < \infty$, then for all sufficiently large n and all $\varepsilon > 0$, $\tilde{\mathbb{P}}_{kn,W_0}(\tilde{f}^{G_{kn}} \in \tilde{F}) > 0$ and

$$\tilde{\mathbb{P}}_{kn,W_0}\left(\delta_\square(\tilde{f}^{G_{kn}}, \tilde{F}^\star) \geq \varepsilon \middle| \tilde{f}^{G_{kn}} \in \tilde{F}\right) \leq e^{-C(\varepsilon,\tilde{F})(kn)^2},$$

where $C(\varepsilon, \tilde{F})$ is a positive constant depending only on ε and $\tilde{F}$. In particular, if $\tilde{F}^\star$ contains only one element $\tilde{h}^\star$ (and $J_{W_0}(\tilde{h}^\star) < \infty$), then the conditional distribution of $\tilde{f}^{G_{kn}}$ given $\tilde{f}^{G_{kn}} \in \tilde{F}$ converges to the point mass at $\tilde{h}^\star$ as $n \to \infty$.

Proof. Since $\widetilde{\mathcal{W}}$ is compact and $\tilde{F}$ is closed, $\tilde{F}$ is also compact. By Lemma 32, the function J_{W_0} is lower semi-continuous on $\tilde{F}$. Since $\tilde{F}$ is compact, J_{W_0} must attain its minimum on $\tilde{F}$. Therefore, $\tilde{F}^\star$ is non-empty. Moreover, by the lower semi-continuity of J_{W_0}, $\tilde{F}^\star$ is closed, and hence compact. Finally, by Theorem 11,

$$\begin{aligned}
-\inf_{\tilde{h}\in\tilde{F}^0} J_{W_0}(\tilde{h}) &\leq \liminf_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}(\tilde{F}^0) \leq \liminf_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}(\tilde{F}) \\
&\leq \limsup_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}(\tilde{F}) \leq -\inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}) = -\min_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}).
\end{aligned}$$

Combined with (18) this proves (19).

Next, assume that the inf in (18) is finite. This is only compatible with (19) if $\tilde{\mathbb{P}}_{kn,W_0}(\tilde{F^0}) > 0$ for all n larger than some n_0. Fix $\varepsilon > 0$ and let

$$\tilde{F}_\varepsilon \triangleq \left\{\tilde{h} \in \tilde{F} : \delta_\square(\tilde{h}, \tilde{F}^\star) \geq \varepsilon\right\},$$

which is also a closed subset. Observe that $\tilde{F}_\varepsilon \cap \tilde{F}^\star = \emptyset$. Then

$$\tilde{\mathbb{P}}_{kn,W_0}\left(\delta_\square(\tilde{f}^{G_{kn}}, \tilde{F}^\star) \geq \varepsilon | \tilde{f}^{G_{kn}} \in \tilde{F}\right) = \frac{\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}_\varepsilon\right)}{\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}\right)} \leq \frac{\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}_\varepsilon\right)}{\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}^0\right)}.$$

Using Theorem 11 again, this shows that,

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}\left(\delta_\square(\tilde{f}^{G_{kn}}, \tilde{F}^\star) \geq \varepsilon | \tilde{f}^{G_{kn}} \in \tilde{F}\right) \\
&\leq \limsup_{n\to\infty}\frac{1}{(kn)^2}\left[\log\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}_\varepsilon\right) - \log\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}^0\right)\right] \\
&= \limsup_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}_\varepsilon\right) - \liminf_{n\to\infty}\frac{1}{(kn)^2}\log\tilde{\mathbb{P}}_{kn,W_0}\left(\tilde{f}^{G_{kn}} \in \tilde{F}^0\right) \\
&\leq \inf_{\tilde{h}\in\tilde{F}^0} J_{W_0}(\tilde{h}) - \inf_{\tilde{h}\in\tilde{F}_\varepsilon} J_{W_0}(\tilde{h}) \\
&= \inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}) - \inf_{\tilde{h}\in\tilde{F}_\varepsilon} J_{W_0}(\tilde{h}).
\end{aligned}$$

It now suffices to show that $\inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}) < \inf_{\tilde{h}\in\tilde{F}_\varepsilon} J_{W_0}(\tilde{h})$. Clearly, $\inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}) \leq \inf_{\tilde{h}\in\tilde{F}_\varepsilon} J_{W_0}(\tilde{h})$. Suppose that equality holds. The compactness of $\tilde{F}_\varepsilon$ and the lower semi-continuity of J_{W_0}

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

(Lemma 32) imply that there exists $\tilde{g} \in \tilde{F}_\varepsilon$ that attains the infimum. It follows that $J_{W_0}(\tilde{g}) = \inf_{\tilde{h}\in\tilde{F}_\varepsilon} J_{W_0}(\tilde{h}) = \inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h})$. But then $\tilde{g} \in \tilde{F}^\star$, and so $\tilde{F}_\varepsilon \cap \tilde{F}^\star \neq \emptyset$, which is a contradiction. □

Proof of Theorem 16. We will prove the theorem by establishing condition (18) in Theorem 52. We first note that the continuity of $\phi_\tau(W_0, \cdot)$ at t excludes the trivial case $\tau(\mathcal{W}) = \{t\}$, since then $\tau(\widetilde{\mathcal{W}}) = \{t\}$ as well, which shows that $\phi_\tau(W_0, \cdot)$ jumps from a finite constant to ∞ at t. Therefore, we may assume the graph parameter τ is not constant.

Next we recall that $\phi_\tau(W_0, t) = \min\{J_{W_0}(\tilde{f}) : \tilde{f} \in \tilde{F}\}$, where $\tilde{F} = \{\tilde{f} \in \widetilde{\mathcal{W}} : \tau(\tilde{f}) \geq t\}$. By the continuity of τ, $\tilde{F}$ is closed. Also by the continuity of τ, the open set $\tilde{O} = \{\tilde{f} \in \widetilde{\mathcal{W}} : \tau(\tilde{f}) > t\} \subseteq \tilde{F}^0$ contains $\tilde{F}_\varepsilon = \{\tilde{f} \in \widetilde{\mathcal{W}} : \tau(\tilde{f}) \geq t + \varepsilon\}$ for all $\varepsilon > 0$. As a consequence,

$$\phi_\tau(W_0, t) = \inf_{\tilde{h}\in\tilde{F}} J_{W_0}(\tilde{h}) \leq \inf_{\tilde{h}\in\tilde{F}^0} J_{W_0}(\tilde{h}) \leq \inf_{\tilde{h}\in\tilde{O}} J_{W_0}(\tilde{h}) \leq \inf_{\tilde{h}\in\tilde{F}_\varepsilon} J_{W_0}(\tilde{h}) = \phi_\tau(W_0, t + \varepsilon).$$

Sending $\varepsilon \to 0$ and using the continuity of $\phi_\tau(W_0, \cdot)$ at t, we see that the first inequality is saturated, proving (18).

Recall that $\mathbb{P}_{kn,W_0}$ is supported in $\mathcal{W}_\Omega$. The proof is completed by noting that $\phi_\tau(W_0, t) < \infty$ if and only if $t \leq t_{\max}$. □

3. ϕ_τ: Monotonicity, continuity, and examples

In this section, we establish some analytical properties of the function ϕ_τ, which will be critical for our discussion of symmetry/symmetry breaking in the subsequent sections. Section 3.1 collects some preliminary properties of homomorphism densities and the cut distance. In Section 3.2, we introduce the 'sufficient increase property', which guarantees the continuity of ϕ_τ. Further, we establish that homomorphism densities satisfy this property, and the operator norm satisfies this property under additional assumptions. Finally, Section 3.3 establishes an alternative variational representation of ϕ_τ at points of continuity. Using this representation, we identify a class of parameters τ such that ϕ_τ is strictly increasing.

3.1 Preliminaries

In subsequent sections, it will be necessary to express the homomorphism density as the sum of interval labelled homomorphisms and identify interval labelled homomorphisms that are always zero on $\mathcal{W}_\Omega$. Given $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$ and $W_0 \in \mathcal{B}^\gamma$, let $I_1, I_2, \ldots I_m$ be the intervals of γ, i.e. $I_1 = [0, \gamma_1]$ and $I_j = (\sum_{i=1}^{j-1} \gamma_i, \sum_{i=1}^{j} \gamma_i]$ for $j \geq 2$. When we proved Theorem 16, we used k to denote the total number of intervals and assumed that all intervals have the same length. Here the intervals need not be the same length – to emphasise this we now use m to denote the total number of intervals. The blocks of W_0 have the form $I_i \times I_j$ for $i, j \in [m]$. Let $v = |V(H)|$ be the number of vertices in H, and let $Y \in [m]^v$ be a vector of vertex interval indices. Define the *interval-labelled homomorphism density* as

$$t(H, g, Y) = \int_{x_1\in I_{Y_1}} \int_{x_2\in I_{Y_2}} \cdots \int_{x_v\in I_{Y_v}} \prod_{\{i,j\}\in E(H)} g(x_i, x_j)\, dx_v \ldots dx_2 dx_1.$$

In other words, $t(H, g, Y)$ accounts for the homomorphisms in which the j^{th} vertex is in I_{Y_j} for all $j \in [v]$ and so

$$t(H, g) = \sum_{Y\in[m]^v} t(H, g, Y).$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

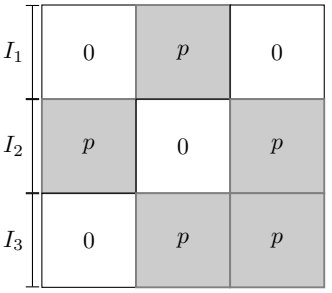


Figure 6. For W_0 the graphon above and H a triangle, the block $I_1 \times I_2$ is not relevant. Increasing the number of edges between I_1 and I_2 will not increase the number of triangles in the graph since any triangle has at least two vertices coming only from I_3.

Next, we define *relevant blocks* to be the blocks whose values may affect the homomorphism density of a graphon in $\mathcal{W}_\Omega$. Increasing the value of a graphon g in $\mathcal{W}_\Omega$ on a relevant block has the potential to increase $t(H, g)$. Figure 6 gives an example of a block that is not relevant.

Definition 53. *Fix a finite graph H and $\mathcal{W}_\Omega$. We say an interval labelling vector Y is irrelevant with respect to W_0 if there exists $\{i, j\} \in E(H)$ such that the block $I_{Y_i} \times I_{Y_j}$ takes value zero on W_0. Equivalently, Y is irrelevant if $t(H, g, Y) = 0$ for all $g \in \mathcal{W}_\Omega$. We say Y is relevant otherwise.*

We say a block $I_a \times I_b$ contributes to the interval-labelled homomorphism density $t(H, g, Y)$ if $Y_i = a$ and $Y_j = b$ for some $\{i, j\} \in E(H)$. We say the block $I_a \times I_b$ is relevant if $I_a \times I_b$ contributes to some $t(H, g, Y)$ with Y relevant. Let $R \subseteq [0, 1]^2$ be the union of all relevant blocks.

Note that a block $I_a \times I_b$ is relevant if and only if $p_{ab} > 0$ and $t(H, W_0)$ strictly decreases when p_{ab} is lowered.

Our next result establishes that if the cut distance between two graphons is at least a constant, one can find a region where the values on the graphons differ by at least a constant. This result will be crucially used to establish the 'sufficient increase property' in this section. In our subsequent discussion, we will use this result to establish the existence of nearby graphons with lower entropy.

Lemma 54. *Let $f, g \in \mathcal{W}$. Let $S^+_\beta = \{(x, y) \in [0, 1]^2 : f(x, y) - g(x, y) \geq \beta\}$ and $S^-_\beta = \{(x, y) \in [0, 1]^2 : g(x, y) - f(x, y) \geq \beta\}$.*

1. *If $f \geq g$ pointwise and $d_\square(f, g) \geq \varepsilon$, then $|S^+_{\varepsilon/2}| > \varepsilon/2$.*
2. *If $d_\square(f, g) > \varepsilon$, then $|S^+_{\varepsilon/4}| \geq \varepsilon/4$ or $|S^-_{\varepsilon/4}| \geq \varepsilon/4$.*

Proof. Suppose $f \geq g$ pointwise and $d_\square(f, g) \geq \varepsilon$. Since $f \geq g$ pointwise, $d_\square(f, g) = \|f - g\|_1$. It follows that

$$\varepsilon \leq d_\square(f, g) = \|f - g\|_1 = \int_{[0,1]^2} f - g \leq |S^+_{\varepsilon/2}| + \frac{\varepsilon}{2}(1 - |S^+_{\varepsilon/2}|),$$

and so $|S^+_{\varepsilon/2}| \geq \varepsilon/(2 - \varepsilon) > \varepsilon/2$.

Next suppose $d_\square(f, g) > \varepsilon$ with no additional assumptions on $f, g \in \mathcal{W}$. Let $S^+ = \{(x, y) \in [0, 1]^2 : f(x, y) \geq g(x, y)\}$ and $S^- = \{(x, y) \in [0, 1]^2 : f(x, y) < g(x, y)\}$. Since $d_\square(f, g) > \varepsilon$, there exists $A, B \subseteq [0, 1]$ such that

$$\left| \int_{A \times B} f - g \right| \geq \varepsilon.$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Observe

$$\left|\int_{A\times B} f-g\right| \le \int_{A\times B} |f-g| = \int_{(A\times B)\cap S^+} f-g + \int_{(A\times B)\cap S^-} g-f.$$

It follows that $\int_{(A\times B)\cap S^+} f-g \ge \varepsilon/2$ or $\int_{(A\times B)\cap S^-} g-f \ge \varepsilon/2$. In the first case, let $\bar{f}(x,y) = f(x,y)$ for $(x,y) \in (A\times B)\cap S^+$, and otherwise set $\bar{f}(x,y) = g(x,y)$. Then $\bar{f} \ge g$ pointwise, and $d_\square(\bar{f}, g) \ge \frac{\varepsilon}{2}$. By the first statement, we obtain $|S^+_{\frac{\varepsilon}{4}}| \ge \frac{\varepsilon}{4}$. A similar argument applies to the second case. □

3.2 Establishing the continuity of ϕ_τ

In this subsection we establish that ϕ_τ is continuous for certain graph parameters τ. When τ is clear from context, we let $\phi(t) = \phi_\tau(W_0, t)$.

Lemma 55. *Let H be a finite graph, let $\tau = t(H,\cdot)$, let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and $W_0 \in \mathcal{B}^\gamma$. Then ϕ is continuous on $\mathbb{R} \setminus \{t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)\}$.*

Lemma 56. *Let $\tau(g) = \|g\|_{op}$, let W_0 be a two-block bipartite graphon with $W_0 \in \mathcal{B}^{(\gamma,1-\gamma)}$ where $\gamma \in (0,1)\cap\mathbb{Q}$. Then ϕ is continuous on $\mathbb{R} \setminus \{t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)\}$.*

In order to establish the above lemmas, we describe the sufficient increase property of τ and W_0, and show that having this property guarantees that ϕ_τ is continuous. Then we show that homomorphism densities have this property (with any block constant base graphon W_0), and the operator norm has this property when W_0 is a two block bipartite graphon.

Definition 57. *We say that τ has the sufficient increase property on $\mathcal{W}_\Omega$ if the following is true. Fix any $\eta > 0$ and $t^\tau_{\min}(\widetilde{\mathcal{W}}_\Omega) \le t < t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$. Then there exist $\alpha = \alpha(t, t_{\max}, \eta)$, $\beta = \beta(t, t_{\max}, \eta) > 0$, such that the following holds for all $g \in \mathcal{W}_\Omega$. If $\tau(g) \ge t - \alpha$, then there exists $g^* \in \{f : \|f-g\|_\infty \le \eta\} \cap \mathcal{W}_\Omega$ such that $\tau(g^*) \ge t + \beta$.*

Lemma 58. *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$, and let τ be a continuous graph parameter that has the sufficient increase property on $\mathcal{W}_\Omega$. Then ϕ is continuous on $\mathbb{R} \setminus \{t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)\}$.*

In [34], Lubetzky and Zhao studied the variational problem (6) when W_0 is a constant graphon. They defined a 'nice graph parameter' as a graph parameter τ that is (i) continuous with respect to $\delta_\square$ and (ii) has the property that every local extremum of τ with respect to L_∞ is necessarily a global extremum. They showed that for any nice graph parameter τ, ϕ_τ is continuous in the setting where the base graphon W_0 is constant. Their proof technique cannot be directly adapted to the setting where W_0 is a block constant graphon with a zero or one block. When W_0 take values zero or one, the entropy function I_{W_0} can be infinite, creating a technical hurdle. In particular, it is not clear how to establish right continuity of ϕ_τ for arbitrary nice graph parameters. We instead use Definition 57 as a sufficient condition for the right continuity of ϕ_τ.

Before proving Lemma 58, we establish the left continuity of ϕ_τ without any assumptions on the block graphon W_0 or the continuous graph parameter τ.

Lemma 59. *Let τ be a continuous graph parameter, and let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and $W_0 \in \mathcal{B}^\gamma$. Then ϕ is left-continuous.*

Proof. We first note that we may assume that $t \le t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$, since $\phi = \infty$ and hence constant above $t_{\max}$. Let $t_n \nearrow t$. Since ϕ is non-decreasing in t, the sequence $\phi(t_n)$ has a limit, and $\lim_{n\to\infty}\phi(t_n) \le \phi(t)$. To prove an upper bound on $\phi(t)$, recall the definition (5) of $\phi(t)$ as a minimum. For each $k \ge 1$, there exists $\tilde{g}_k$ such that $\tau(\tilde{g}_k) \ge t_k$ and $J_{W_0}(\tilde{g}_k) = \phi(t_k)$. By the compactness of $\widetilde{\mathcal{W}}$, there exists a convergent subsequence $\tilde{g}_{k_j}$ such that $\delta_\square(\tilde{g}_{k_j}, \tilde{g}) \to 0$ for some $\tilde{g} \in \widetilde{\mathcal{W}}$. Since $\tau(\tilde{g}_{k_j}) \ge t_{k_j}$ and $t_{k_j} \nearrow t$, it follows that $\tau(\tilde{g}) \ge t$, and thus $\phi(t) \le J_{W_0}(\tilde{g})$. Combined with the lower

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

semi-continuity of J_{W_0} (Lemma 32), we get

$$\phi(t) \le J_{W_0}(\tilde{g}) \le \liminf_{j\to\infty} J_{W_0}(\tilde{g}_{k_j}) = \liminf_{j\to\infty} \phi(t_{k_j}) = \lim_{n\to\infty} \phi(t_n).$$

□

Remark 60. Since ϕ is left-continuous (Lemma 59) and non-decreasing, ϕ can have at most countably many points of discontinuity.

We now prove Lemma 58, which establishes the continuity of ϕ_τ when τ has the sufficient increase property.

Proof of Lemma 58. By Lemma 59, it suffices to establish the right-continuity of ϕ at t. By assumption, $t \neq t_{\max}$. Since ϕ is constant on $(-\infty, t_{\min}]$ and $(t_{\max}, \infty)$ (where it is 0 and ∞, respectively), we may assume that $t_{\min} \le t < t_{\max}$. Consider a sequence $t_n \searrow t$, and an arbitrary $\varepsilon > 0$. We need to show that there exists n sufficiently large such that $\phi(t_n) \le \phi(t) + \varepsilon$.

Let $\eta > 0$ be such that if $f, g \in \mathcal{W}_\Omega$ and $\|f - g\|_\infty \le \eta$, then $|I_{W_0}(f) - I_{W_0}(g)| < \varepsilon$; Proposition 35 guarantees the existence of such an η. Let $\tilde{g} \in \widetilde{\mathcal{W}}_\Omega$ be such that $\tau(\tilde{g}) \ge t$ and $\phi(t) = J_{W_0}(\tilde{g})$. By definition of $J_{W_0}(\,\cdot\,)$ there exists a sequence $f_k \in \mathcal{W}_\Omega$ such that

$$I_{W_0}(f_k) \to J_{W_0}(\tilde{g}) \quad \text{and} \quad \delta_\square(f_k, \tilde{g}) \to 0.$$

Since τ has the sufficient increase property on $\mathcal{W}_\Omega$, there exist $\alpha, \beta > 0$ such that

$$\tau(h) \ge t - \alpha \implies \exists\, h^* \text{ with } \|h^* - h\|_\infty \le \eta \text{ and } \tau(h^*) \ge t + \beta.$$

Since τ is continuous in $(\widetilde{\mathcal{W}}_\Omega, \delta_\square)$ and $\tau(g) \ge t$, there exists k_0 sufficiently large such that for all $k \ge k_0$, $\tau(f_k) \ge t - \alpha$. Thus, for all $k \ge k_0$, there exists f'_k such that

$$\tau(f'_k) \ge t + \beta \quad \text{and} \quad \|f'_k - f_k\|_\infty \le \eta.$$

The choice of η implies

$$|I_{W_0}(f'_k) - I_{W_0}(f_k)| \le \varepsilon.$$

By compactness of $(\widetilde{\mathcal{W}}_\Omega, \delta_\square)$, there exists a convergent subsequence such that $\tilde{f}'_{k_j} \to \tilde{h}$ for some $\tilde{h} \in \widetilde{\mathcal{W}}_\Omega$. Since τ is continuous with respect to $\delta_\square$, $\tau(\tilde{f}'_{k_j}) \to \tau(\tilde{h})$, and so $\tau(\tilde{h}) \ge t + \beta$. It follows that

$$\phi(t+\beta) \le J_{W_0}(\tilde{h}) \le \liminf_{j\to\infty} I_{W_0}(f'_{k_j}) \le \liminf_{j\to\infty} I_{W_0}(f_{k_j}) + \varepsilon = J_{W_0}(\tilde{g}) + \varepsilon = \phi(t) + \varepsilon.$$

Taking n sufficiently large such that $t_n \le t + \beta$ and noting $\phi(t_n) \le \phi(t+\beta)$ yields the desired statement. □

Next we establish that homomorphism densities have the sufficient increase property. To this end, we introduce the following graphon $g^{+\eta}$,

$$g^{+\eta}(x,y) = \begin{cases} g(x,y) & (x,y) \notin \Omega \\ \min\{g(x,y) + \eta, 1\} & \text{otherwise} \end{cases} \tag{20}$$

and note the following fact.

Fact 61. *Let $\ell, u, w, z \in \mathbb{R}$ with $u \ge 0$. Suppose that $\ell \ge uz$ and $\ell \ge w - z$. Then $\ell \ge \frac{uw}{u+1}$.*

Proof. Note that for all $z \in \mathbb{R}$, $uz \ge \frac{uw}{u+1}$ or $w - z \ge \frac{uw}{u+1}$. The fact follows directly. □

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Lemma 62. *Let $\tau = t(H, \cdot)$ where H is a finite graph, let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and $W_0 \in \mathcal{B}^\gamma$, and let $t_{\max}$ denote $t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$. Fix $\eta \in (0, 1]$. Then there exists $c = c(\gamma, H, \eta) \in (0, 1]$ such that $\tau(g^{+\eta}) \geq \tau(g) + c(t_{\max} - \tau(g))^2$ for all $g \in \mathcal{W}_\Omega$.*

Proof. Define $g^{\max}$ as follows

$$g^{\max}(x, y) = \begin{cases} g(x, y) & (x, y) \in ([0, 1]^2 \setminus R) \cup \Omega^c \\ 1 & \text{otherwise,} \end{cases}$$

where R is the union of relevant blocks, as defined above. Note that since $g = W_0$ on $[0, 1]^2 \setminus \Omega$, $g^{\max} \in \mathcal{W}_\Omega$. Also note that $\tau(g^{\max}) = \max_{f \in \mathcal{W}_\Omega} \tau(f) = t_{\max}$.

Let $e_H = |E(H)|$ and $v = |V(H)|$ be the numbers of edges and vertices in H. Let

$$d = \frac{t(H, g^{\max}) - t(H, g)}{e_H} = \frac{t_{\max} - \tau(g)}{e_H}.$$

Since the statement of the lemma is trivial if $\tau(g) = t_{\max}$ we may assume w.l.o.g. that $d > 0$. The Counting Lemma [8, Theorem 3.7] [32, Lemma 10.23] implies that $\delta_\square(g^{\max}, g) \geq d$, and so

$$d_\square(g, g^{\max}) \geq \delta_\square(g, g^{\max}) \geq d.$$

Let $S = \{(x, y) \in [0, 1]^2 : g^{\max} - g \geq d/2\}$. Since $g^{\max} \geq g$ pointwise, Lemma 54 implies that $|S| \geq d/2$. Let $\eta' \triangleq \min\{\eta, d/2\}$. It follows that $g^{+\eta} - g \geq \eta'$ on S. By construction, $S \subseteq R$. Recall that m denotes the number of blocks in W_0. Therefore, there are at most m^2 relevant blocks of $[0, 1]^2$ of the form $I_i \times I_j$ for $i, j \in [m]$. Thus, there exists $a, b \in [m]$ such that $I_a \times I_b$ is relevant and $|(I_a \times I_b) \cap S| \geq d/(2m^2)$.

It suffices to show that increasing g to $g^{+\eta}$ on $(I_a \times I_b) \cap S$ yields a constant increase in the homomorphism density. Since $I_a \times I_b$ is a relevant block, there exists a relevant $Y \in [m]^v$ such that $Y_p = a$, $Y_q = b$ for some $\{p, q\} \in E(H)$.

Define

$$Z_{\overline{S}}(H, g, Y) = \int_{x_1 \in I_{Y_1}} \cdots \int_{x_v \in I_{Y_v}} \prod_{\{i,j\} \in E(H)} g(x_i, x_j) \mathbb{1}\{(x_p, x_q) \notin S\} \, dx_v \ldots dx_1.$$

$$Z_S(H, g, Y) = \int_{x_1 \in I_{Y_1}} \cdots \int_{x_v \in I_{Y_v}} \prod_{\{i,j\} \in E(H)} g(x_i, x_j) \mathbb{1}\{(x_p, x_q) \in S\} \, dx_v \ldots dx_1.$$

In other words, $Z_S(H, g, Y)$ accounts for the homomorphisms in which the $\{p, q\}$ edge lies in S, and $Z_{\overline{S}}(H, g, Y)$ accounts for the homomorphisms in which the $\{p, q\}$ edge does not lie in S. Therefore,

$$t(H, g, Y) = Z_{\overline{S}}(H, g, Y) + Z_S(H, g, Y).$$

Since $g^{+\eta} \geq g$ pointwise, $Z_{\overline{S}}(H, g^{+\eta}, Y) \geq Z_{\overline{S}}(H, g, Y)$. Next we derive two lower bounds on $Z_S(H, g^{+\eta}, Y)$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

First observe

$$\begin{aligned}
Z_S(H, g^{+\eta}, Y) &= \int_{x_1 \in I_{Y_1}} \cdots \int_{x_v \in I_{Y_v}} \prod_{\{i,j\} \in E(H)} g^{+\eta}(x_i, x_j) \mathbb{1}\{(x_p, x_q) \in S\}\, dx_v \ldots dx_1 \\
&= \int_{x_1 \in I_{Y_1}} \cdots \int_{x_v \in I_{Y_v}} g^{+\eta}(x_p, x_q) \prod_{\substack{\{i,j\} \in E(H) \\ \{i,j\} \neq \{p,q\}}} g^{+\eta}(x_i, x_j) \mathbb{1}\{(x_p, x_q) \in S\}\, dx_v \ldots dx_1 \\
&\geq \int_{x_1 \in I_{Y_1}} \cdots \int_{x_v \in I_{Y_v}} (g(x_p, x_q) + \eta') \prod_{\substack{\{i,j\} \in E(H) \\ \{i,j\} \neq \{p,q\}}} g(x_i, x_j) \mathbb{1}\{(x_p, x_q) \in S\}\, dx_v \ldots dx_1 \\
&\geq Z_S(H, g, Y)(1 + \eta').
\end{aligned}$$

The final inequality follows from noting that $g + \eta' \geq (1 + \eta')g$. Our goal is to lower bound the difference $t(H, g^{+\eta}) - t(H, g)$ by a constant. The above computation implies that $t(H, g^{+\eta}) - t(H, g) \geq \eta' Z_S(H, g, Y)$. This lower bound may not be sufficient if $Z_S(H, g, Y)$ is too small. We derive another lower bound for this case.

Let $\beta = \min_{j \in [m]} |I_j|$. Observe that for all $\{i, j\} \in E(H)$, $g^{+\eta}(x_i, x_j) \geq \eta$ when $x_i \in I_{Y_i}$ and $x_j \in I_{Y_j}$, as Y is relevant. It follows that

$$Z_S(H, g^{+\eta}, Y) \geq \eta^{e_H} |S \cap (I_a \times I_b)| \prod_{j \in [m] \setminus \{p,q\}} |I_{Y_j}| \geq \eta^{e_H} |S \cap (I_a \times I_b)| \beta^{v-2}.$$

We have shown that

$$Z_S(H, g^{+\eta}, Y) - Z_S(H, g, Y) \geq \eta' Z_S(H, g, Y)$$

and

$$Z_S(H, g^{+\eta}, Y) - Z_S(H, g, Y) \geq \eta^{e_H} |S \cap (I_a \times I_b)| \beta^{v-2} - Z_S(H, g, Y).$$

Applying Fact 61 with $u = \eta'$, $w = \eta^{e_H} |S \cap (I_a \times I_b)| \beta^{v-2}$, $z = Z_S(H, g, Y)$, and $\ell = Z_S(H, g^{+\eta}, Y) - Z_S(H, g, Y)$, we obtain

$$Z_S(H, g^{+\eta}, Y) - Z_S(H, g, Y) \geq \frac{\eta'(\eta^{e_H} |S \cap (I_a \times I_b)| \beta^{v-2})}{\eta' + 1}.$$

We now simplify our lower bound using the facts that $|S \cap (I_a \times I_b)| \geq d/(2m^2)$, $\eta'/(1 + \eta') \geq \eta'/2$, $\eta' \geq \eta d/2$, and $d = (t_{\max} - \tau(g))/e_H$, obtaining

$$\frac{\eta'(\eta^{e_H} |S \cap (I_a \times I_b)| \beta^{v-2})}{\eta' + 1} \geq \frac{\eta^{e_H+1} d^2 \beta^{v-2}}{8m^2} = \frac{\eta^{e_H+1} (t_{\max} - \tau(g))^2 \beta^{v-2}}{8 e_H^2 m^2}.$$

Set $c = \frac{\eta^{e_H+1} \beta^{v-2}}{8 e_H^2 m^2}$. Note that c is a function only of H, γ, and η. It follows that

$$\begin{aligned}
t(H, g^{+\eta}, Y) &= Z_{\overline{S}}(H, g^{+\eta}, Y) + Z_S(H, g^{+\eta}, Y) \\
&\geq Z_{\overline{S}}(H, g, Y) + Z_S(H, g, Y) + c(t_{\max} - \tau(g))^2 \\
&= t(H, g, Y) + c(t_{\max} - \tau(g))^2.
\end{aligned}$$

Thus, $t(H, g^{+\eta}) \geq t(H, g) + c(t_{\max} - \tau(g))^2$. □

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Next we establish that ϕ_τ is continuous when τ is a homomorphism density by using Lemma 62 to show that homomorphism densities have the sufficient increase property.

Proof of Lemma 55. By Lemma 58 it suffices to show that τ has the sufficient increase property. Fix η and t. Let $\alpha = \beta = \frac{c}{8}(t_{\max} - t)^2$, where $c \in (0, 1]$ is from Lemma 62. If $\tau(g) > t + \frac{1}{2}(t_{\max} - t)$, we can choose $g^* = g$. Next, suppose $\tau(g) \le t + \frac{1}{2}(t_{\max} - t)$. But then $t_{\max} - \tau(g) \ge \frac{1}{2}(t_{\max} - t)$ and by Lemma 62, $\tau(g^{+\eta}) \ge \tau(g) + \frac{c}{4}(t_{\max} - t)^2$. The assumption $\tau(g) \ge t - \alpha$ then implies

$$\tau(g^{+\eta}) \ge t + \frac{c}{8}(t_{\max} - t)^2 = t + \beta.$$

Noting that $\beta \le \frac{1}{2}(t_{\max} - t)^2 \le \frac{1}{2}(t_{\max} - t)$ completes the proof. □

Next we show that the operator norm has the sufficient increase property. We begin with the following lemma.

Lemma 63. *Let $\tau(g) = \|g\|_{op}$ and let $W_0 \in \mathcal{B}^{(\gamma,1-\gamma)}$ where $\gamma \in (0,1) \cap \mathbb{Q}$ be a bipartite graphon. Let $g^{+\eta}$ be as defined in (20). Fix any $\eta \in (0, 1]$ and $g \in \mathcal{W}_\Omega$ such that $\tau(g) < t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$. Then*

$$\tau(g^{+\eta})^2 \ge \max\left\{\eta^2 t^2_{\max}, \left(1 + \frac{\eta^4}{2^{25}}\left(t_{\max} - \tau(g)\right)^{20}\right)\tau(g)^2\right\}.$$

Proof. Recalling the definition of a bipartite graphon f^γ_p from Section 1.6, we note that except for the trivial case $p \in \{0, 1\}$ (in which case $\mathcal{W}_\Omega = \{W_0\}$ and $\tau(g) = t_{\max}$ for all $g \in \mathcal{W}_\Omega$), the set Ω is $[0, \gamma] \times (\gamma, 1] \cup (\gamma, 1] \times [0, \gamma]$. Let $h^{\max} = f^\gamma_1$ be the graphon that takes value 1 on Ω and agrees with W_0 on Ω^c (where both are 0). Note that if $f \le g$ pointwise, then $\|f\|_{\mathrm{op}} \le \|g\|_{\mathrm{op}}$. It follows that $\tau(h^{\max}) = t_{\max}$. The graphon $h^{\max}$ satisfies $f \le h^{\max}$ for every $f \in \mathcal{W}_\Omega$.

To prove the first lower bound, we note that $g^{+\eta} \ge f^\gamma_\eta$ pointwise, implying that $\|g^{+\eta}\|_{\mathrm{op}} \ge \|f^\gamma_\eta\|_{\mathrm{op}} = \eta\|f^\gamma_1\|_{\mathrm{op}} = \eta t_{\max}$.

To prove the second lower bound, we note that it follows from the proof of [34, Lemma 3.6] that for $f, g \in \mathcal{W}$

$$\frac{(\|f\|_{\mathrm{op}} - \|g\|_{\mathrm{op}})^4}{4} \le \delta_\square(f, g).$$

Let $d = \frac{(t_{\max} - \|g\|_{\mathrm{op}})^4}{4}$. It follows that

$$d \le \delta_\square(h^{\max}, g) \le d_\square(h^{\max}, g).$$

Next, let $\eta' = \min\{\eta, d/2\}$ and $S = \{(x, y) \in \Omega : h^{\max}(x, y) - g(x, y) \ge \eta'\}$. Since $h^{\max} = g$ on $[0, 1]^2 \setminus \Omega$ and $h^{\max} \ge g$ pointwise, Lemma 54 implies that $|S| \ge d/2$.

Since T_g is a self-adjoint compact nonnegative linear operator, there exists $u \in L_2([0, 1])$ such that $u(x) \ge 0$, $\|u\|_2 = 1$ and $T_g u(x) = \|g\|_{\mathrm{op}} u(x)$ for almost all $x \in [0, 1]$ [11, Proposition 2.11]. Let $P \subseteq [0, 1]$ be a measure one subset where this proposition holds. We will derive a lower bound on $\tau(g^{+\eta})^2$ by showing that for some $c > 0$,

$$T_{g^{+\eta}} u(x) \ge T_g u(x) + c, \tag{21}$$

for x in some subset of $[0, 1]$. The construction of this subset will depend on S and u. For ease of notation, we let $\varepsilon = \eta/2$. Define

$$A_\varepsilon = \{x \in [0, 1] : u(x) \ge \varepsilon\} \quad \text{and} \quad A^c_\varepsilon = [0, 1] \setminus A_\varepsilon.$$

We will consider two cases that depend on the size of A^c_ε. In each case we find a subset of $[0, 1]$ and $c > 0$ satisfying (21).

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Before proceeding to the cases, we establish a useful property of u. Define

$$u_1 = \int_0^\gamma u(x)dx \quad \text{and} \quad u_2 = \int_\gamma^1 u(x)dx.$$

Since $\|u\|_2 = 1$, there exists a subset of $[0, 1]$ with positive measure where $u \geq 1$ on the subset. Let z be an element of the intersection of this subset with P. We assume that $z \in [0, \gamma]$. After completing the proof under this assumption, we will discuss how a similar argument applies when $z \in (\gamma, 1]$. Since $z \in [0, \gamma]$, g is zero on $\{z\} \times [0, \gamma]$ and so

$$\|g\|_{\text{op}} \leq \|g\|_{\text{op}} u(z) = T_g u(z) = \int_0^1 g(z, y)u(y)dy = \int_\gamma^1 g(z, y)u(y)dy \leq u_2. \tag{22}$$

Case 1: $|A_\varepsilon^c \cap [0, \gamma]| \geq d/16$. Let $x \in A_\varepsilon^c \cap [0, \gamma] \cap P$. Observe

$$T_g u(x) = \|g\|_{\text{op}} u(x) \leq \|g\|_{\text{op}} \varepsilon = \frac{\eta \|g\|_{\text{op}}}{2}.$$

Note that $g^{+\eta}(x, y) \geq \eta$ for all $y \in (\gamma, 1]$ and $g^{+\eta}(x, y) = 0$ for all $y \in [0, \gamma]$ by construction. It follows that

$$T_{g^{+\eta}} u(x) = \int_\gamma^1 g^{+\eta}(x, y)u(y)dy \geq \eta u_2 \geq \eta \|g\|_{\text{op}}.$$

Thus for all $x \in A_\varepsilon^c \cap [0, \gamma] \cap P$

$$T_{g^{+\eta}} u(x) - T_g u(x) \geq \frac{\eta \|g\|_{\text{op}}}{2}.$$

Observe that

$$\begin{aligned}
\tau(g^{+\eta})^2 = \|g^{+\eta}\|_{\text{op}}^2 \geq \|T_{g^{+\eta}} u\|_2^2 &= \int_0^1 (T_{g^{+\eta}} u(x))^2 dx \\
&\geq \int_{A_\varepsilon^c \cap [0,\gamma]} \left(T_g u(x) + \frac{\eta \|g\|_{\text{op}}}{2} \right)^2 dx + \int_{(A_\varepsilon^c \cap [0,\gamma])^c} T_g u(x)^2 dx \\
&\geq \|T_g u\|_2^2 + |A_\varepsilon^c \cap [0, \gamma]| \frac{\eta^2 \|g\|_{\text{op}}^2}{4} \geq \tau(g)^2 + \frac{d\eta^2 \|g\|_{\text{op}}^2}{64}.
\end{aligned}$$

Case 2: $|A_\varepsilon^c \cap [0, \gamma]| < d/16$. Recall $|S| \geq d/2$. For each $x \in [0, 1]$, define $S_x = \{y \in [0, 1] : (x, y) \in S\}$, and let $X = \{x \in (\gamma, 1] : |S_x| > d/8\}$. Since g is symmetric, $(x, y) \in S$ if and only if $(y, x) \in S$. It follows that

$$\frac{d}{4} \leq \frac{|S|}{2} = \int_\gamma^1 |S_x| \leq |X| + \frac{d}{8}(1 - \gamma - |X|) \leq |X| + \frac{d}{8}(1 - |X|) \leq \frac{d}{8} + |X|. \tag{23}$$

Therefore $|X| \geq d/8$. Note that for all $x \in X$, $|S_x \cap A_\varepsilon| > d/16$ because $|S_x| > d/8$, $S_x \subseteq [0, \gamma]$ and $|A_\varepsilon^c \cap [0, \gamma]| < d/16$. Recall that for $y \in S_x$, $g^{+\eta}(x, y) \geq g(x, y) + \eta'$ and that for $y \in A_\varepsilon$, $u(y) \geq \varepsilon$. Therefore, for all $x \in X$,

$$\begin{aligned}
T_{g^{+\eta}} u(x) = \int_0^1 g^{+\eta}(x, y)u(y)dy &\geq \int_{S_x \cap A_\varepsilon} (g(x, y) + \eta')u(y)dy + \int_{(S_x \cap A_\varepsilon)^c} g(x, y)u(y)dy \\
&\geq T_g u(x) + \varepsilon \eta' |S_x \cap A_\varepsilon| \geq T_g u(x) + \frac{d\varepsilon\eta'}{16}.
\end{aligned}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Finally, observe that

$$\tau(g^{+\eta})^2 \geq \int_0^1 (T_{g^{+\eta}}u(x))^2 dx \geq \int_X \left(T_g u(x) + \frac{d\varepsilon\eta'}{16}\right)^2 dx + \int_{X^c} T_g u(x)^2 dx$$
$$\geq \|T_g u\|_2^2 + |X|\left(\frac{d\varepsilon\eta'}{16}\right)^2 \geq \tau(g)^2 + \frac{d^3\varepsilon^2\eta'^2}{2^{11}}.$$

Recalling that $\varepsilon = \eta/2$, $\eta' \geq \eta d/2$, and $d = (t_{\max} - \tau(g))^4/4$, we see that in the first case, we have

$$\tau(g^{+\eta})^2 - \tau(g)^2 \geq \frac{d\eta^2\|g\|_{\mathrm{op}}^2}{64} = \frac{\eta^2(t_{\max} - \tau(g))^4\tau(g)^2}{2^8}$$

and in the second case,

$$\tau(g^{+\eta})^2 - \tau(g)^2 \geq \frac{d^3\varepsilon^2\eta'^2}{2^{11}} \geq \frac{d^3\eta^2}{2^{13}}\left(\frac{\eta d}{2}\right)^2 = \frac{\eta^4 d^5}{2^{15}} = \frac{\eta^4(t_{\max} - \tau(g))^{20}}{2^{25}}.$$

Therefore, in both cases

$$\tau(g^{+\eta})^2 - \tau(g)^2 \geq \frac{\eta^4}{2^{25}}(t_{\max} - \tau(g))^{20}\tau(g)^2.$$

Finally we revisit our assumption prior to the case work that the value $z \in P$ such that $u(z) \geq 1$ is in $[0, \gamma]$. Suppose instead that $z \in (\gamma, 1]$. The equation analogous to (22) implies that $u_1 \geq \|g\|_{\mathrm{op}}$. Now switching the roles of $[0, \gamma]$ and $(\gamma, 1]$ in Case 1 gives the same lower bounds on $\tau(g^{+\eta})^2$. □

Next we establish the continuity of ϕ_τ when τ is the operator norm and W_0 is a bipartite graphon with $W_0 \in \mathcal{B}^{(\gamma,1-\gamma)}$. We use Lemma 63 to establish the sufficient increase property.

Proof of Lemma 56. By Lemma 58, it suffices to show that τ has the sufficient increase property. Fix η and t. If $t = 0$, the choice $\beta = \eta t_{\max}$ and the first bound in Lemma 63 implies that $\tau(g^{+\eta}) \geq \beta$.

If $t > 0$, the proof proceeds along the same lines as the proof of Lemma 55. Set $\alpha = \min\{t/2, ct/4\}$ and $\beta = \min\{ct/4, (t_{\max} - t)/2\}$, where

$$c = \sqrt{1 + \frac{\eta^4}{2^{25}}(t_{\max} - t)^{20}} - 1.$$

As before, the required bound is easy if $\tau(g) > t + \frac{1}{2}(t_{\max} - t)$. In this case, we again choose $g^* = g$. Next, suppose $\tau(g) \leq t + \frac{1}{2}(t_{\max} - t)$. But then $t_{\max} - \tau(g) \geq \frac{1}{2}(t_{\max} - t)$ and by the second bound in Lemma 63 , $\tau(g^{+\eta}) \geq (1 + c)\tau(g)$. The assumption $\tau(g) \geq t - \alpha$ implies

$$\tau(g^{+\eta}) \geq (t - \alpha) + (t - \alpha)c \geq t - ct/4 + ct/2 \geq t + ct/4 \geq t + \beta.$$ □

3.3 Properties of ϕ_τ at points of continuity

Lemma 64 below states that at points of continuity, the function $\phi_\tau(W_0, t)$ can be alternatively expressed as the minimum of I_{W_0} over a subset of $\mathcal{W}$. We use this characterisation of ϕ_τ to establish that ϕ_τ is strictly increasing when τ is an increasing, uniformly continuous graph parameter and ϕ_τ is continuous (Lemma 65).

Lemma 64. *Fix $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$, and let $t \in \mathbb{R}$. If $\phi_\tau(W_0, \cdot)$ is continuous at t, then*

$$\phi_\tau(W_0, t) = \inf\{I_{W_0}(f) : f \in \mathcal{W}, \tau(f) \geq t\}. \tag{24}$$

Lemma 65. *Fix $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$, and let τ be a graph parameter that is uniformly continuous (with respect to $\delta_\square$) and increasing, meaning that if $f \geq g$ pointwise then $\tau(f) \geq \tau(g)$. Suppose*

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

that $\phi_\tau(W_0,\cdot)$ *is continuous on the open interval* $(\tau(W_0), t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega))$. *Then* $\phi_\tau(W_0,\cdot)$ *is strictly increasing on* $[\tau(W_0), t_{\max}]$.

Therefore, if τ *is a uniformly continuous increasing graph parameter,* $\phi_\tau(W_0,\cdot)$ *is continuous on* $(\tau(W_0), t_{\max})$, *and* $\tilde{f}$ *is a minimiser of the variational problem (5), then* $\tau(f) = t$ *for all* $t \in [\tau(W_0), t_{\max}]$.

Remark 66. Note that the Counting Lemma [33, Lemma 10.23] and [34, Lemma 3.6] imply that homomorphism densities and the operator norm are each uniformly continuous with respect to $\delta_\square$.

Proof of Lemma 64. Let $h(t) = \inf\{I_{W_0}(f) : f \in \mathcal{W}, \tau(f) \geq t\}$. It is clear from the definition that $\phi \leq h$. We will show that the right continuity of ϕ at t implies that $\phi(t) \geq h(t)$.

To this end, we claim that for all $a \in \mathbb{R}$,

$$\inf\{J_{W_0}(\tilde{f}) : \tau(\tilde{f}) > a\} = \inf\{I_{W_0}(f) : \tau(f) > a\}. \tag{25}$$

Indeed, it is clear the left hand side is at most the right hand side. Since both sides are infinite if the set $\{\tilde{f} \in \widetilde{\mathcal{W}}_\Omega : \tau(\tilde{f}) > a\}$ is empty, we may assume that this set contains at least one $\tilde{f} \in \widetilde{\mathcal{W}}_\Omega$ such that $\tau(\tilde{f}) > a$. By the definition of J_{W_0}, there exists g_n such that $\delta_\square(g_n, \tilde{f}) \to 0$ and $I_{W_0}(g_n) \to J_{W_0}(\tilde{f})$. By continuity of τ, there exists n_0 sufficiently large such that for all $n > n_0$, $\tau(g_n) > a$. Thus

$$J_{W_0}(\tilde{f}) \geq \inf\{I_{W_0}(h) : \tau(f) > a\},$$

and so (25) follows.

We now turn to showing that if ϕ is right continuous at t, then $\phi(t) \geq h(t)$. Applying (25), we obtain

$$\begin{aligned}
\phi(t+\varepsilon) &\geq \inf\{J_{W_0}(\tilde{f}) : \tau(\tilde{f}) > t + \varepsilon/2\} \\
&= \inf\{I_{W_0}(f) : \tau(f) > t + \varepsilon/2\} \\
&\geq \inf\{I_{W_0}(f) : \tau(f) \geq t\} \\
&= h(t)
\end{aligned}$$

for any $\varepsilon > 0$. It follows by right continuity of ϕ at t that

$$\phi(t) = \lim_{\varepsilon \to 0} \phi(t+\varepsilon) \geq h(t).$$

□

Lemma 67. *Fix* $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, *and* $W_0 \in \mathcal{B}^\gamma$, *and let* $f, g \in \mathcal{W}_\Omega$. *For any* $\varepsilon > 0$, *there exists* $\eta = \eta(\varepsilon, W_0) > 0$ *such that if* $f \geq g \geq W_0$ *pointwise and* $d_\square(f,g) \geq \varepsilon$, *then*

$$I_{W_0}(g) \leq I_{W_0}(f) - \eta.$$

Proof. We may assume that $\varepsilon < 2(1-p)$ for $p \in Im(W_0) \setminus \{0,1\}$. Recall the definition

$$S^+_{\varepsilon/2} = \{(x,y) \in [0,1]^2 : f - g \geq \varepsilon/2\}.$$

Lemma 54 implies that $|S^+_{\varepsilon/2}| > \varepsilon/2$. Let

$$\eta' = \min_{p \in Im(W_0)\setminus\{0,1\}} \min_{x \in [p, 1-\varepsilon/2]} [h_p(x+\varepsilon/2) - h_p(x)] > 0.$$

Since $f \geq g \geq W_0$ and $h_p(\,\cdot\,)$ is increasing on $[p,1]$, we have that

$$h_{W_0(x,y)}(f(x,y)) \geq h_{W_0(x,y)}(g(x,y) + \varepsilon/2) \geq h_{W_0(x,y)}(g(x,y)) + \eta'$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

for all $(x,y) \in S^+_{\varepsilon/2}$. As a consequence,

$$\begin{aligned} I_{W_0}(f) &= \int_{[0,1]^2} h_{W_0(x,y)}(f(x,y))\,dx\,dy \\ &\geq \int_{[0,1]^2 \setminus S^+_{\varepsilon/2}} h_{W_0(x,y)}(g(x,y))\,dx\,dy + \int_{S^+_{\varepsilon/2}} (h_{W_0(x,y)}(g(x,y)) + \eta')\,dx\,dy \\ &\geq I_{W_0}(g) + \eta' |S^+_{\varepsilon/2}| \geq I_{W_0}(g) + \frac{\varepsilon \eta'}{2}. \end{aligned}$$

Taking $\eta = \eta' \varepsilon/2$ completes the proof. □

Proof of Lemma 65. The lemma is trivial if $\tau(W_0) = t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$ so we may assume that $\tau(W_0) < t_{\max}$. Furthermore, since $\phi_\tau(W_0, \cdot)$ is non-decreasing, it is enough to prove that it is strictly increasing on the open interval $(\tau(W_0), t_{\max})$. Let $\tau(W_0) < t_1 < t_2 < t_{\max}$. We will prove that $\phi_\tau(W_0, t_1) < \phi_\tau(W_0, t_2)$ by applying Lemma 64 and showing that

$$\inf\{I_{W_0}(f) : \tau(f) \geq t_1\} < \inf\{I_{W_0}(f) : \tau(f) \geq t_2\}.$$

To establish the above statement, it suffices to show that there exists $\eta = \eta(t_1, t_2) > 0$ such that the following is true. If $f \in \mathcal{W}_\Omega$ is such that $\tau(f) \geq t_2$, then there exists $g' \in \mathcal{W}_\Omega$ such that $\tau(g') = t_1$ and $I_{W_0}(g') \leq I_{W_0}(f) - \eta$.

Given $f \in \mathcal{W}_\Omega$ with $\tau(f) \geq t_2$, we will define g and g' satisfying $\delta_\square(g, g') \geq \beta$ where β is a function of $t_2 - t_1$. We will then show that $I_{W_0}(g) \leq I_{W_0}(f)$ and $I_{W_0}(g') \leq I_{W_0}(g) - \eta$, for $\eta = \eta(\beta)$. Define $g \in \mathcal{W}_\Omega$ such that $g(x,y) = \max\{W_0(x,y), f(x,y)\}$. Since $g \geq f$ pointwise and τ is increasing, $\tau(g) \geq \tau(f) \geq t_2$. Moreover $g \geq W_0$ pointwise. Next define $g_\alpha \in \mathcal{W}_\Omega$ where

$$g_\alpha(x,y) = W_0(x,y) + \alpha(g(x,y) - W_0(x,y)).$$

By construction $g_1 = g$ and $g_0 = W_0$, and so $\tau(g_1) \geq t_2$, and $\tau(g_0) \leq t_1$. Since $\tau(g_\alpha)$ decreases continuously as $\alpha \to 0$, there exists some α_0 such that $\tau(g_{\alpha_0}) = t_1$. Let $g' = g_{\alpha_0}$. Since $\tau(g) - \tau(g') \geq t_2 - t_1$, the uniform continuity of τ implies that $\delta_\square(g, g') \geq \beta$ for some positive $\beta = \beta(t_2 - t_1)$. It follows that $d_\square(g, g') \geq \beta$. Note also that $g, g' \in \mathcal{W}_\Omega$ and $g \geq g'$ pointwise. Therefore Lemma 67 implies that there exists $\eta = \eta(\beta)$ such that $I_{W_0}(g') \leq I_{W_0}(g) - \eta$. By construction, $I_{W_0}(g) \leq I_{W_0}(f)$ and so $I_{W_0}(g') \leq I_{W_0}(f) - \eta$. □

4. The symmetric regime in general block models

In this section we prove Theorems 19 and 20, which establish that for any d-regular graph H and $W_0 \in \mathcal{B}^\gamma$, there exists a symmetric regime for W_0 and $t(H, \cdot)$.

Notation. Throughout this section we fix a particular d-regular graph H with $d \geq 2$. Since we consider d to be fixed, we suppress the dependence on d in our notation.

Definition 68. *Let $p \in (0,1)$ and $d \geq 2$. We define $\psi_p : [0,1] \to \mathbb{R}$ as*

$$\psi_p(x) = h_p(x^{1/d}),$$

and let $\hat{\psi}_p(x)$ denote the convex minorant of $\psi_p(x)$.

Proposition 98 in the Appendix collects some useful properties of ψ_p.

Next, we introduce notation that allows us to reason about individual blocks within a graphon. Let $m \in \mathbb{Z}^+$ and let $\gamma \in \Delta^m$ be a vector of interval widths that determines block membership. Let I_j for $j \in [m]$ be as given in Definition 6. For $x \in [0,1]$, recall that $\vartheta(x)$ denotes the membership of x, so that $x \in I_{\vartheta(x)}$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

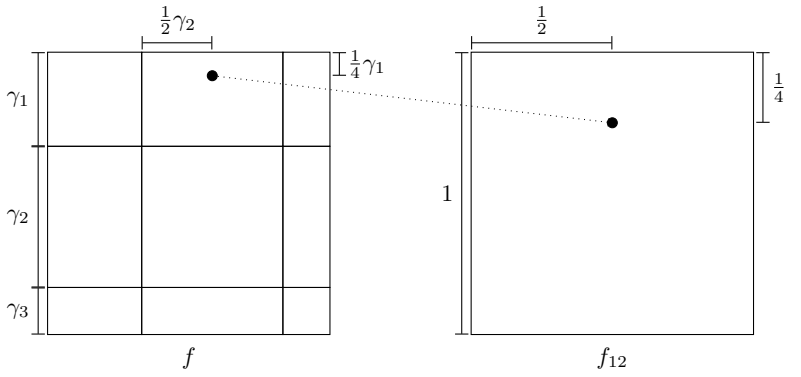


Figure 7. Illustration of the graphon $f = (f_{ij})_{i,j\in[m]}$ for $m = 3$. The indicated point in the graphon f is in the $(1, 2)$ block. The point is mapped to a point in the function f_{12}, with a scaled position.

Let $f \in \mathcal{W}$. For each $i, j \in [m]$, we define a function $f_{ij} : [0, 1]^2 \to [0, 1]$ that describes f restricted to the block $I_i \times I_j$ by[2]

$$f_{ij}(x, y) = f\left(\sum_{k=0}^{i-1} \gamma_k + x\gamma_i, \sum_{k=0}^{j-1} \gamma_k + y\gamma_j\right) \tag{26}$$

with $\gamma_0 = 0$. We write $f = (f_{ij})_{i,j\in[m]}$ to indicate that

$$f(x, y) = f_{\vartheta(x),\vartheta(y)}(r(x), r(y)), \tag{27}$$

where

$$r(x) = \frac{x - \sum_{i=0}^{\vartheta(x)-1} \gamma_i}{\gamma_{\vartheta(x)}} \tag{28}$$

(see Figure 7). By an abuse of notation, when a graphon f takes constant values on the blocks defined by γ (as in Definition 6), we write $f = (f_{ij})_{i,j\in[m]}$ where each $f_{ij} \in \mathbb{R}$ is a constant rather than a constant function.

In this section, we will utilise the restricted functions f_{ij}. Note that in contrast to the original graphon $f \in \mathcal{W}$, the f_{ij} functions are not necessarily symmetric. However, we will continue to use the cut distance $d_\square$ on these functions. In particular, we will crucially use Lemma 54 – we note that the proof of this result does not utilise the symmetry of the functions, and thus continues to hold in this extended setting.

Definition 69. *Let $d > 0$ and $m \in \mathbb{Z}^+$. For a graphon $f = (f_{ij})_{i,j\in[m]}$, we define the corresponding d-averaged block constant graphon*

$$f^* = (\|f_{ij}\|_d)_{i,j\in[m]},$$

where the d-norm is defined as $\|g\|_d = \left(\int_{[0,1]^2} g(x, y)^d dxdy\right)^{\frac{1}{d}}$.

Definition 70. *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and $W_0 \in \mathcal{B}^{\gamma,*}$ with $W_0 = (p_{ij})_{i,j\in[m]}$. We say that the graphon $f = (f_{ij})_{i,j\in[m]}$ satisfies the ε-neighbourhood minorant condition with respect to W_0 if for all (i, j) such*

[2]Strictly speaking, the function f_{ij} contains a little more information than is contained in f restricted to $I_i \times I_j$, namely, it represents the function f restricted to the closure of $I_i \times I_j$. But this difference only appears on a set of measure zero, and is thus inconsequential; furthermore, the relation (27) holds for all $(x, y) \in [0, 1]^2$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

that $p_{ij} \in (0,1)$, $\psi_{p_{ij}} = \hat{\psi}_{p_{ij}}$ *in an* ε*-neighbourhood around* $\|f_{ij}\|_d^d$, *meaning*

$$x \in \left(\|f_{ij}\|_d^d - \varepsilon, \|f_{ij}\|_d^d + \varepsilon \right) \cap [0,1] \implies \psi_{p_{ij}}(x) = \hat{\psi}_{p_{ij}}(x).$$

Remark 71. To avoid ambiguity, we define the ε-neighbourhood minorant condition for W_0 with respect to the minimal block structure describing W_0. For this reason we require $W_0 \in B^{\gamma,*}$ in Definitions 18 and 70 and Theorems 19 and 20.

4.1 A key lemma for establishing the symmetric regime

We will establish the symmetric regimes for δ small and δ large (Theorems 19 and 20 respectively) by applying the following lemma.

Lemma 72. *Let* $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, *and* $W_0 \in \mathcal{B}^{\gamma,*}$, $\varepsilon > 0$, *and* H *be a* d*-regular graph. Suppose* $\tilde{f}$ *is a minimiser of the variational problem (6) for* $\tau = t(H,\cdot)$ *and there exists a sequence of graphons* $f_n \in \mathcal{W}_\Omega$ *such that each* f_n *satisfies the* ε*-neighbourhood minorant condition,* $\delta_\square(f_n, \tilde{f}) \to 0$, *and* $I_{W_0}(f_n) \to J_{W_0}(\tilde{f})$. *Then* $\tilde{f} \in \tilde{\mathcal{B}}^\gamma$ *and* $J_{W_0}(\tilde{f}) = I_{W_0}(g)$ *for some* $g \in \mathcal{B}^\gamma$ *with* $\tilde{g} = \tilde{f}$.

The proof relies on the following two lemmas. We delay their proofs until the end of the subsection.

Lemma 73. *Let* H *be a* d*-regular graph. Let* $f = (f_{ij})_{i,j\in[m]}$ *and* $f^* = (\|f_{ij}\|_d)_{i,j\in[m]}$. *Then*

$$t(H,f) \le t(H,f^*).$$

Lemma 74. *Let* H *be a* d*-regular graph,* $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, *and* $W_0 \in \mathcal{B}^{\gamma,*}$. *Let* $f = (f_{ij})_{i,j\in[m]} \in \mathcal{W}_\Omega$. *Assume that* $f \in \mathcal{W}_\Omega$ *satisfies the* ε*-neighbourhood minorant condition with respect to* W_0 *for some* $\varepsilon > 0$. *If* $d_\square(f,f^*) \ge \alpha > 0$, *there exists* $\eta = \eta(\varepsilon, d, W_0, \alpha) > 0$, *such that*

$$I_{W_0}(f) \ge I_{W_0}(f^*) + \eta.$$

Consequently, $I_{W_0}(f) = I_{W_0}(f^*)$ *if and only if* $f \in \mathcal{B}^\gamma$.

Proof of Lemma 72. Let $\tilde{f}$ be a minimiser of the variational problem (6). Let $f_n \in \mathcal{W}_\Omega$ be such that each f_n satisfies the ε-neighbourhood minorant condition, $\delta_\square(f_n, \tilde{f}) \to 0$, and $I_{W_0}(f_n) \to J_{W_0}(\tilde{f})$. For each f_n, define f_n^* to be the corresponding d-averaged graphon. We claim that $d_\square(f_n^*, f_n) \to 0$.

Indeed, suppose for the sake of contradiction that there exists $\beta > 0$ and a subsequence $\{f_{n_i}\}_{i\ge1}$ such that $d_\square(f_{n_i}^*, f_{n_i}) \ge \beta$ for all i. By Lemma 74, there exists some $\eta > 0$ such that $I_{W_0}(f_{n_i}) \ge I_{W_0}(f_{n_i}^*) + \eta$. Consider the sequence $\{\tilde{f}_{n_i}^*\}_{i\ge1}$. By the compactness of $\tilde{\mathcal{W}}_\Omega$, there exists a convergent subsequence $\tilde{f}_{n_{i_k}}^* \to \tilde{f}'$ for some $f' \in \mathcal{W}_\Omega$. By Lemma 73, $t(H, f_{n_{i_k}}^*) \ge t(H, f_{n_{i_k}})$ and so $t(H,f') \ge t(H,f) \ge t$. It follows that $J_{W_0}(\tilde{f}') \ge \min\{J_{W_0}(\tilde{g}) : \tau(g) \ge t\}$. Observe

$$\begin{aligned} J_{W_0}(\tilde{f}) &= \liminf_{k\to\infty} I_{W_0}(f_{n_{i_k}}) \ge \liminf_{k\to\infty} I_{W_0}(f_{n_{i_k}}^*) + \eta \ge J_{W_0}(\tilde{f}') + \eta \\ &\ge \min\{J_{W_0}(\tilde{g}) : \tau(g) \ge t\} + \eta, \end{aligned}$$

and thus we have reached a contradiction.

We have shown that $d_\square(f_n^*, f_n) \to 0$. It follows that $\delta_\square(f_n^*, \tilde{f}) \to 0$. Since each $f_n^* \in \mathcal{B}^\gamma$, we can write $f_n^* = (\alpha_{ij}^n)_{i,j\in[m]}$ where each $\alpha_{ij}^n \in [0,1]$. By the compactness of $[0,1]^{m^2}$, there exists a subsequence such that

$$\alpha_{ij}^{n_k} \to \beta_{ij} \text{ for all } i,j \in [m].$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Let $g = (\beta_{ij})_{i,j\in[m]}$, $g \in \mathcal{B}^\gamma$. Since $f^*_{n_k} \to g$ pointwise and $d_\square(f^*_{n_k}, g) \le \|f^*_{n_k} - g\|_1$, the Dominated Convergence Theorem implies that $d_\square(f^*_{n_k}, g) \to 0$. Since $\delta_\square(f^*_{n_k}, \tilde{f}) \to 0$, we have $\delta_\square(\tilde{g}, \tilde{f}) = 0$. Thus $\tilde{f} \in \tilde{\mathcal{B}}^\gamma$. Note $\tau(g) \ge t$.

Next we show that $I_{W_0}(g) = J_{W_0}(\tilde{f})$. Since $\delta_\square(f^*_n, \tilde{f}) \to 0$, $\liminf_{n\to\infty} I_{W_0}(f^*_n) \ge J_{W_0}(\tilde{f})$. Further, since each f_n satisfies the ε-neighbourhood minorant condition, then by Lemma 74, $\limsup_{n\to\infty} I_{W_0}(f^*_n) \le \lim_{n\to\infty} I_{W_0}(f_n) = J_{W_0}(\tilde{f})$. Thus, $\lim_{n\to\infty} I_{W_0}(f^*_n) = J_{W_0}(\tilde{f})$. Since $f^*_{n_k} \to g$ pointwise and $f_{n_k}, g \in \mathcal{W}_\Omega$, the continuity of h_p for any fixed $p \in (0,1)$ implies that

$$I_{W_0}(g) = \lim_{k\to\infty} I_{W_0}(f^*_{n_k}) = J_{W_0}(\tilde{f}).$$

□

4.1.1. Proofs of supporting lemmas

We turn to the proofs of Lemmas 73 and 74.

Proof of Lemma 73. The lemma is a direct consequence of the generalised Hölder inequality of [23] (stated as Theorem 97 in the Appendix). Let $v = |V(H)|$ be the number of vertices in H. Recall the definition of $t(H,f)$:

$$t(H,f) = \int_{[0,1]^v} \prod_{(i,j)\in E(H)} f(x_i, x_j) dx_1 \cdots dx_v.$$

We break up the integration over the blocks specified by the vector γ. Recall the definition of I_j (Definition 6). We have

$$t(H,f) = \sum_{i_1=1}^{m} \cdots \sum_{i_v=1}^{m} \int_{x_1\in I_{i_1}} \cdots \int_{x_v\in I_{i_v}} \prod_{(a,b)\in E(H)} f(x_a, x_b) dx_1 \cdots dx_v.$$

Recall that $f(x,y) = f_{\vartheta(x),\vartheta(y)}(r(x), r(y))$ as stated in (28) and (27). It follows that

$$f(x_a, x_b) = f_{\vartheta(x_a),\vartheta(x_b)}(r(x_a), r(x_b)) = f_{i_a,i_b}(r(x_a), r(x_b)).$$

Substituting and applying a change of variables, we obtain

$$t(H,f) = \sum_{i_1=1}^{m} \cdots \sum_{i_v=1}^{m} \int_{x_1\in I_{i_1}} \cdots \int_{x_v\in I_{i_v}} \prod_{(a,b)\in E(H)} f_{i_a,i_b}(r(x_a), r(x_b)) dx_1 \cdots dx_v$$

$$= \sum_{i_1=1}^{m} \cdots \sum_{i_v=1}^{m} \left(\prod_{j=1}^{v} \gamma_{i_j} \right) \int_{x\in[0,1]^v} \prod_{(a,b)\in E(H)} f_{i_a,i_b}(x_a, x_b) dx_1 \cdots dx_v.$$

By the generalised Hölder inequality (Theorem 97),

$$t(H,f) \le \sum_{i_1=1}^{m} \cdots \sum_{i_v=1}^{m} \left(\prod_{j=1}^{v} \gamma_{i_j} \right) \prod_{(a,b)\in E(H)} \|f_{i_a,i_b}\|_d = t(H,f^*).$$

□

We will apply Lemmas 75, 77 and 78 in the proof of Lemma 74.

Lemma 75. *Fix $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, and let $f, g \in \mathcal{B}^\gamma$ be two graphons with $f = (f_{ij})_{(i,j)\in[m]^2}$ and $g = (g_{ij})_{(i,j)\in[m]^2}$. Then $d_\square(f,g) \le \max_{i,j} d_\square(f_{ij}, g_{ij})$.*

Proof. Recalling the definition of $d_\square$, we will need to bound $\sup_{S,T} |\int_{S\times T} (f-g)|$ over all measurable subsets $S, T \subset [0,1]$. Fix two such subsets, and recall that I_i is the ith block. By the triangle

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

inequality,

$$\left|\int_{S\times T}\big(f(x,y)-g(x,y)\big)\,dxdy\right| = \left|\sum_{(i,j)\in[m]^2}\int_{\{S\times T\}\cap\{I_i\times I_j\}}\big(f(x,y)-g(x,y)\big)\,dxdy\right|$$

$$\leq \sum_{(i,j)\in[m]^2}\left|\int_{(S\times T)\cap(I_i\times I_j)}\big(f(x,y)-g(x,y)\big)\,dxdy\right|.$$

Setting $T_i = r(T\cap I_i)$ and $S_i = r(T\cap I_i)$, where r is as described in (28) and (27), we write the right hand side as

$$\sum_{(i,j)\in[m]^2}\left|\gamma_i\gamma_j\int_{(S_i\times T_i)}\big(f_{ij}(x,y)-g_{ij}(x,y)\big)\,dxdy\right| \leq \sum_{i,j}\gamma_i\gamma_j\max_{k,\ell}d_\square(f_{k\ell},g_{k\ell}) = \max_{k,\ell}d_\square(f_{k\ell},g_{k\ell}).$$

Since $S, T\subset[0,1]$ were arbitrary, this completes the proof. □

Definition 76. *Given $f:[0,1]^2\to[0,1]$ measurable, $c\in\mathbb{R}$, and $\varepsilon\in[0,1]$, define the sets*

$$A^+_\varepsilon(f,c) = \{(x,y)\in[0,1]^2 : f(x,y)-c\geq\varepsilon\}$$

$$A^-_\varepsilon(f,c) = \{(x,y)\in[0,1]^2 : c-f(x,y)\geq\varepsilon\}.$$

Lemma 77. *Let $f:[0,1]^2\to[0,1]$ be measurable, and let g be the constant graphon that takes value $\|f\|_d$. There exists $\beta=\beta(\varepsilon)$ such that if $d_\square(f,g)\geq\varepsilon$, then*

$$|A^+_\beta(f^d,\|f\|_d^d)|\geq\beta \quad \textit{and} \quad |A^-_\beta(f^d,\|f\|_d^d)|\geq\beta$$

where f^d denotes the function $f^d(x,y)=f(x,y)^d$.

Proof. Without loss of generality, we may assume that ε is small enough such that $1\geq\varepsilon/4\geq d\,(\varepsilon/8)^d$. We begin by observing that for $u,v\geq0$ satisfying $u-v\geq\varepsilon/4$ and $a(x)=x^d$,

$$a(u)-a(v) = \int_v^u a'(x)dx \geq \int_{v+\varepsilon/8}^{v+\varepsilon/4}a'(x)dx \geq \frac{\varepsilon}{8}a'\left(\frac{\varepsilon}{8}\right) = d\left(\frac{\varepsilon}{8}\right)^d. \tag{29}$$

Let $f:[0,1]^2\to[0,1]$ satisfy the hypotheses of the lemma. Lemma 54 implies that $|A^+_{\varepsilon/4}(f,\|f\|_d)|\geq\varepsilon/4$ or $|A^-_{\varepsilon/4}(f,\|f\|_d)|\geq\varepsilon/4$. We will establish the result in the case that $|A^+_{\varepsilon/4}(f,\|f\|_d)|\geq\varepsilon/4$. The other case follows by an analogous argument. Note that if $f(x,y)-\|f\|_d\geq\varepsilon/4$, then (29) implies

$$f^d(x,y)-\|f\|_d^d\geq d\left(\frac{\varepsilon}{8}\right)^d.$$

Let $c=d(\varepsilon/8)^d$. Since $|A^+_{\varepsilon/4}(f,\|f\|_d)|\geq\varepsilon/4$ and $A^+_{\varepsilon/4}(f,\|f\|_d)\subseteq A^+_c(f^d,\|f\|_d^d)$, it follows that $|A^+_c(f^d,\|f\|_d^d)|\geq\varepsilon/4\geq c$. For ease of notation, let $A^+=A^+_0(f^d,\|f\|_d^d)$ and $A^-=A^-_0(f^d,\|f\|_d^d)$. Observe

$$c^2\leq\int_{A^+_c(f^d,\|f\|_d^d)}f^d-\|f\|_d^d\leq\int_{A^+}f^d-\|f\|_d^d=\int_{A^-}\|f\|_d^d-f^d\leq|A^-_{c^2/2}(f^d,\|f\|_d^d)|+\frac{c^2}{2}.$$

It follows that $|A^-_{c^2/2}(f^d,\|f\|_d^d)|\geq c^2/2$. Since $|A^+_{c^2/2}(f^d,\|f\|_d^d)|\geq|A^+_c(f,\|f\|_d)|\geq c\geq c^2/2$, taking $\beta=c^2/2$ completes the proof. □

Lemma 78. *Let $p\in(0,1)$, $\varepsilon>0$, and $f:[0,1]^2\to[0,1]$ measurable. There exists $\eta=\eta(p,\varepsilon,\beta)>0$ such that if $|A^-_\beta(f,\|f\|_1)|, |A^+_\beta(f,\|f\|_1)|\geq\beta$ and $\psi_p(x)=\hat\psi_p(x)$ for all $x\in\big(\|f\|_1-\varepsilon,\|f\|_1+\varepsilon\big)$,*

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

then

$$\int_{[0,1]^2} \hat{\psi}_p(f) \geq \hat{\psi}_p\left(\|f\|_1\right) + \eta.$$

Proof. For ease of notation let $z = \|f\|_1$. Since ψ_p is differentiable, it follows that $\hat{\psi}_p$ is differentiable at z, and so $\hat{\psi}_p'(z) = \psi_p'(z)$ is a subdifferential of $\hat{\psi}_p$ at z. Let

$$g(w) = \psi_p(z) + \psi_p'(z)(w - z).$$

Moreover, since $\psi_p'(z)$ is a subdifferential of $\hat{\psi}_p$ at z, $\hat{\psi}_p(w) \geq g(w)$ for all $w \in [0,1]$. Since g is a linear function,

$$\int_{[0,1]^2} g(f(x,y)) = g\left(\int_{[0,1]^2} f(x,y)\right) = g\left(\|f\|_1\right) = \psi_p\left(\|f\|_1\right) = \hat{\psi}_p\left(\|f\|_1\right). \tag{30}$$

Define

$$d(w) = \psi_p(w) - g(w),$$

and note that $d'(w) = \psi_p'(w) - \psi_p'(z)$. Assuming $w \geq z + \beta/2$ and applying the Fundamental Theorem of Calculus twice, we obtain

$$\begin{aligned} d(w) = d(w) - d(z) &= \int_z^w d'(a)\, da = \int_z^w \int_z^a \psi_p''(b)\, db\, da \\ &= \int_z^{z+\beta/2} \int_z^a \psi_p''(b)\, db\, da + \int_{z+\beta/2}^w \int_z^a \psi_p''(b)\, db\, da, \\ &\geq \frac{(w - z - \beta/2)^2}{2} \min\{\psi_p''(x) : x \in [z + \beta/2, w]\}, \end{aligned}$$

provided that $\psi_p''(x)$ is non-negative on $[z, w]$. The same argument applies for $w \leq z$, and one obtains a similar lower bound, with $\min\{\psi_p''(x) : x \in [w, z - \beta/2]\}$ instead.

Next, we construct a set $S \subseteq [0,1]^2$ and choose $\eta' > 0$ such that for all $(x,y) \in S$,

(i) $\psi_p(f(x,y)) = \hat{\psi}_p(f(x,y))$,
(ii) $|f(x,y) - z| \geq \beta$,
(iii) $\psi_p''(b) \geq \eta'$ for all $b \in [f(x,y), z - \beta/2]$ if $f(x,y) \leq z$, or for all $b \in [z + \beta/2, f(x,y)]$ if $f(x,y) \geq z$,
(iv) ψ_p'' is non-negative on $[f(x,y), z]$ if $f(x,y) \leq z$ or non-negative on $[z, f(x,y)]$ if $f(x,y) \geq z$, and
(v) $|S| \geq \beta$.

Our construction of S depends on ψ_p and z. Let p_0 be as given in Proposition 98. There are three cases concerning p_0. In each case, η' is well-defined because it is the minimum of a continuous function over a compact set.

(1) If $p > p_0$, then ψ_p'' is positive on $[0,1]$. Let $S = A_\beta^+(f,z)$ and $\eta' = \min\{\psi_p''(x) : x \in [0,1]\}$.
(2) If $p < p_0$, then the function ψ_p has two inflection points r_1 and r_2, and $\psi_p''(x) > 0$ on $[0, r_1)$ and $(r_2, 1]$ and $\psi_p''(x) < 0$ on (r_1, r_2). Note that $z \notin (r_1 - \varepsilon, r_2 + \varepsilon)$, since $\psi_p = \hat{\psi}_p$ is convex in an ε-neighbourhood around z.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

- If $z \le r_1 - \varepsilon$, let $\eta' = \min\{\psi_p''(x) : x \in [0, r_1 - \varepsilon]\}$. Note $\eta' > 0$ since ψ_p'' is positive on $[0, r_1 - \varepsilon]$. Let $S = A_\beta^-(f, z)$.
- If $z \ge r_2 + \varepsilon$, let $\eta' = \min\{\psi_p''(x) : x \in [r_2 + \varepsilon, 1]\}$. Note $\eta' > 0$ since ψ_p'' is positive on $[r_2 + \varepsilon, 1]$. Let $S = A_\beta^+(f, z)$.

(3) If $p = p_0$, then the function ψ_p has one point r such that $\psi_p''(r) = 0$ and $\psi_p''(x) > 0$ on $[0, r)$ and $(r, 1]$. If $z < r$, let $\eta' = \min\{\psi_p''(x) : x \in [0, z - \beta/2]\}$ and $S = A_\beta^-(f, z)$. If $z \ge r$, let $\eta' = \min\{\psi_p''(x) : x \in [z + \beta/2, 1]\}$ and $S = A_\beta^+(f, z)$. Since $A_\beta^+(f, z), A_\beta^-(f, z)$ are non-empty, $0 < z - \beta/2 < z + \beta/2 < 1$.

By our choice of $S = A_\beta^-(f, z)$ or $S = A_\beta^+(f, z)$ as needed, it is easy to see that properties (i), (ii), and (v) are satisfied. To see how the remaining properties are satisfied, consider for example Case (2), where $z \le r_1 - \varepsilon$. In this case, $(x, y) \in S = A_\beta^-(f, z)$ implies $f(x, y) \le z - \beta \le r_1 - \varepsilon - \beta$. Since $\psi_p''(x) > 0$ on $[0, r_1)$, property (iv) is satisfied. Property (iii) is satisfied since $\psi_p''(b) \ge \eta'$ for all $b \in [0, r_1 - \varepsilon] \supset [f(x, y), z - \beta/2]$, where the inclusion holds for any $(x, y) \in S$.

Note that if $(x, y) \in S$ and $f = f(x, y) \ge z$, then properties (i)-(iv) imply that

$$\hat{\psi}_p(f) - g(f) = \psi_p(f) - g(f) = d(f) \ge \frac{(f - z - \beta/2)^2}{2} \min\{\psi_p''(b) : b \in [z + \beta/2, f]\} \ge \frac{\eta' \beta^2}{8}. \tag{31}$$

The same bound holds if $f(x, y) \le z$.

Recall that $\hat{\psi}_p(w) \ge g(w)$ for all $w \in [0, 1]$. It follows by (30), (31) and property (v) that

$$\int_{[0,1]^2} \hat{\psi}_p(f(x, y))\, dx\, dy - \hat{\psi}_p\left(\|f\|_1\right) = \int_{[0,1]^2} \hat{\psi}_p(f(x, y)) - g(f(x, y))\, dx\, dy$$
$$\ge \int_S \hat{\psi}_p(f(x, y)) - g(f(x, y))\, dx\, dy \ge |S| \frac{\eta' \beta^2}{8} \ge \frac{\eta' \beta^3}{8}.$$

Taking $\eta = \eta' \beta^3 / 8$ completes the proof. □

Proof of Lemma 74. Suppose f is a graphon that satisfies the ε-neighbourhood minorant condition, and has the property that $d_\square(f, f^*) \ge \alpha$. Since $d_\square(f, f^*) \ge \alpha$, Lemma 75 implies that there exists some $a, b \in [m]^2$ such that $d_\square(f_{ab}, \|f_{ab}\|_d) \ge \alpha$ (where by an abuse of notation, $\|f_{ab}\|_d$ denotes the constant function that takes that value $\|f_{ab}\|_d$). Since $f \in \mathcal{W}_\Omega$, f_{ij} is constant whenever $p_{ij} \in \{0, 1\}$ and the values a, b are such that $p_{ab} \in (0, 1)$. Lemma 77 implies that there exists $\beta = \beta(\alpha)$ such that $|A_\beta^+(f_{ab}^d, \|f_{ab}\|_d^d)| \ge \beta$ and $|A_\beta^-(f_{ab}^d, \|f_{ab}\|_d^d)| \ge \beta$.

For each (i, j) such that $p_{ij} \in (0, 1)$, Lemma 78 implies that there exists $\eta_{ij} = \eta_{ij}(p_{ij}, \varepsilon, \beta)$ such that if $g : [0, 1]^2 \to [0, 1]$, $|A_\beta^-(g, \|g\|_1)|, |A_\beta^+(g, \|g\|_1)| \ge \beta$ and $\psi_{p_{ij}}(x) = \hat{\psi}_{p_{ij}}(x)$ for all $x \in \left(\|g\|_1 - \varepsilon, \|g\|_1 + \varepsilon\right)$, then

$$\int_{[0,1]^2} \hat{\psi}_{p_{ij}}(g) \ge \hat{\psi}_{p_{ij}}\left(\|g\|_1\right) + \eta_{ij}. \tag{32}$$

Let $\eta' = \min_{(i,j)}\{\eta_{ij} : p_{ij} \in (0, 1)\}$.

Since f satisfies the ε-neighbourhood minorant condition, $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ in an ε-neighbourhood around the value $\|f_{ab}\|_d^d = \|f_{ab}^d\|_1$. Applying (32) to f_{ab}^d, we obtain

$$\int_{[0,1]^2} h_{p_{ab}}(f_{ab}(x, y))dxdy = \int_{[0,1]^2} \psi_{p_{ab}}\left(f_{ab}^d(x, y)\right) dxdy \ge \int_{[0,1]^2} \hat{\psi}_{p_{ab}}\left(f_{ab}^d(x, y)\right) dxdy$$
$$\ge \hat{\psi}_{p_{ab}}\left(\left\|f_{ab}\right\|_d^d\right) + \eta' = \psi_{p_{ab}}\left(\left\|f_{ab}\right\|_d^d\right) + \eta' = h_{p_{ab}}\left(\left\|f_{ab}\right\|_d\right) + \eta'.$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

For all i, j such that $p_{ij} \in (0,1)$, the point $\left(\|f_{ij}\|_d^d, h_p(\|f_{ij}\|_d)\right)$ lies on the convex minorant of $\psi_{p_{ij}}(x) = h_{p_{ij}}(x^{1/d})$, and so applying Jensen's Inequality we obtain

$$\int_{[0,1]^2} h_{p_{ij}}(f_{ij}(x,y))dxdy = \int_{[0,1]^2} \psi_{p_{ij}}\left(f_{ij}^d(x,y)\right)dxdy \geq \int_{[0,1]^2} \hat{\psi}_{p_{ij}}\left(f_{ij}^d(x,y)\right)dxdy$$
$$\geq \hat{\psi}_{p_{ij}}\left(\int_{[0,1]^2} f_{ij}^d(x,y)dxdy\right) = \hat{\psi}_{p_{ij}}\left(\|f_{ij}\|_d^d\right) = \psi_{p_{ij}}\left(\|f_{ij}\|_d^d\right) = h_{p_{ij}}\left(\|f_{ij}\|_d\right).$$

Since $f \in \mathcal{W}_\Omega$, it holds that $f(x,y) = W_0(x,y)$ for all (x,y) such that $W_0(x,y) \in \{0,1\}$. Let c_{ij} be the indicator that $p_{ij} \in (0,1)$. Observe

$$I_{W_0}(f) = \sum_{i=1}^m \sum_{j=1}^m \gamma_i\gamma_j c_{ij} \int_{[0,1]^2} h_{p_{ij}}(f_{ij}(x,y))dxdy$$
$$\geq \eta'\gamma_a\gamma_b + \sum_{i=1}^m \sum_{j=1}^m \gamma_i\gamma_j c_{ij} h_{p_{ij}}\left(\|f_{ij}\|_d\right)$$
$$= \eta'\gamma_a\gamma_b + I_{W_0}\left(f^*\right).$$

Taking $\eta = \eta' \min_{(i,j)}\{\gamma_i\gamma_j : p_{ij} \in (0,1)\}$ yields the desired result. □

4.2 Symmetry for small δ

We now prove Theorem 19, which establishes the existence of a symmetric regime when δ is small. We state the key lemmas used in the proof and defer the proofs of these lemmas to the end of the subsection.

Lemma 79. *Given $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^{\gamma,*}$ and a finite d-regular graph H, there exist $\delta_0 = \delta_0(H, W_0)$ and $\varepsilon = \varepsilon(W_0) > 0$ such that the following is true. For all $0 < \delta < \delta_0$, if $\tilde{f}$ is a minimiser of the variational problem (6) with $t = (1+\delta)t(H, W_0)$, then there exists a sequence of graphons $f_n \in \mathcal{W}_\Omega$ such that $\delta_\square(f_n, \tilde{f}) \to 0$, $I_{W_0}(f_n) \to J_{W_0}(\tilde{f})$ and each f_n satisfies the ε-neighbourhood minorant condition.*

We now use Lemma 79 to prove Theorem 19.

Proof of Theorem 19. Let $W_0 \in \mathcal{B}^{\gamma,*}$, and let H be a finite d-regular graph. Let $\delta_0 > 0$ and $\varepsilon > 0$ be as in Lemma 79, and assume that $\delta < \delta_0$. Suppose that $\tilde{f}$ is a minimiser of the variational problem (6) with $t = (1+\delta)t(H, W_0)$. By Lemma 79, there exists a sequence of graphons $f_n \in \mathcal{W}_\Omega$ such that each f_n satisfies the ε-neighbourhood minorant condition, $\delta_\square(f_n, \tilde{f}) \to 0$, and $I_{W_0}(f_n) \to J_{W_0}(\tilde{f})$. Lemma 72 implies that $\tilde{f} \in \tilde{\mathcal{B}}^\gamma$ and $J_{W_0}(\tilde{f}) = I_{W_0}(g)$ for some $g \in \mathcal{B}^\gamma$ with $\tilde{g} = \tilde{f}$. Thus, the problem is in the symmetric regime.

Let $\tilde{h}$ be a minimiser of the variational problem (6) with $t = (1+\delta)t(H, W_0)$. By the above argument we may assume that $h \in \mathcal{B}^\gamma$ and $I_{W_0}(h) = J_{W_0}(\tilde{h})$. Note that

$$\min\{I_{W_0}(g) : g \in \mathcal{B}^\gamma, t(H,g) \geq (1+\delta)t(H, W_0)\} \leq I_{W_0}(h) = J_{W_0}(\tilde{h})$$

and

$$\min\{I_{W_0}(g) : g \in \mathcal{B}^\gamma, t(H,g) \geq (1+\delta)t(H, W_0)\} \geq \min\{J_{W_0}(\tilde{g}) : \tilde{g} \in \tilde{\mathcal{B}}^\gamma, t(H,g) \geq (1+\delta)t(H, W_0)\}$$
$$= J_{W_0}(\tilde{h}) = I_{W_0}(h).$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

It follows that

$$J_{W_0}(\tilde{h}) = I_{W_0}(h) = \min\{I_{W_0}(g) : g \in \mathcal{B}^\gamma, t(H,g) \geq (1+\delta)t(H,W_0)\}.$$ □

4.2.1. Proofs of supporting lemmas

We now prove Lemma 79. We first show that graphons with homomorphism density close to $t(H,W_0)$ must be close to W_0 pointwise except possibly on a small set.

Lemma 80. *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$ such that $\gamma \in \mathbb{R}^m$, and let H be a finite d-regular graph. Suppose $f \geq W_0$ pointwise and $t(H,f) \leq (1+\delta)t(H,W_0)$ for $\delta > 0$. Let Y be a relevant interval labelling vector with respect to W_0. Suppose $I_a \times I_b$ contributes to Y, i.e. $Y_i = a$ and $Y_j = b$ for some $\{i,j\} \in E(H)$. (See Definition 53.) Let $S_\varepsilon = \{(x,y) \in I_a \times I_b : f(x,y) \geq W_0(x,y) + \varepsilon\}$. Then*

$$\varepsilon|S_\varepsilon| \leq \frac{\delta t(H,W_0)p_{ab}\gamma_a\gamma_b}{t(H,W_0,Y)},$$

where p_{ab} is the value of W_0 on $I_a \times I_b$, $\gamma_a = |I_a|$, and $\gamma_b = |I_b|$.

Proof. Let $v = |V(H)|$ be the number of vertices in H. Observe

$$\begin{aligned}
t(H,f,Y) &= \int_{x_1 \in I_{Y_1}} \cdots \int_{x_v \in I_{Y_v}} \prod_{\{u,w\} \in E(H)} f(x_u,x_w)\, dx_v \ldots dx_1 \\
&\geq \left(\prod_{u \in [v]\setminus\{i,j\}} |I_{Y_u}| \prod_{\{u,w\} \in E(H)\setminus\{i,j\}} p_{Y_uY_w} \right) \int_{x_i \in I_a} \int_{x_j \in I_b} f(x_i,x_j) dx_i\, dx_j \\
&\geq \left(\prod_{u \in [v]\setminus\{i,j\}} |I_{Y_u}| \prod_{\{u,w\} \in E(H)\setminus\{i,j\}} p_{Y_uY_w} \right) \left(\int_{(x_i,x_j) \in I_a \times I_b \setminus S_\varepsilon} p_{ab} + \int_{(x_i,x_j) \in S_\varepsilon} (\varepsilon + p_{ab}) \right) \\
&= \left(\prod_{u \in [v]\setminus\{i,j\}} |I_{Y_u}| \prod_{\{u,w\} \in E(H)\setminus\{i,j\}} p_{Y_uY_w} \right) \left(|I_a||I_b|p_{ab} + \varepsilon|S_\varepsilon| \right) \\
&= t(H,W_0,Y) + \frac{\varepsilon|S_\varepsilon|}{p_{ab}\gamma_a\gamma_b} t(H,W_0,Y).
\end{aligned}$$

We use this to lower bound the homomorphism density of f, and obtain

$$\begin{aligned}
(1+\delta)t(H,W_0) \geq t(H,f) &= \sum_{Z \in [m]^v} t(H,f,Z) \\
&\geq \frac{\varepsilon|S_\varepsilon|}{p_{ab}\gamma_a\gamma_b} t(H,W_0,Y) + \sum_{Z \in [m]^v} t(H,W_0,Z) \\
&= \frac{\varepsilon|S_\varepsilon|}{p_{ab}\gamma_a\gamma_b} t(H,W_0,Y) + t(H,W_0).
\end{aligned}$$

The statement of the lemma follows directly. □

Lemma 81. *Fix $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$ and H a finite d-regular graph. There exist $\delta_0 = \delta_0(H,W_0)$ and $\varepsilon = \varepsilon(W_0) > 0$ such that the following is true. If $f \in \mathcal{W}_\Omega$, $f \geq W_0$ pointwise, $f = W_0$ on irrelevant blocks (Definition 53), and $t(H,f) \leq (1+\delta_0)t(H,W_0)$, then f satisfies the ε-neighbourhood minorant condition.*

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Proof. Let $W_0 = (p_{ij})_{i,j\in[m]}$. By Proposition 98, for $p_{ij} \in (0,1)$ the function $\psi_{p_{ij}}:(0,1) \to \mathbb{R}$ is either convex or $\hat{\psi}_{p_{ij}}$ is constructed by replacing $\psi_{p_{ij}}$ with its lower common tangent on exactly one interval. If $\psi_{p_{ij}}$ is convex, then the minorant condition is trivially satisfied for the (i,j) block. Otherwise, let q_{ij} be such that q_{ij}^d is the left point of intersection between $\psi_{p_{ij}}$ and its lower common tangent. Let $P = \{(i,j) : \psi''_{p_{ij}}(x) = 0 \text{ for some } x \in [p_{ij}^d, 1]\}$ and define

$$\varepsilon = \min_{(i,j)\in P} q_{ij}^d - p_{ij}^d.$$

Since $\psi'_{p_{ij}}(p_{ij}^d) = 0$ for all (i,j), we have $q_{ij} > p_{ij}$ for all $(i,j) \in P$, so that $\varepsilon > 0$.

Let $f \in \mathcal{W}_\Omega, f \geq W_0$ pointwise, $f = W_0$ on irrelevant blocks (Definition 53), and $t(H,f) \leq (1+\delta)t(H, W_0)$. Let $(a,b) \in P$. When $I_a \times I_b$ is irrelevant, $\|f_{ab}\|_d^d = p_{ab}^d \leq q_{ab}^d - \varepsilon$. Since $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ on $[0, q_{ab}^d]$, $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ in an ε-neighbourhood around $\|f_{ab}\|_d^d$. It suffices to show that for δ sufficiently small, $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ in a neighbourhood around $\|f_{ab}\|_d^d$ for all $(a,b) \in P$ such that $I_a \times I_b$ is relevant.

Fix such a pair (a,b). Let $S_\eta = \{(x,y) \in I_a \times I_b : f(x,y) \geq W_0(x,y) + \eta\}$. Let $S_\eta^c = (I_a \times I_b) \setminus S_\eta$. Observe

$$\begin{aligned}
\|f_{ab}\|_d^d &= \frac{1}{\gamma_a\gamma_b}\left(\int_{S_\eta^c} f(x,y)^d\,dx\,dy + \int_{S_\eta} f(x,y)^d\,dx\,dy\right)\\
&\leq \frac{1}{\gamma_a\gamma_b}\left(|S_\eta^c|(p_{ab}+\eta)^d + |S_\eta|\right)\\
&\leq (p_{ab}+\eta)^d + \frac{1}{\gamma_a\gamma_b}|S_\eta|.
\end{aligned}$$

Let $\eta(\delta) = \sqrt{\frac{\delta t(H,W_0)p_{ab}\gamma_a\gamma_b}{t(H,W_0,Y)}}$. By Lemma 80,

$$\eta(\delta)|S_{\eta(\delta)}| \leq \frac{\delta t(H,W_0)p_{ab}\gamma_a\gamma_b}{t(H,W_0,Y)} \implies |S_{\eta(\delta)}| \leq \eta(\delta).$$

It follows that

$$\|f_{ab}\|_d^d \leq \big(p_{ab} + \eta(\delta)\big)^d + \frac{\eta(\delta)}{\gamma_a\gamma_b}.$$

Note that as $\delta \to 0$, $\eta(\delta) \to 0$, and the above expression approaches p_{ab}^d. Thus we can pick δ sufficiently small so that $\|f_{ab}\|_d^d < q_{ab}^d - \varepsilon$. Since $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ on $[0, q_{ab}^d]$, $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ in an ε-neighbourhood around $\|f_{ab}\|_d^d$. Repeating the argument for each block concludes the proof. □

We will combine Lemma 81 with the following lemma to prove Lemma 79.

Lemma 82. *Let $\tilde{f}$ be a minimiser of the variational problem (6) with $t = (1+\delta)t(H, W_0)$ for $\delta > 0$. Then there exists a sequence $\{f_n\}_{n\geq 1}$ such that $f_n \in \mathcal{W}_\Omega$, $f_n \geq W_0$ pointwise, $f_n = W_0$ on irrelevant blocks, $I_{W_0}(f_n) \to J_{W_0}(\tilde{f})$ and $\delta_\square(f_n, \tilde{f}) \to 0$.*

Proof. Let $\tilde{f}$ be a minimiser of the variational problem (6) with $t = (1+\delta)t(H, W_0)$. There exists a sequence $\{f_n\}_{n\geq 1}$ such that $f_n \in \mathcal{W}_\Omega$, $I_{W_0}(f_n) \to J_{W_0}(\tilde{f})$ and $\delta_\square(\tilde{f}, f_n) \to 0$. Let $R \subseteq [0,1]^2$ be the union of the relevant blocks (Definition 53). Define $f_n' \in \mathcal{W}_\Omega$ such that

$$f_n'(x,y) = \begin{cases} \max\{f_n(x,y), W_0(x,y)\} & (x,y) \in R \\ W_0(x,y) & (x,y) \notin R \end{cases}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Note that this ensures that $f'_n \geq W_0$ pointwise, and note further that $f'_n = W_0$ whenever $f'_n \neq f_n$, which in turn implies that $h_{W_0(x,y)}(f'_n(x,y)) \leq h_{W_0(x,y)}(f_n(x,y))$ for all $(x,y) \in [0,1]^2$ and hence $I_{W_0}(f'_n) \leq I_{W_0}(f_n)$.

We claim that the lemma follows from showing that $d_\square(f'_n, f_n) \to 0$. Indeed, if $d_\square(f'_n, f_n) \to 0$, then $\delta_\square(f'_n, \tilde{f}) \to 0$. Thus $\liminf I_{W_0}(f'_n) \geq J_{W_0}(\tilde{f})$. Since $I_{W_0}(f'_n) \leq I_{W_0}(f_n)$ for each n, $\limsup I_{W_0}(f'_n) \leq \limsup I_{W_0}(f_n) = J_{W_0}(\tilde{f})$. Thus, $I_{W_0}(f'_n) \to J_{W_0}(\tilde{f})$ as $n \to \infty$. Note that by construction $f'_n \geq W_0$ pointwise and $f'_n = W_0$ on irrelevant blocks. This completes the proof.

It remains to prove that $d_\square(f'_n, f_n) \to 0$. Suppose for contradiction that there exists $\varepsilon > 0$ and a subsequence such that $d_\square(f'_{n_i}, f_{n_i}) \geq \varepsilon$ for all $i \geq 1$. Let

$$\eta = \min_{p \in Im(W_0) \setminus \{0,1\}} \min\{h_p(p + \varepsilon/4), h_p(p - \varepsilon/4)\},$$

where we have chosen $\varepsilon > 0$ sufficiently small so that $p + \varepsilon/4, p - \varepsilon/4 \in (0,1)$ for all $p \in \mathrm{Im}(W_0) \setminus \{0,1\}$. We will show that $I_{W_0}(f'_{n_i}) \leq I_{W_0}(f_{n_i}) - \varepsilon\eta/4$, and then use this to derive a contradiction.

Indeed, let $S^+_{\varepsilon/4} = \{(x,y) \in [0,1]^2 : f'_{n_i}(x,y) - f_{n_i}(x,y) \geq \varepsilon/4\}$ and $S^-_{\varepsilon/4} = \{(x,y) \in [0,1]^2 : f_{n_i}(x,y) - f'_{n_i}(x,y) \geq \varepsilon/4\}$. Let $S = S^+_{\varepsilon/4} \cup S^-_{\varepsilon/4}$. Lemma 54 implies that $|S| \geq \varepsilon/4$.

If $(x,y) \in S^+_{\varepsilon/4} \cap R$, then $\max\{f_{n_i}(x,y), W_0(x,y)\} - f_{n_i}(x,y) \geq \varepsilon/4$. It follows that $f'_{n_i}(x,y) = W_0(x,y)$. If $(x,y) \in S^+_{\varepsilon/4} \setminus R$, then $W_0(x,y) - f_{n_i}(x,y) \geq \varepsilon/4$. In both cases $f_{n_i}(x,y) \leq W_0(x,y) - \varepsilon/4$, and so $h_{W_0(x,y)}(f_{n_i}(x,y)) \geq \eta$. Therefore

$$h_{W_0(x,y)}(f'_{n_i}(x,y)) = 0 \leq h_{W_0(x,y)}(f_{n_i}(x,y)) - \eta.$$

If $(x,y) \in S^-_{\varepsilon/4}$, then $(x,y) \notin R$ because $f'_n \geq f_n$ on R. It follows that $f'_{n_i}(x,y) = W_0(x,y)$, and so $f_{n_i}(x,y) \geq W_0(x,y) + \varepsilon/4$. Thus $h_{W_0(x,y)}(f_{n_i}(x,y)) \geq \eta$, which implies

$$h_{W_0(x,y)}(f'_{n_i}(x,y)) = 0 \leq h_{W_0(x,y)}(f_{n_i}(x,y)) - \eta.$$

Recall that $h_{W_0(x,y)}(f'_{n_i}(x,y)) \leq h_{W_0(x,y)}(f_{n_i}(x,y))$ for all $(x,y) \in [0,1]^2$ and that $|S| \geq \varepsilon/4$. Therefore

$$I_{W_0}(f'_{n_i}) \leq \int_{[0,1]^2 \setminus S} h_{W_0(x,y)}(f_{n_i}(x,y)) + \int_S \left[h_{W_0(x,y)}(f_{n_i}(x,y)) - \eta\right] \leq I_{W_0}(f_{n_i}) - \frac{\varepsilon\eta}{4}.$$

Next, consider the sequence $\{\tilde{f}'_{n_i}\}_{i \geq 1}$. By the compactness of $\widetilde{\mathcal{W}}_\Omega$, there exists a convergent subsequence $\tilde{f}'_{n_{i_k}} \to \tilde{h}$ for some $\tilde{h} \in \widetilde{\mathcal{W}}_\Omega$. Since $f'_{n_i} \geq f_{n_i}$ on all relevant blocks, $t(H, f'_{n_i}) \geq t(H, f_{n_i})$, and so $t(H,h) \geq t(H,f) \geq (1+\delta)t(H, W_0)$. It follows that $J_{W_0}(\tilde{h}) \geq \min\{J_{W_0}(\tilde{g}) : t(H,g) \geq (1+\delta)t(H, W_0)\}$, which in turn implies

$$\begin{aligned} J_{W_0}(\tilde{f}) = \liminf I_{W_0}(f_{n_{i_k}}) &\geq \liminf I_{W_0}(f'_{n_{i_k}}) + \frac{\varepsilon\eta}{4} \geq J_{W_0}(\tilde{h}) + \frac{\varepsilon\eta}{4} \\ &\geq \min\{J_{W_0}(\tilde{g}) : t(H,g) \geq (1+\delta)t(H, W_0)\} + \frac{\varepsilon\eta}{4} = J_{W_0}(\tilde{f}) + \frac{\varepsilon\eta}{4}. \end{aligned}$$

We thus have reached a contradiction. □

Proof of Lemma 79. By Lemma 81, there exists δ_0, ε such that if $f \in \mathcal{W}_\Omega$, $f \geq W_0$ pointwise, $f = W_0$ on irrelevant blocks (Definition 53), and $t(H,f) \leq (1+\delta_0)t(H, W_0)$, then f satisfies the ε-neighbourhood minorant condition. Let $0 < \delta < \delta_0$, and let $\tilde{f}$ be a minimiser of the variational problem (6) with $t = (1+\delta)t(H, W_0)$.

Lemmas 55 and 65 imply that $t(H,f) = (1+\delta)t(H, W_0)$. By Lemma 82, there exists a sequence $\{f_n\}_{n \geq 1}$ such that $f_n \in \mathcal{W}_\Omega$, $f_n \geq W_0$ pointwise, $f_n = W_0$ on irrelevant blocks, $I_{W_0}(f_n) \to J_{W_0}(\tilde{f})$, and $\delta_\square(\tilde{f}, f_n) \to 0$. Since $\delta_\square(\tilde{f}, f_n) \to 0$, $t(H, f_n) \to (1+\delta)t(H, W_0)$. Since $\delta < \delta_0$, there exists n_0

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

such that for all $n \geq n_0$, $t(H, f_n) \leq (1+\delta_0)t(H, W_0)$. It follows by the assumption on δ_0 that for all $n \geq n_0$, f_n satisfies the ε-neighbourhood minorant condition. □

The next proposition derives crucial properties for the minimiser of the relative entropy problem, which will be useful in the subsequent analysis.

Proposition 83. *Let $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$, a finite regular graph H, $\tau = t(H, \cdot)$, and $t(H, W_0) \leq t \leq t^\tau_{\max}(\widetilde{\mathcal{W}}_\Omega)$. If f is a minimiser of*

$$\min\{I_{W_0}(g) : g \in \mathcal{B}^\gamma, t(H, g) \geq t\},$$

then $W_0 \leq f$ pointwise, with equality on the irrelevant set. Furthermore, $t(H, f) = t$.

Proof. Let $W_0 = (p_{ij})_{i,j\in[m]}$ and let $f = (\alpha_{ij})_{i,j\in[m]}$ be a minimiser. First, if $\alpha_{ij} < p_{ij}$, we can decrease I_{W_0} while maintaining the constraint $t(H, g) \geq t$ by increasing α_{ij}. Next observe that if $I_i \times I_j$ is irrelevant, then $t(H, f)$ does not depend on f_{ij}; since $h_p(\beta)$ has a unique minimum at $\beta = p$, this implies $f_{ij} = p_{ij}$.

To prove the last statement, suppose for contradiction that $t(H, f) > t$. There exists a relevant block $I_a \times I_b$ such that $\alpha_{ab} > p_{ab}$. Let R be the union of the relevant blocks. Let f_β be the following graphon

$$f_\beta(x, y) = \begin{cases} \beta & (x, y) \in I_a \times I_b \\ f(x, y) & (x, y) \in R \setminus I_a \times I_b \\ W_0(x, y) & \text{otherwise.} \end{cases}$$

Since $h_p(\beta)$ is strictly increasing for $\beta \in [p, 1]$, $I_{W_0}(f_\beta)$ is strictly increasing for $\beta \geq p_{ab}$. Combined with the continuity of $t(H, f_\beta)$ as a function of β, we conclude that there exists $\beta \in [p_{ab}, \alpha_{ab})$ such that $t(H, f_\beta) > t$ and $I_{W_0}(f) > I_{W_0}(f_\beta)$. This is a contradiction. □

4.3 A symmetric regime for larger δ

We now establish the existence of a symmetric regime for larger δ (Theorem 20).

Lemma 84. *Fix $m \in \mathbb{Z}^+$, $\gamma \in \Delta_m$, $W_0 \in \mathcal{B}^\gamma$ and H a finite d-regular graph. There exists $\delta = \delta(W_0, H), \varepsilon = \varepsilon(W_0) > 0$ for which $(1+\delta)t(H, W_0) < \max_{g\in\mathcal{W}_\Omega} t(H, g)$ such that the following is true: If $f \in \mathcal{W}_\Omega, f \geq W_0$ pointwise, $f = W_0$ on irrelevant blocks, and $t(H, f) \geq (1+\delta)t(H, W_0)$, then f satisfies the ε-neighbourhood minorant condition.*

Proof. Let $W_0 = (p_{ij})_{i,j\in[m]}$. By Proposition 98, for $p_{ij} \in (0, 1)$ the function $\psi_{p_{ij}}:(0, 1) \to \mathbb{R}$ is either convex or $\hat{\psi}_{p_{ij}}$ is constructed by replacing $\psi_{p_{ij}}$ with its lower common tangent on exactly one interval. If $\psi_{p_{ij}}$ is convex, then the minorant condition is trivially satisfied for the (i, j) block. Otherwise, let $\underline{q}_{ij}$ and $\overline{q}_{ij}$ be such that $\underline{q}^d_{ij}$ and $\overline{q}^d_{ij}$ are respectively the left and right points of intersection between $\psi_{p_{ij}}$ and its lower common tangent. Let $P = \{(i, j) : \psi''_{p_{ij}}(x) = 0 \text{ for some } x \in [p^d_{ij}, 1]\}$, let R denote the relevant blocks, and define

$$\varepsilon_1 = \min_{(i,j)\in P} \left(1 - \overline{q}_{ij}\right)^d$$

$$\varepsilon_2 = \min_{(i.j)\in P\setminus R} \underline{q}^d_{ij} - p^d_{ij}$$

$$\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}.$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Since $\psi'_{p_{ij}}(p_{ij}^d) = 0$ and $\psi'_{p_{ij}}(1) = \infty$ for all i, j, we have $\underline{q}_{ij} > p_{ij}$ and $\overline{q}_{ij} < 1$ for $(i,j) \in P$, so that $\varepsilon > 0$. Let g_{ij} be the graphon in $\mathcal{W}_\Omega$ that takes value $\overline{q}_{ij} + \varepsilon^{1/d}$ on the block $I_i \times I_j$ and takes value 1 on the rest of Ω. Let δ be such that $(1+\delta)t(H, W_0) = \max_{(i,j)\in P} t(H, g_{ij})$.

Suppose $f \in \mathcal{W}_\Omega$, $f \geq W_0$ pointwise, $f = W_0$ on irrelevant blocks, and $t(H,f) \geq (1+\delta)t(H, W_0) \geq t(H, g_{ij})$. Lemma 73 implies that for $(i,j) \in P$, $t(H, g_{ij}) \leq t(H,f) \leq t(H, f^*)$ where $f^* = (\|f_{ij}\|_d)_{i,j\in[m]}$ is the d-averaged graphon. Let $(a,b) \in P$. When $I_a \times I_b$ is irrelevant, $\|f_{ab}\|_d^d = p_{ab}^d \leq \underline{q}_{ab}^d - \varepsilon$. Since $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ on $[0, \underline{q}_{ab}^d]$, $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ in an ε-neighbourhood around $\|f_{ab}\|_d^d$. Next suppose $I_a \times I_b$ is relevant. Since $t(H, f^*) \geq t(H, g_{ab})$ and $g_{ab} \geq f^*_{ab}$ on $[0,1]^2 \setminus (I_a \times I_b)$, f^* must be greater than g_{ab} on $I_a \times I_b$. It follows that $\|f_{ab}\|_d \geq \overline{q}_{ab} + \varepsilon^{1/d}$, and so $\|f_{ab}\|_d^d \geq \overline{q}_{ab}^d + \varepsilon$. Since $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ on $[\overline{q}_{ab}^d, 1]$, $\psi_{p_{ab}} = \hat{\psi}_{p_{ab}}$ in an ε-neighbourhood around $\|f_{ab}\|_d^d$. □

Proof of Theorem 20. Let $W_0 \in \mathcal{B}^{\gamma,*}$. By Lemma 84, there exist $\delta, \varepsilon > 0$ such that if $f \in \mathcal{W}_\Omega$, $f \geq W_0$ pointwise, $f = W_0$ on irrelevant blocks, and $t(H,f) \geq (1+\delta)t(H, W_0)$, then f satisfies the ε-neighbourhood minorant condition. Let η be such that $(1+\delta)t(H, W_0) = (1-\eta)t_{\max}$. Since $(1+\delta)t(H, W_0) < \max_{g\in\mathcal{W}_\Omega} t(H, g)$, it holds that $\eta > 0$. Let $t \in ((1-\eta)t_{\max}, t_{\max}]$

Let g be a minimiser of the variational problem (6) with this value of t. Lemma 82 implies that there exists a sequence $g_n \in \mathcal{W}_\Omega$ such that $I_{W_0}(g_n) \to J_{W_0}(\tilde{g})$, each $g_n \geq W_0$ pointwise, and $\delta_\square(g_n, \tilde{g}) \to 0$. It follows that $t(H, g_n) \to t(H, g) \geq t$. Since $t > (1+\delta)t(H, W_0)$, there exists some n_0 such that for all $n \geq n_0$, $t(H, g_n) \geq (1+\delta)t(H, W_0)$. It follows by Lemma 84 that for all $n \geq n_0$, g_n satisfies the ε-neighbourhood minorant condition. Lemma 72 implies that $\tilde{g} \in \mathcal{B}^\gamma$, which establishes symmetry in this case. The rest of the proof is the same as that of Theorem 19. □

5. Symmetry breaking in special cases

In previous sections we let $\gamma \in \Delta_m$ and defined $\mathcal{B}^\gamma$ as the set of block graphons in which the interval structure is given by the vector γ. In this section, we let $\gamma \in (0,1) \cap \mathbb{Q}$ and let $\mathcal{B}^{(\gamma,1-\gamma)}$ denote the set of block graphons with two intervals, the first of length γ and the second of length $1-\gamma$.

Recall the definition

$$f^\gamma_{p,q,r}(x,y) = \begin{cases} p & \text{if } (x,y) \in [0,\gamma]^2 \\ r & \text{if } (x,y) \in (\gamma,1]^2 \\ q & \text{otherwise.} \end{cases}$$

We prove Theorem 23, which establishes the existence of a non-symmetric regime for graphons of the form $f^\gamma_{0,p,p}$, $f^\gamma_{1,p,p}$, and $f^\gamma_{1,p,0}$ when p is sufficiently small.

Lemma 85. *Let $\gamma \in (0,1) \cap \mathbb{Q}$ and $W_0 \in \mathcal{B}^{(\gamma,1-\gamma)}$ be a graphon with $Im(W_0) \in \{0, p, 1\}$. If $g \in \mathcal{W}_\Omega$, then $J_{W_0}(\tilde{g}) = I_{W_0}(g)$. Moreover, if τ is a continuous graph parameter and W_0 is a graphon of the form $f^\gamma_{z,p,p}$ with $z \in \{0,1\}$, then*

$$\min\{J_{W_0}(\tilde{f}) : \tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)},\ \tau(\tilde{f}) \geq t\} = \min\{I_{W_0}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)},\ \tau(f) \geq t\},$$

provided $t \in \mathbb{R}$ is such that the above minima are finite. If W_0 is a graphon of the form $f^\gamma_{z_1,p,z_2}$ with $z_1, z_2 \in \{0,1\}$, then

$$\min\{J_{W_0}(\tilde{f}) : \tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)},\ \tau(\tilde{f}) \geq t\} = \min\{I_{W_0}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)},\ \tau(f) \geq t\},$$

again provided $t \in \mathbb{R}$ is such that these minima are finite.

Proposition 43 establishes that sets of the form $\{\tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)} : \tau(\tilde{f}) \geq t\}$ are compact under $\delta_\square$, and therefore the above left-hand minima are well-defined. Since $f \in \mathcal{B}^{(\gamma,1-\gamma)}$ can be identified

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

with $[0,1]^3$, the set $\mathcal{B}^{(\gamma,1-\gamma)}$ is compact with respect the Euclidean metric. Since I_{W_0} is continuous, it follows that the righthand minima are well-defined.

Lemma 86. *Let $\gamma \in (0,1) \cap \mathbb{Q}$ and H be a d-regular graph with v vertices. Assume that*

1. *$0 < t < t(H, f^{\gamma}_{0,0,1})$ and denote $W_p = f^{\gamma}_{0,p,p}$, $\mathscr{C}(\gamma) = \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}$, or*
2. *$t(H, f^{\gamma}_{1,0,0}) < t < 1$ and denote $W_p = f^{\gamma}_{1,p,p}$, $\mathscr{C}(\gamma) = \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}$, or*
3. *$t(H, f^{\gamma}_{1,0,0}) < t < t(H, f^{\gamma}_{1,1,0})$ and denote $W_p = f^{\gamma}_{1,p,0}$, $\mathscr{C}(\gamma) = \mathcal{B}^{(\gamma,1-\gamma)}$.*

Separately, under each of these assumptions, there exists $p_0 > 0$ such that if $p < p_0$,

$$\inf\{I_{W_p}(f) : t(H,f) \geq t\} < \min\{I_{W_p}(f) : f \in \mathscr{C}(\gamma),\ t(H,f) \geq t\}.$$

Proof of Theorem 23. We apply Lemma 86 to conclude that there exists $p_0 > 0$ such that if $p < p_0$,

$$\inf\{I_{W_p}(f) : t(H,f) \geq t\} < \min\{I_{W_p}(f) : f \in \mathscr{C}(\gamma),\ t(H,f) \geq t\}.$$

Lemmas 55, 64 and 85 imply that

$$\min\{J_{W_p}(\tilde{g}) : t(H,g) \geq t\} < \min\{J_{W_p}(\tilde{f}) : \tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)},\ t(H,f) \geq t\}.$$

Therefore if $\tilde{g}$ is a minimiser of (6), then $\tilde{g} \notin \tilde{\mathcal{B}}^{(\gamma,1-\gamma)}$, meaning t is not in the symmetric regime. □

5.1 Proof of Lemma 85

Proposition 87. *Let $\gamma \in (0,1) \cap \mathbb{Q}$ and $W_0 \in \mathcal{B}^{(\gamma,1-\gamma)}$ be a graphon such that $Im(W_0) \subseteq \{0, p, 1\}$. Let $I_p(f) = \int_{[0,1]^2} h_p(f(x,y))dxdy$, and let $\Omega_q = \{(x,y) \in [0,1]^2 : W_0(x,y) = q\}$ for $q \in \{0,p,1\}$. For all $f \in \mathcal{W}_\Omega$,*

$$I_{W_0}(f) = I_p(f) - |\Omega_0| h_p(0) - |\Omega_1| h_p(1).$$

Proof. Suppose $f \in \mathcal{W}_\Omega$. Then for $q \in \{0,1\}$, $f = q$ almost everywhere on Ω_q. It follows that

$$\begin{aligned} I_{W_0}(f) &= \int_{\Omega_p} h_p(f(x,y))dxdy + \int_{\Omega_0} h_0(f(x,y))dx\,dy + \int_{\Omega_1} h_1(f(x,y))dx\,dy \\ &= \int_{\Omega_p} h_p(f(x,y))dx\,dy \\ &= \int_{[0,1]^2} h_p(f(x,y))dx\,dy - |\Omega_0| h_p(0) - |\Omega_1| h_p(1) \\ &= I_p(f) - |\Omega_0| h_p(0) - |\Omega_1| h_p(1). \end{aligned}$$

□

Lemma 88. *Let $\gamma \in (0,1) \cap \mathbb{Q}$ and $W_0 \in \mathcal{B}^{(\gamma,1-\gamma)}$. Let $\widetilde{\mathcal{W}}_\Omega$ be defined with respect to W_0.*

(i) *Suppose $W_0 = f^{\gamma}_{z,p,p}$, and $z \in \{0,1\}$. If $\tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)} \cap \widetilde{\mathcal{W}}_\Omega$, there exists g such that $\delta_\square(\tilde{f}, g) = 0$ and $g \in (\mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}) \cap \mathcal{W}_\Omega$.*

(ii) *Suppose $W_0 = f^{\gamma}_{z_1,p,z_2}$, and $z_1, z_2 \in \{0,1\}$. If $\tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)} \cap \widetilde{\mathcal{W}}_\Omega$, there exists g such that $\delta_\square(\tilde{f}, g) = 0$ and $g \in \mathcal{B}^{(\gamma,1-\gamma)} \cap \mathcal{W}_\Omega$.*

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Proof. Since $\tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)} \cap \widetilde{\mathcal{W}}_\Omega$, there exists some $g \in \mathcal{B}^{(\gamma,1-\gamma)}$ and $h \in \mathcal{W}_\Omega$ such that $\delta_\square(\tilde{f}, h) = 0$ and $\delta_\square(\tilde{f}, g) = 0$. Thus $\delta_\square(\tilde{h}, \tilde{g}) = 0$. Since g is of the form $g = \sum_{ij} \alpha_{ij} 1_{Y_i} \times 1_{Y_j}$ with $Y_1 = [0, \gamma]$ and $Y_2 = (\gamma, 1]$, we can use Lemma 44 to conclude that h must be of the same form with appropriate sets Y_1', Y_2' of sizes γ and $1 - \gamma$.

First suppose that W_0 has the form $f^\gamma_{z,p,p}$ for $z \in \{0, 1\}$. We consider cases:

(a) If $|Y_1' \cap [0, \gamma]| > 0$, then $\alpha_{11} = z$ and $g \in \mathcal{W}_\Omega$, which completes the proof.
(b) If $|Y_1' \cap [0, \gamma]| = 0$, it must be the case that $\gamma \le 1/2$ and $|Y_2' \cap [0, \gamma]| > 0$, implying that $\alpha_{22} = z$. Note that we can always re-define g on the measure zero set $(\{\gamma\} \times [0, 1]) \cup ([0, 1] \times \{\gamma\})$ so that $g^\phi \in \mathcal{B}^{(1-\gamma,\gamma)}$ for $\phi(x) = 1 - x$. Since $\gamma \le 1/2$, $g^\phi = z$ on $[0, \gamma] \times [0, \gamma]$ meaning $g^\phi \in \mathcal{W}_\Omega$. By construction, $g^\phi \in \mathcal{B}^{(1-\gamma,\gamma)}$ and $\delta_\square(g^\phi, \tilde{f}) = 0$.

Next, suppose that W_0 has the form $f^\gamma_{z_1,p,z_2}$ for $z_1, z_2 \in \{0, 1\}$. As $h \in \mathcal{W}_\Omega$, h takes value z_1 on $[0, \gamma]^2$ and value z_2 on $(\gamma, 1]^2$.

(a) If $|Y_1' \cap [0, \gamma]| > 0$ and $|Y_2' \cap (\gamma, 1]| > 0$, then $\alpha_{11} = z_1$, $\alpha_{22} = z_2$ and again $g \in \mathcal{W}_\Omega$.
(b) If $|Y_1' \cap [0, \gamma]| = 0$, then $\gamma \le 1/2$, $|Y_2' \cap [0, \gamma]| > 0$ and $|Y_1' \cap (\gamma, 1]| > 0$. It follows that $\alpha_{11} = z_2$ and $\alpha_{22} = z_1$.
   – Suppose $\gamma = 1/2$. As before, by re-defining g on the boundary if necessary, we note that for $\phi(x) = 1 - x$, g^ϕ takes value z_1 on $[0, \gamma] \times [0, \gamma]$, and value z_2 on $(\gamma, 1] \times (\gamma, 1]$. Thus $g^\phi \in \mathcal{W}_\Omega$. By construction $g^\phi \in \mathcal{B}^{(\gamma,1-\gamma)}$ and $\delta_\square(g^\phi, \tilde{f}) = 0$.
   – If $\gamma < 1/2$, then $|Y_2'| = 1 - \gamma = |(\gamma, 1]| > 1/2$. Thus $|Y_2' \cap (\gamma, 1]| > 0$, and $\alpha_{22} = z_2$. Thus $z_2 = z_1$ and h must take value $z_1 = z_2$ almost everywhere, meaning g does as well, so $g \in \mathcal{W}_\Omega \cap \mathcal{B}^{(\gamma,1-\gamma)}$.
(c) The case that $|Y_2' \cap (\gamma, 1]| = 0$ follows analogously to the above case. Note that if $\gamma \ne 1/2$, cases (b) and (c) only occur when $z_1 = z_2$. Therefore when $z_1 \ne z_2$, $g \in \mathcal{B}^{(\gamma,1-\gamma)} \cap \mathcal{W}_\Omega$. □

Proof of Lemma 85. First we show that if $g \in \mathcal{W}_\Omega$, then $J_{W_0}(\tilde{g}) = I_{W_0}(g)$. Let $g \in \mathcal{W}_\Omega$. There exists a sequence of graphons $\{g_n\}_{n\ge1}$ with each $g_n \in \mathcal{W}_\Omega$ such that $I_{W_0}(g_n) \to J_{W_0}(\tilde{g})$ and $\delta_\square(g_n, \tilde{g}) \to 0$. It follows that there exists a sequence $\phi_n \in \mathcal{M}$ such that $d_\square(g_n^{\phi_n}, g) \to 0$. Note that $I_p(g_n^{\phi_n}) = I_p(g_n)$. Let $c = |\Omega_0| h_p(0) + |\Omega_1| h_p(1)$. By Proposition 87, $I_{W_0}(g) = I_p(g) - c$ and $I_{W_0}(g_n) = I_p(g_n) - c$ since $g_n, g \in \mathcal{W}_\Omega$. Leveraging the lower semi-continuity of I_p with respect to $d_\square$ (Lemma 39), we obtain

$$J_{W_0}(\tilde{g}) = \liminf_{n\to\infty} I_{W_0}(g_n) = \liminf_{n\to\infty} I_p(g_n) - c = \liminf_{n\to\infty} I_p(g_n^{\phi_n}) - c \ge I_p(g) - c = I_{W_0}(g).$$

Since the definition of J_{W_0} implies that $J_{W_0}(\tilde{g}) \le I_{W_0}(g)$, it follows that $J_{W_0}(\tilde{g}) = I_{W_0}(g)$.

Next suppose W_0 is of the form $f^\gamma_{z,p,p}$ or $f^\gamma_{z_1,p,z_2}$ where $z, z_1, z_2 \in \{0, 1\}$ and $z_1 = z_2$. Clearly $\min\{I_{W_0}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)},\ \tau(f) \ge t\} \ge \min\{J_{W_0}(\tilde{g}) : \tilde{g} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)}, \tau(g) \ge t\}$. Let h be such that $J_{W_0}(\tilde{h}) = \min\{J_{W_0}(\tilde{g}) : \tilde{g} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)}, \tau(\tilde{g}) \ge t\}$. Lemma 88 implies that we may assume $h \in (\mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}) \cap \mathcal{W}_\Omega$. Observe

$$\begin{aligned}\min\{J_{W_0}(\tilde{f}) : \tilde{f} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)}, \tau(f) \ge t\} &= J_{W_0}(\tilde{h}) = I_{W_0}(h) \\ &\ge \min\{I_{W_0}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)},\ \tau(f) \ge t\}.\end{aligned}$$

This establishes the claim in these cases. The proof for $W_0 = f^\gamma_{z_1,p,z_2}$, $z_1, z_2 \in \{0, 1\}$, $z_1 \ne z_2$ is analogous. □

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

5.2 Proof of Lemma 86

Our construction of a non-symmetric graphon with lower entropy than any symmetric graphon is different for each of the three cases. Each proof uses the following proposition.

Proposition 89. *Let $\{W_p\}_{p\in(0,1)}$ be a family of graphons where W_p takes value in $\{0,p,1\}$. Assume further that the sets where W_p assumes the values $\{0,p,1\}$ is the same for all $p\in(0,1)$. For $i\in\{0,p,1\}$, let $\Omega_i=\{(x,y)\in[0,1]: W_p(x,y)=i\}$. Let E be the graph with two vertices and one edge. If $f\in\mathcal{W}_\Omega$, then*

$$\lim_{p\to 0}\frac{I_{W_p}(f)}{\log 1/p}=\frac{1}{2}\left(t(E,f)-|\Omega_1|\right).$$

Proof. First observe that for a fixed α,

$$\lim_{p\to 0}\frac{h_p(\alpha)}{\log\frac{1}{p}}=\lim_{p\to 0}\frac{\alpha\log\frac{\alpha}{p}+(1-\alpha)\log\frac{1-\alpha}{1-p}}{\log\frac{1}{p}}=\alpha. \tag{33}$$

Since $\frac{h_p(f)}{\log 1/p}$ is bounded as $p\to 0$ (using Lemma 33) and $f\in\mathcal{W}_\Omega$, the Dominated Convergence Theorem implies that

$$\lim_{p\to 0}\frac{I_{W_p}(f)}{\log 1/p}=\lim_{p\to 0}\frac{1}{2}\int_{\Omega_p}\frac{h_p(f)}{\log 1/p}=\frac{1}{2}\int_{\Omega_p}\lim_{p\to 0}\frac{h_p(f)}{\log 1/p}=\frac{1}{2}\int_{\Omega_p}f=\frac{1}{2}\left(t(E,f)-|\Omega_1|\right).$$

□

5.2.1. Erdős–Rényi graphs with a planted independent set

In this subsection, we prove the existence of a non-symmetric regime for d-regular subgraph counts in graphons of the form $f^{\gamma}_{0,p,p}$ when p is sufficiently small. We will do this by showing that the union of isolated vertices with a clique will have lower relative entropy than the minimum in $\mathcal{B}^{(\gamma,1-\gamma)}\cup\mathcal{B}^{(1-\gamma,\gamma)}$.

Proof of Lemma 86, Statement 1. Let $W_p=f^{\gamma}_{0,p,p}$. First note that

$$\begin{aligned}&\min\{I_{W_p}(f): f\in\mathcal{B}^{(\gamma,1-\gamma)}\cup\mathcal{B}^{(1-\gamma,\gamma)},\ t(H,f)\geq t\}\\&=\min\{I_{W_p}(f^z_{0,\alpha,\beta}):\alpha,\beta\in[p,1], z\in\{\gamma,1-\gamma\},\ t(H,f^z_{0,\alpha,\beta})\geq t\}.\end{aligned} \tag{34}$$

We can restrict to graphons of the form $f^z_{0,\alpha,\beta}$ since $I_{W_p}(f^z_{\eta,\alpha,\beta})=\infty$ when $\eta\neq 0$. We can further restrict to $\alpha,\beta\geq p$ since $h_p(\cdot)$ is decreasing on $[0,p]$ and $t(H,\cdot)$ is an increasing function.

Define the non-symmetric graphon χ_t as follows.

$$\chi_t(x,y)=\begin{cases}1 & (x,y)\in[1-t^{\frac{1}{\nu}},1]^2\\ 0 & \text{otherwise.}\end{cases}$$

In other words, the graphon χ_t is the union of a clique and isolated vertices that has the required subgraph density (by the fact that $t(H,\chi_t)=t$). Note that the assumption that $t\leq t(H,f^{\gamma}_{0,0,1})$ implies that $t^{1/\nu}\leq 1-\gamma$ and so $\chi_t\in\mathcal{W}_\Omega$. By (34), it suffices to show

$$\lim_{p\to 0}\frac{I_{W_p}(\chi_t)}{\log\frac{1}{p}}<\lim_{p\to 0}\frac{I_{W_p}(f^z_{0,\alpha,\beta})}{\log\frac{1}{p}} \tag{35}$$

for each triple $z\in\{\gamma,1-\gamma\}$, $\alpha,\beta\in[p,1]$ such that $t(H,f^z_{0,\alpha,\beta})\geq t$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Let E be the graph with two vertices and one edge. By Proposition 89, for $z \in \{\gamma, 1-\gamma\}$

$$\lim_{p\to 0} \frac{I_{W_p}(f^z_{0,\alpha,\beta})}{\log \frac{1}{p}} = \frac{1}{2} t(E, f^z_{0,\alpha,\beta}) \quad \text{and} \quad \lim_{p\to 0} \frac{I_{W_p}(\chi_t)}{\log \frac{1}{p}} = \frac{1}{2} t(E, \chi_t) = \frac{1}{2} t^{\frac{2}{v}}.$$

Therefore to establish (35), it suffices to show that

$$t^{\frac{2}{v}} \le t(H, f^z_{0,\alpha,\beta})^{\frac{2}{v}} < t(E, f^z_{0,\alpha,\beta}).$$

By the generalised Hölder inequality (Theorem 97) and the facts that $E(H) = dv/2$ and $g^d \le g$ for any graphon g,

$$t(H,g) = \int_{[0,1]^v} \prod_{(i,j)\in E(H)} g(x_i, x_j) dx_1 \dots dx_v \le \left(\int_{[0,1]^2} \big(g(x,y)\big)^d \, dx\, dy \right)^{\frac{e(H)}{d}}$$

$$\le \left(\int_{[0,1]^2} g(x,y)\, dx\, dy \right)^{\frac{v}{2}} = t(E,g)^{\frac{v}{2}} \tag{36}$$

The second inequality is strict if g takes values in $(0,1)$ on a set with positive measure, as is the case for all graphons in the set of potential minimisers described on the right hand side of (34), with the exception of graphons of the form $f^z_{0,1,1}$. We claim that for such graphons, the first inequality is strict. When we apply Theorem 97, we take $f_1(x_i, x_j) = f^z_{0,1,1}(x_i, x_j)$. To have equality, f_1 must have a product representation $f_1(x_i, x_j) = f_{1i}(x_i) f_{1j}(x_j)$. Suppose such a representation exists. For $x_i, x_j \in [0, z)^2$, $f_1(x_i, x_j) = 0$, and so $f_{1i}(x_i) = 0$ or $f_{1j}(x_j) = 0$. However, then $f_1(x_i, \cdot) = 0$ or $f_1(\cdot, x_j) = 0$ on a set of positive measure, which is not consistent with the structure of the graphon $f^z_{0,1,1}$. Finally, taking $g = f^z_{0,\alpha,\beta}$ and rearranging establishes (35). □

5.2.2. Erdős–Rényi graphs with a planted clique

In this subsection, we prove the existence of a non-symmetric regime for d-regular subgraph counts in graphons of the form $f^\gamma_{1,p,p}$ when p is sufficiently small. Again it will be the union of a clique and isolated vertices which has lower relative entropy than the minimiser in $\mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}$.

Proof of Lemma 86, Statement 2. Let $W_p = f^\gamma_{1,p,p}$. First note that

$$\min\{I_{W_p}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)},\ t(H,f) \ge t\}$$

$$= \min\{I_{W_p}(f^z_{1,\alpha,\beta}) : z \in \{\gamma, 1-\gamma\}, \alpha, \beta \in [p,1],\ t(H, f^z_{1,\alpha,\beta}) \ge t\}, \tag{37}$$

by similar reasoning to the proof of Lemma 86, Statement 1. Define the non-symmetric graphon χ_t as follows.

$$\chi_t(x,y) = \begin{cases} 1 & (x,y) \in [0, t^{\frac{1}{v}}]^2 \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the graphon χ_t is the union of a clique and isolated vertices that has the required subgraph density. Note that since $t \ge t(H, f^\gamma_{1,0,0})$, $t^{1/v} \ge \gamma$ and so $\chi_t \in \mathcal{W}_\Omega$. By (37), it suffices to show that

$$\lim_{p\to 0} \frac{I_{W_p}(\chi_t)}{\log \frac{1}{p}} < \lim_{p\to 0} \frac{I_{W_p}(f^\gamma_{1,\alpha,\beta})}{\log \frac{1}{p}}$$

for each triple $z \in \{\gamma, 1-\gamma\}, \alpha, \beta \in [p,1]$ such that $t(H, f^z_{1,\alpha,\beta}) \ge t$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Let E be the graph with two vertices and one edge. By Proposition 89, for $z \in \{\gamma, 1-\gamma\}$

$$\lim_{p\to 0} \frac{I_{W_p}(f^z_{1,\alpha,\beta})}{\log \frac{1}{p}} = \frac{t(E, f^z_{1,\alpha,\beta}) - \gamma^2}{2} \quad \text{and} \quad \lim_{p\to 0} \frac{I_{W_p}(\chi_t)}{\log \frac{1}{p}} = \frac{t(E,\chi_t) - \gamma^2}{2} = \frac{t^{\frac{2}{v}} - \gamma^2}{2}.$$

Therefore, it suffices to show that

$$\frac{1}{2}\left(t^{\frac{2}{v}} - \gamma^2\right) < \frac{1}{2}\left(t(E, f^z_{1,\alpha,\beta}) - \gamma^2\right) \impliedby t^{\frac{2}{v}} \le t(H, f^z_{1,\alpha,\beta})^{\frac{2}{v}} < t(E, f^z_{1,\alpha,\beta}).$$

This holds by (36), as long as one of α or β is not equal to 1. Finally, note that $f^z_{1,1,1}$ is clearly not an optimiser since $t(H, f^z_{1,1,1}) > t$ and so we can always produce $g \in \mathcal{B}^{(z,1-z)} \setminus \{f^z_{1,1,1}\}$ such that $t(H,g) \ge t$ and $I_{W_p}(g) < I_{W_p}(f^z_{1,1,1})$. □

5.2.3. Erdős–Rényi graphs with a planted clique and independent set

In this subsection, we prove the existence of a non-symmetric regime for d-regular subgraph counts in graphons of the form $f^\gamma_{1,p,0}$ when p is sufficiently small. This time, it will be the union of a clique, a bipartite complete graph and isolated vertices which has lower relative entropy than the minimiser in $\mathcal{B}^{(\gamma,1-\gamma)}$.

Proof of Lemma 86, Statement 3. Let $W_p = f^\gamma_{1,p,0}$ and $t(H, f^\gamma_{1,0,0}) < t < t(H, f^\gamma_{1,1,0})$. First note that

$$\begin{aligned}
&\min\{I_{W_p}(f) : f \in \mathcal{B}^{(\gamma,1-\gamma)},\ t(H,f) \ge t\} \\
&= \min\{I_{W_p}(f^\gamma_{1,\alpha,0}) : \alpha \in [0,1],\ t(H, f^z_{1,\alpha,0}) \ge t\} \\
&= \min\{I_{W_p}(f^\gamma_{1,\alpha,0}) : \alpha \in [0,1],\ t(H, f^z_{1,\alpha,0}) = t\}. \qquad (38)
\end{aligned}$$

The first equality follows because $I_{W_p}(f^z_{\eta,\alpha,\beta}) = \infty$ when $\eta \ne 1$ or $\beta \ne 0$. The second equality follows by Proposition 83.

We construct a non-symmetric graphon χ_α such that $t(H, \chi_\alpha) = t(H, f^\gamma_{1,\alpha,0})$. Let

$$\chi_\alpha(x,y) = \begin{cases} 1 & (x,y) \in [0, \gamma + (1-\gamma)\alpha^d]^2 \setminus (\gamma, 1]^2 \\ 0 & \text{otherwise.} \end{cases}$$

Let s_k be the number of labelled independent sets of size k in H, and let $v = |V(H)|$ be the number of vertices of H. In any homomorphism of H in $f^\gamma_{1,\alpha,0}$, the vertices of H mapped to the interval $(\gamma, 1]$ must form an independent set. Counting homomorphisms by the number of vertices that map to $(\gamma, 1]$, we obtain

$$t(H, f^\gamma_{1,\alpha,0}) = \sum_{k=0}^{v} s_k \gamma^{v-k}(1-\gamma)^k \alpha^{dk} = \sum_{k=0}^{v} s_k \gamma^{v-k}((1-\gamma)\alpha^d)^k 1^{dk} = t(H, \chi_\alpha).$$

To establish symmetry breaking, it suffices to show that χ_α has lower entropy than the class of symmetric graphons $f^\gamma_{1,\alpha,0}$. Thus by (38), it is enough to show that

$$\lim_{p\to 0} \frac{I_{W_p}(\chi_\alpha)}{\log \frac{1}{p}} < \lim_{p\to 0} \frac{I_{W_p}(f^\gamma_{1,\alpha,0})}{\log \frac{1}{p}} \qquad (39)$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

for all α such that $t(H, f^{\gamma}_{1,\alpha,0}) = t$. Since $t(H, f^{\gamma}_{1,0,0}) < t < t(H, f^{\gamma}_{1,1,0})$, $t(H, f^{z}_{1,\alpha,0}) \neq t$ when $\alpha \in \{0, 1\}$. Thus, it suffices to establish (39) when $\alpha \in (0, 1)$.

Observe using (33) that

$$\lim_{p\to 0} \frac{I_{W_p}(\chi_\alpha)}{\log \frac{1}{p}} = \lim_{p\to 0} \frac{(1-\gamma)\alpha^d \gamma h_p(1) + (1-\gamma-(1-\gamma)\alpha^d)\gamma h_p(0)}{\log \frac{1}{p}} = (1-\gamma)\gamma\alpha^d$$

and

$$\lim_{p\to 0} \frac{I_{W_p}(f^{\gamma}_{1,\alpha,0})}{\log \frac{1}{p}} = \lim_{p\to 0} \frac{(1-\gamma)\gamma h_p(\alpha)}{\log \frac{1}{p}} = (1-\gamma)\gamma\alpha.$$

Noting that $0 < \alpha < 1$ establishes (39), and completes the proof. □

6. Bipartite Erdős–Rényi graphs

In this section, we prove Theorems 24 and 28, which precisely identify the symmetric and non-symmetric regimes for d-regular subgraph counts and the operator norm in bipartite Erdős–Rényi graphs respectively. Throughout this section, we fix $p \in (0, 1)$ and $\gamma \in (0, 1) \cap \mathbb{Q}$. We use the notation f^{γ}_{p} to denote the bipartite graphon with density p and blocks of size γ and $1-\gamma$, as illustrated in Figure 2.

6.1 Density of d-regular subgraphs

We will apply Lemmas 90 and 91 to identify the symmetric and non-symmetric regimes respectively.

Lemma 90. *Let $p \in (0, 1)$, $\gamma \in (0, 1) \cap \mathbb{Q}$ and $W_0 = f^{\gamma}_{p}$. Let H be a d-regular graph with $d \geq 1$. Let $0 < p \leq r \leq 1$ be such that $(r^d, h_p(r))$ is on the convex minorant of ψ_p. If $f \in \mathcal{W}_\Omega$ and $t(H, f) \geq t(H, f^{\gamma}_{r})$, then $I_{W_0}(f) \geq I_{W_0}(f^{\gamma}_{r})$ with equality if and only if $f = f^{\gamma}_{r}$ almost everywhere.*

Lemma 91. *Let $p \in (0, 1)$, $\gamma \in (0, 1) \cap \mathbb{Q}$ and $W_0 = f^{\gamma}_{p}$. Let H be a d-regular graph with $d \geq 1$. Let $0 < p < r < 1$ be such that $(r^d, h_p(r))$ is not on the convex minorant of ψ_p. Then there exists $g \in \mathcal{W}$ such that $t(H, g) > t(H, f^{\gamma}_{r})$ and $I_{W_0}(g) < I_{W_0}(f^{\gamma}_{r})$.*

We now prove Theorem 24, which completely characterises the symmetric and non-symmetric regimes for d-regular homomorphism densities in bipartite Erdős–Rényi graphons.

Proof of Theorem 24. Suppose that the point $(r^d, h_p(r))$ lies on the convex minorant of ψ_p. We will show that $t^{\gamma}_{r} = t(H, f^{\gamma}_{r})$ is in the symmetric regime for $t(H, \cdot)$. Let $\tilde{g} \in \widetilde{\mathcal{W}}_\Omega$ be such that $J_{W_0}(\tilde{g}) = \min\{J_{W_0}(\tilde{f}) : t(H, f) \geq t^{\gamma}_{r}\}$. We may assume that $g \in \mathcal{W}_\Omega$. By Lemma 85, $J_{W_0}(\tilde{g}) = I_{W_0}(g)$. Since $t(H, g) \geq t^{\gamma}_{r}$ and $(r^d, h_p(r))$ lies on the convex minorant of ψ_p, Lemma 90 implies that $I_{W_0}(g) \geq I_{W_0}(f^{\gamma}_{r})$. Since $I_{W_0}(g) = J_{W_0}(\tilde{g}) = \min\{J_{W_0}(\tilde{g}) : t(H, g) \geq t^{\gamma}_{r}\} \leq I_{W_0}(f^{\gamma}_{r})$, it follows that $I_{W_0}(f^{\gamma}_{r}) = I_{W_0}(g)$. Lemma 90 implies that $g = f^{\gamma}_{r}$, meaning that f^{γ}_{r} is the unique symmetric solution.

Next, suppose that the point $(r^d, h_p(r))$ does not lie on the convex minorant of ψ_p. We will show that $t^{\gamma}_{r} = t(H, f^{\gamma}_{r})$ is not in the symmetric regime for $t(H, \cdot)$. Lemma 91 implies that there exists $g \in \mathcal{W}_\Omega$ such that $t(H, g) > t(H, f^{\gamma}_{r})$ and $I_{W_0}(g) < I_{W_0}(f^{\gamma}_{r})$. By Lemma 85, $J_{W_0}(\tilde{g}) = I_{W_0}(g)$. We apply Lemma 85 and obtain

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$$\begin{aligned}\min\{J_{W_0}(\tilde{g}) : t(H,g) \geq t_r^\gamma, \tilde{g} \in \tilde{\mathcal{B}}^{(\gamma,1-\gamma)}\} &= \min\{I_{W_0}(g) : t(H,g) \geq t_r^\gamma, g \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}\} \\ &= \min\{I_{W_0}(f_q^\gamma) : q \in [0,1], t(H,f_q^\gamma) \geq t_r^\gamma\} = I_{W_0}(f_r^\gamma) \\ &> I_{W_0}(g) \geq \min\{J_{W_0}(\tilde{g}) : t(H,g) \geq t_r^\gamma\}.\end{aligned}$$

The second equality follows by noting that if $\gamma \neq 1/2$ and $g \in \mathcal{B}^{(1-\gamma,\gamma)}$, then $I_{W_0}(g) = \infty$ or g is the zero graphon. The third equality follows by noting that $I_{W_0}(f_q^\gamma)$ and $t(H,f_q^\gamma)$ are increasing functions of q. □

6.1.1. Proof for the symmetric regime

The following lemma describes a norm condition on f that implies that the graphon f_r^γ has lower entropy.

Lemma 92. *Suppose that $d \geq 1$ and $p \leq r \leq 1$ are such that the point $(r^d, h_p(r))$ lies on the convex minorant of ψ_p and*

$$\|f\|_d^d \geq 2\gamma(1-\gamma)r^d.$$

Then $I_{W_0}(f) \geq I_{W_0}(f_r^\gamma)$, with equality occurring if and only if $f = f_r^\gamma$ almost everywhere.

Proof. The statement is trivial if $f \notin \mathcal{W}_\Omega$. For $f \in \mathcal{W}_\Omega$,

$$\begin{aligned}I_{W_0}(f) &= \int_0^\gamma \int_\gamma^1 h_p(f(x,y))dxdy = \int_0^\gamma \int_\gamma^1 \psi_p\left(f^d(x,y)\right)dxdy \\ &\geq \int_0^\gamma \int_\gamma^1 \hat{\psi}_p\left(f^d(x,y)\right)dxdy \\ &\geq \gamma(1-\gamma)\hat{\psi}_p\left(\frac{1}{\gamma(1-\gamma)}\int_0^\gamma \int_\gamma^1 f^d(x,y)dxdy\right) \qquad (40) \\ &= \gamma(1-\gamma)\hat{\psi}_p\left(\frac{1}{\gamma(1-\gamma)}\frac{\|f\|_d^d}{2}\right) \\ &\geq \gamma(1-\gamma)\hat{\psi}_p(r^d) \\ &= \gamma(1-\gamma)\psi_p(r^d) = \gamma(1-\gamma)h_p(r) = I_{W_0}(f_r^\gamma). \qquad (41)\end{aligned}$$

Note that (40) is an application of Jensen's inequality, and (41) is due to $\hat{\psi}_p$ being an increasing function on $[r^d, 1]$. If $f \neq f_r^\gamma$, then the step using Jensen's inequality is a strict inequality. Therefore, $I_{W_0}(f) \geq I_{W_0}(f_r^\gamma)$, with equality occurring if and only if $f = f_r^\gamma$ almost everywhere. □

The following lemma establishes a norm condition on graphons that satisfy the subgraph density requirement.

Lemma 93. *Let $p, r \in (0,1)$, $\gamma \in (0,1) \cap \mathbb{Q}$ and $W_0 = f_p^\gamma$. Let H be a d-regular graph with $d \geq 1$. Let $f \in \mathcal{W}_\Omega$ be such that $t(H,f) \geq t(H,f_r^\gamma)$. Then $\|f\|_d^d \geq 2\gamma(1-\gamma)r^d$.*

Proof. We may assume H is bipartite. Since H is d-regular, H must have m vertices in each partition class and dm edges for some $m \in \mathbb{Z}^+$. Let c be the number of connected components of H. Note that $t(H,f_r^\gamma) = 2^c(\gamma(1-\gamma))^m r^{dm}$. Let f be any graphon such that $I_{W_0}(f) < \infty$ and $t(H,f) \geq t(H,f_r^\gamma)$. Then $t(H,f_r^\gamma) \leq t(H,f)$ implies

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$$t(H,f_r^\gamma) = 2^c\gamma^m(1-\gamma)^m r^{dm} \le \int_{[0,1]^{2m}} \prod_{(i,j)\in E(H)} f(x_i,x_j)dx_1,\dots,dx_{2m}$$

$$= 2^c \int_0^\gamma \int_\gamma^1 \cdots \int_0^\gamma \int_\gamma^1 \prod_{(i,j)\in E(H)} f(x_i,x_j)dx_1,\dots,dx_{2m} \tag{42}$$

$$\le 2^c\gamma^m(1-\gamma)^m \prod_{(i,j)\in E(H)} \left(\frac{1}{\gamma(1-\gamma)}\int_{x_i=0}^\gamma \int_{x_j=\gamma}^1 f(x_i,x_j)^d dx_j\,dx_i\right)^{\frac{1}{d}} \tag{43}$$

$$= 2^c\gamma^m(1-\gamma)^m \left(\frac{1}{\gamma(1-\gamma)}\int_{x=0}^\gamma \int_{y=\gamma}^1 f(x,y)^d dy\,dx\right)^{\frac{md}{d}}$$

$$= 2^c \left(\int_{x=0}^\gamma \int_{y=\gamma}^1 f(x,y)^d dy\,dx\right)^m = 2^c\left(\frac{1}{2}\|f\|_d^d\right)^m.$$

In (42) we rewrite the density by ordering the vertices so that they alternate between the sides of the bipartition. The factor 2^c accounts for the fact that within each component, a partition class of vertices can map to either $[0,\gamma]$ or $(\gamma,1]$, and the other partition class will map to the other interval. The generalised Hölder inequality from Theorem 97 implies (43). In the application of Theorem 97, we set $p_i = d$ for every $i \in [2m]$. The Radon–Nikodym derivatives of the measures are given by

$$\frac{d\mu_{2k+1}}{dx} = \begin{cases} \frac{1}{\gamma} & 0 \le x \le \gamma \\ 0 & \gamma < x \le 1 \end{cases} \quad \text{and} \quad \frac{d\mu_{2k}}{dx} = \begin{cases} 0 & 0 \le x \le \gamma \\ \frac{1}{1-\gamma} & \gamma < x \le 1 \end{cases}$$

for $k \in \{0,1,\dots,m\}$, i.e. μ_{2k+1} is uniform on $[0,\gamma]$ and μ_{2k} is uniform on $(\gamma,1]$. The sets $A_1,\dots,A_{e(H)}$ correspond to the set $E(H)$. We conclude that $\|f\|_d^d \ge 2\gamma(1-\gamma)r^d$. □

Proof of Lemma 90. Let $f \in \mathcal{W}_\Omega$ be such that $t(H,f) \ge t(H,f_r^\gamma)$. Lemma 93 implies that $\|f\|_d^d \ge 2\gamma(1-\gamma)r^d$. It follows by Lemma 92 that $I_{W_0}(f) \ge I_{W_0}(f_r^\gamma)$ with equality if and only if $f = f_r^\gamma$. □

6.1.2. *Proof for the non-symmetric regime*

Proof of Lemma 91. Since $(r^d, h_p(r))$ is not on the convex minorant of $\psi_p(x) = h_p(r^{1/d})$, we may use Proposition 98 to conclude there exist r_1, r_2, r such that $p < r_1 < r < r_2 \le 1$ and $(r^d, h_p(r))$ lies strictly above the line segment joining $(r_1^d, h_p(r_1))$ and $(r_2^d, h_p(r_2))$. Let $s \in (0,1)$ be such that

$$r^d = sr_1^d + (1-s)r_2^d,$$

and thus

$$sh_p(r_1) + (1-s)h_p(r_2) < h_p(r). \tag{44}$$

We use the values r_1, r_2, and s to define a family of graphons $(g^\varepsilon)_{\varepsilon>0}$. We will prove that for $\varepsilon > 0$ sufficiently small (i) $t(H,g^\varepsilon) > t(H,f)$ and (ii) $I_{W_0}(g^\varepsilon) < I_{W_0}(f_r^\gamma)$. Define

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

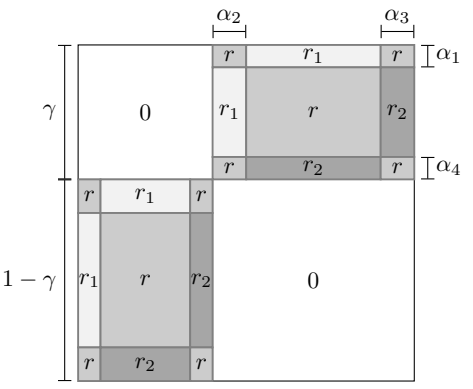


Figure 8. Construction of g^ε.

$$\begin{aligned}
\alpha_1 &= \gamma s\varepsilon^2 & I_1 &= [0, \alpha_1] \\
\alpha_2 &= (1-\gamma)s\varepsilon^2 & I_2 &= (\gamma, \gamma + \alpha_2] \\
\alpha_3 &= (1-\gamma)\left((1-s)\varepsilon^2 + \varepsilon^3\right) & I_3 &= (1-\alpha_3, 1] \\
\alpha_4 &= \gamma\left((1-s)\varepsilon^2 + \varepsilon^3\right) & I_4 &= (\gamma - \alpha_4, \gamma].
\end{aligned} \tag{45}$$

Let

$$I_{c14} = [0,\gamma] \setminus (I_1 \cup I_4) \quad \text{and} \quad I_{c23} = (\gamma, 1] \setminus (I_2 \cup I_3).$$

Define

$$g^\varepsilon(x,y) = \begin{cases} 0 & (x,y) \in ([0,\gamma) \times [0,\gamma)) \cup ((\gamma,1] \times (\gamma,1]) \\ r_1 & (x,y) \in (I_1 \times I_{c23}) \cup (I_{c23} \times I_1) \cup (I_2 \times I_{c14}) \cup (I_{c14} \times I_2) \\ r_2 & (x,y) \in (I_3 \times I_{c14}) \cup (I_{c14} \times I_3) \cup (I_4 \times I_{c23}) \cup (I_{c23} \times I_4) \\ r & \text{otherwise.} \end{cases}$$

Figure 8 illustrates the construction of the graphon g^ε.

Next we claim that $t(H, g^\varepsilon) > t(H, f_r^\gamma)$ for sufficiently small ε. Let m be such that H has $2m$ vertices and dm edges. Let c be the number of connected components of H. Note that the only embeddings of H into g^ε which contribute a value other than $r^{e(H)}$ to the integral in $t(H, g^\varepsilon)$ are such that at least one vertex of H is mapped to $\bigcup_{j=1}^4 I_j$. Since each α_i is of order ε^2, in order to compute $t(H, g^\varepsilon) - t(H, f_r^\gamma)$ up to error $O\left(\varepsilon^4\right)$ it suffices to consider embeddings in which only one vertex is mapped to $\bigcup_{j=1}^4 I_j$. Observe

$$\begin{aligned}
&t(H, g^\varepsilon) - t(H, f_r^\gamma) \\
&= 2^c \int_0^\gamma \int_\gamma^1 \cdots \int_0^\gamma \int_\gamma^1 \left(\prod_{(i,j) \in E(H)} g^\varepsilon(x_i, y_j) - r^{e(H)} \right) dx_1 \ldots dx_{2m}
\end{aligned}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$$
\begin{aligned}
&= 2^c m \Big[\alpha_1 \gamma^{m-1}(1-\gamma)^m (r_1^d - r^d) r^{e(H)-d} + \alpha_2 (1-\gamma)^{m-1}\gamma^m (r_1^d - r^d) r^{e(H)-d} \\
&\qquad + \alpha_3 (1-\gamma)^{m-1}\gamma^m (r_2^d - r^d) r^{e(H)-d} + \alpha_4 \gamma^{m-1}(1-\gamma)^m (r_2^d - r^d) r^{e(H)-d}\Big] + O\left(\varepsilon^4\right) \\
&= 2^{c+1} m \gamma^m (1-\gamma)^m r^{e(H)-d} \left(s\varepsilon^2 (r_1^d - r^d) + \left((1-s)\varepsilon^2 + \varepsilon^3\right)\left(r_2^d - r^d\right)\right) + O\left(\varepsilon^4\right) \\
&= 2^{c+1} m \gamma^m (1-\gamma)^m r^{e(H)-d} \left(r_2^d - r^d\right)\varepsilon^3 + O\left(\varepsilon^4\right).
\end{aligned}
$$

Since $r_2 > r$, the above computation implies that $t(H, g^\varepsilon) - t(H, f_r^\gamma) > 0$ for ε sufficiently small.

Next we show that $I_{W_0}(g^\varepsilon) < I_{W_0}(f_r^\gamma)$ for sufficiently small ε. Observe

$$
\begin{aligned}
&I_{W_0}(g^\varepsilon) - I_{W_0}(f_r^\gamma) \\
&= (\alpha_1 (1-\gamma-\alpha_2-\alpha_3) + \alpha_2 (\gamma - \alpha_1 - \alpha_4))\left(h_p(r_1) - h_p(r)\right) \\
&\quad + (\alpha_3 (\gamma - \alpha_1 - \alpha_4) + \alpha_4 (1-\gamma-\alpha_2-\alpha_3))\left(h_p(r_2) - h_p(r)\right) \\
&= 2\gamma(1-\gamma)\left(1-\varepsilon^2-\varepsilon^3\right)\left[s\varepsilon^2\left(h_p(r_1) - h_p(r)\right) + \left((1-s)\varepsilon^2 + \varepsilon^3\right)\left(h_p(r_2) - h_p(r)\right)\right] \\
&= 2\gamma(1-\gamma)\left(1-\varepsilon^2-\varepsilon^3\right)\varepsilon^2\left[s h_p(r_1) + (1-s)h_p(r_2) - h_p(r) + \varepsilon\left(h_p(r_2) - h_p(r)\right)\right].
\end{aligned}
$$

Using the condition (44), we conclude that there exists ε sufficiently small such that $I_{W_0}(g^\varepsilon) - I_{W_0}(f_r^\gamma) < 0$, as desired. □

6.2 *Largest eigenvalue*

In this subsection we prove Theorem 28, which characterises the symmetric and non-symmetric regimes for the largest eigenvalue of the adjacency matrix of a bipartite graph. Recall that $\|\cdot\|_{\mathrm{op}}$ is a continuous extension of the normalised graph spectral norm (Lemma 27). Note that $\|f_r^\gamma\|_{\mathrm{op}} = r\sqrt{\gamma(1-\gamma)}$. We will use the following two lemmas to prove Theorem 28.

Lemma 94. *Let $\gamma \in (0,1) \cap \mathbb{Q}$ and let $W_0 = f_p^\gamma$. For every f such that $f \in \mathcal{W}_\Omega$, we have $\|f\|_1 \le \|f\|_{op} \le \frac{1}{\sqrt{2}}\|f\|_2$.*

Lemma 95. *Let $0 < p \le r < 1$ be such that $(r^2, h_p(r))$ does not lie on the convex minorant of $x \mapsto h_p(\sqrt{x})$. Then there exists some $g \in \mathcal{W}_\Omega$ with $\|g\|_{op} > r\sqrt{\gamma(1-\gamma)}$ and $I_{W_0}(g) < I_{W_0}(f_r^\gamma)$.*

Proof of Theorem 28. Let $\psi_p(x) = h_p(\sqrt{x})$. Suppose that the point $(r^2, h_p(r))$ lies on the convex minorant of ψ_p. We will show that $t_r^\gamma = \|f_r^\gamma\|_{\mathrm{op}}$ is in the symmetric regime for $t(H,\cdot)$. Let $\tilde{g} \in \widetilde{\mathcal{W}}_\Omega$ be such that $J_{W_0}(\tilde{g}) = \min\{J_{W_0}(\tilde{g}) : \|g\|_{\mathrm{op}} \ge t_r^\gamma\}$. We may assume that $g \in \mathcal{W}_\Omega$. By Lemma 85, $J_{W_0}(\tilde{g}) = I_{W_0}(g)$. Since $\|g\|_{\mathrm{op}} \ge t_r^\gamma = r\sqrt{\gamma(1-\gamma)}$, Lemma 94 implies that $\|g\|_2 \ge r\sqrt{2\gamma(1-\gamma)}$. Next, by Lemma 92, we have $I_{W_0}(g) \ge I_{W_0}(f_r^\gamma)$ with equality if and only if $g = f_r^\gamma$. Since $I_{W_0}(g) = J_{W_0}(\tilde{g}) = \min\{J_{W_0}(\tilde{g}) : t(H,g) \ge t_r^\gamma\} \le I_{W_0}(f_r^\gamma)$, it follows that $I_{W_0}(f_r^\gamma) = I_{W_0}(g)$. By Lemma 92, we conclude that $g = f_r^\gamma$, meaning that f_r^γ is the unique symmetric solution.

Next, suppose that the point $(r^2, h_p(r))$ does not lie on the convex minorant of ψ_p. We will show that $t_r^\gamma = \|f_r^\gamma\|_{\mathrm{op}}$ is not in the symmetric regime for $t(H,\cdot)$. Lemma 95 implies that there exists $g \in \mathcal{W}_\Omega$ such that $\|g\|_{\mathrm{op}} > r\sqrt{\gamma(1-\gamma)}$ and $I_{W_0}(g) < I_{W_0}(f_r^\gamma)$. By Lemma 85, $J_{W_0}(\tilde{g}) = I_{W_0}(g)$. We apply Lemma 85 and obtain

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$$\begin{aligned}\min\{J_{W_0}(\tilde{g}) : \|g\|_{\mathrm{op}} \geq t_r^\gamma, \tilde{g} \in \tilde{\mathcal{B}}^\gamma\} &= \min\{I_{W_0}(g) : \|g\|_{\mathrm{op}} \geq t_r^\gamma, g \in \mathcal{B}^{(\gamma,1-\gamma)} \cup \mathcal{B}^{(1-\gamma,\gamma)}\} \\ &= \min\{I_{W_0}(f_q^\gamma) : q \in [0,1], \|f_q^\gamma\|_{\mathrm{op}} \geq t_r^\gamma\} = I_{W_0}(f_r^\gamma) \\ &> I_{W_0}(g) \geq \min\{J_{W_0}(\tilde{g}) : \|g\|_{\mathrm{op}} \geq t_r^\gamma\}. \end{aligned} \tag{46}$$

The second equality follows by noting that if $\gamma \neq 1/2$ and $g \in \mathcal{B}^{(1-\gamma,\gamma)}$, then $I_{W_0}(g) = \infty$ or g is the zero graphon. The third equality follows by noting that $I_{W_0}(f_q^\gamma)$ and $\|f_q^\gamma\|_{\mathrm{op}}$ are increasing functions of q.

It follows by (46) that any minimiser $\tilde{g} \notin \tilde{\mathcal{B}}^{(\gamma,1-\gamma)}$, and so the problem is not in the symmetric regime. □

Proof of Lemma 94. As stated in [34], the left inequality follows from the observation that

$$\|f\|_1 = \|T_f\mathbf{1}\|_1 \leq \|T_f\mathbf{1}\|_2 \leq \|f\|_{\mathrm{op}}.$$

To derive the upper bound, we use the Cauchy–Schwarz inequality. Observe that for any $u : [0,1] \to \mathbb{R}$,

$$\begin{aligned}\|T_f u\|_2^2 &= \int_0^1 \left(\int_0^1 f(x,y)u(y)dy\right)^2 dx \\ &= \int_0^\gamma \left(\int_\gamma^1 f(x,y)u(y)dy\right)^2 dx + \int_\gamma^1 \left(\int_0^\gamma f(x,y)u(y)dy\right)^2 dx \\ &\leq \int_\gamma^1 u(y)^2 dy \left(\int_0^\gamma \int_\gamma^1 f(x,y)^2 dydx\right) + \int_0^\gamma u(y)^2 dy \left(\int_\gamma^1 \int_0^\gamma f(x,y)^2 dydx\right) \\ &= \frac{1}{2}\|u\|_2^2\|f\|_2^2. \end{aligned}$$

It follows that $\|T_f u\|_2 \leq \frac{1}{\sqrt{2}}\|f\|_2\|u\|_2$ for all u, and thus $\|f\|_{\mathrm{op}} \leq \frac{1}{\sqrt{2}}\|f\|_2$. □

Proof of Lemma 95. Let g^ε be as defined in the proof of Lemma 91. We have already shown that $I_{W_0}(g^\varepsilon) < I_{W_0}(f_r^\gamma)$ for small enough $\varepsilon > 0$. It remains to show that $\|g^\varepsilon\|_{\mathrm{op}} > r\sqrt{\gamma(1-\gamma)}$.

For this claim, it suffices to exhibit a function $u \in L^2([0,1])$ such that $(T_{g^\varepsilon}u)(x) > r\sqrt{\gamma(1-\gamma)}u(x)$ for all $x \in [0,1]$. Recall the definitions given in (45). Let

$$u(x) = \begin{cases} \frac{\sqrt{1-\gamma}}{\gamma}(\gamma - \alpha_1 - \alpha_4)r_1 & x \in I_1 \\ r\sqrt{1-\gamma} & x \in I_{c14} \\ \frac{\sqrt{1-\gamma}}{\gamma}(\gamma - \alpha_1 - \alpha_4)r_2 & x \in I_4 \\ \frac{\sqrt{\gamma}}{1-\gamma}(1 - \gamma - \alpha_2 - \alpha_3)r_1 & x \in I_2 \\ r\sqrt{\gamma} & x \in I_{c23} \\ \frac{\sqrt{\gamma}}{1-\gamma}(1 - \gamma - \alpha_2 - \alpha_3)r_2 & x \in I_3. \end{cases}$$

Recall that

$$1 - \varepsilon^2 - \varepsilon^3 = \frac{1}{\gamma}(\gamma - \alpha_1 - \alpha_4) = \frac{1}{1-\gamma}(1 - \gamma - \alpha_2 - \alpha_3).$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

We consider six cases, and assume ε is sufficiently small in each. For $x \in I_1$,

$$\begin{aligned}
T_{g^\varepsilon} u(x) &= \int_0^1 g^\varepsilon(x,y)u(y)dy \\
&= \alpha_2 r \cdot \frac{\sqrt{\gamma}}{1-\gamma}(1-\gamma-\alpha_2-\alpha_3)r_1 + (1-\gamma-\alpha_2-\alpha_3)r_1 \cdot \sqrt{\gamma} r \\
&\quad + \alpha_3 r \cdot \frac{\sqrt{\gamma}}{1-\gamma}(1-\gamma-\alpha_2-\alpha_3)r_2 \\
&> (1-\gamma-\alpha_2-\alpha_3)r_1\sqrt{\gamma}r = \frac{1-\gamma}{\gamma}(\gamma-\alpha_1-\alpha_4)\sqrt{\gamma}r_1 r \\
&= r\sqrt{\gamma(1-\gamma)}\frac{\sqrt{1-\gamma}}{\gamma}(\gamma-\alpha_1-\alpha_4)r_1 = r\sqrt{\gamma(1-\gamma)}u(x).
\end{aligned}$$

For $x \in I_4$,

$$\begin{aligned}
T_{g^\varepsilon} u(x) &> (1-\gamma-\alpha_2-\alpha_3)r_2\sqrt{\gamma}r = \frac{1-\gamma}{\gamma}(\gamma-\alpha_1-\alpha_4)r_2\sqrt{\gamma}r \\
&= r\sqrt{\gamma(1-\gamma)}\frac{\sqrt{1-\gamma}}{\gamma}(\gamma-\alpha_1-\alpha_4)r_2 = r\sqrt{\gamma(1-\gamma)}u(x).
\end{aligned}$$

For $x \in I_2$,

$$\begin{aligned}
T_{g^\varepsilon} u(x) &> (\gamma-\alpha_1-\alpha_4)r_1\sqrt{1-\gamma}r = \frac{\gamma}{1-\gamma}(1-\gamma-\alpha_2-\alpha_3)r_1\sqrt{1-\gamma}r \\
&= r\sqrt{\gamma(1-\gamma)}\frac{\sqrt{\gamma}}{1-\gamma}(1-\gamma-\alpha_2-\alpha_3)r_1 = r\sqrt{\gamma(1-\gamma)}u(x).
\end{aligned}$$

For $x \in I_3$,

$$\begin{aligned}
T_{g^\varepsilon} u(x) &> (\gamma-\alpha_1-\alpha_4)r_2\sqrt{1-\gamma}r = \frac{\gamma}{1-\gamma}(1-\gamma-\alpha_2-\alpha_3)r_2\sqrt{1-\gamma}r \\
&= r\sqrt{\gamma(1-\gamma)}\frac{\sqrt{\gamma}}{1-\gamma}(1-\gamma-\alpha_2-\alpha_3)r_2 = r\sqrt{\gamma(1-\gamma)}u(x).
\end{aligned}$$

Next consider $x \in I_{c14}$. Using the fact that $r^2 = sr_1^2 + (1-s)r_2^2$, along with $\alpha_2 = (1-\gamma)s\varepsilon^2$ and $\alpha_3 = (1-\gamma)\left((1-s)\varepsilon^2 + \varepsilon^3\right)$, we obtain

$$\begin{aligned}
T_{g^\varepsilon} u(x) &= (1-\gamma-\alpha_2-\alpha_3)\left(\alpha_2\frac{\sqrt{\gamma}}{1-\gamma}r_1^2 + \sqrt{\gamma}r^2 + \alpha_3\frac{\sqrt{\gamma}}{1-\gamma}r_2^2\right) \\
&= (1-\gamma)(1-\varepsilon^2-\varepsilon^3)\left(\frac{\sqrt{\gamma}}{1-\gamma}\left(\alpha_2 r_1^2 + \alpha_3 r_2^2\right) + \sqrt{\gamma}r^2\right) \\
&= (1-\gamma)(1-\varepsilon^2-\varepsilon^3)\left(\frac{\sqrt{\gamma}}{1-\gamma}\left((1-\gamma)s\varepsilon^2 r_1^2 + (1-\gamma)\left((1-s)\varepsilon^2+\varepsilon^3\right)r_2^2\right) + \sqrt{\gamma}r^2\right) \\
&= \sqrt{\gamma}(1-\gamma)(1-\varepsilon^2-\varepsilon^3)\left(s\varepsilon^2 r_1^2 + \left((1-s)\varepsilon^2+\varepsilon^3\right)r_2^2 + r^2\right) \\
&= \sqrt{\gamma}(1-\gamma)(1-\varepsilon^2-\varepsilon^3)\left(\varepsilon^2 r^2 + \varepsilon^3 r_2^2 + r^2\right) = \sqrt{\gamma}(1-\gamma)\left(r^2 + \left(r_2^2 - r^2\right)\varepsilon^3 + O(\varepsilon^4)\right) \\
&> \sqrt{\gamma}(1-\gamma)r^2 = r\sqrt{\gamma(1-\gamma)}\sqrt{1-\gamma}r = r\sqrt{\gamma(1-\gamma)}u(x),
\end{aligned}$$

where the inequality holds for sufficiently small $\varepsilon > 0$ since $r_2 > r$.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Similarly, for $x \in I_{c23}$,

$$\begin{aligned}
T_{g^\varepsilon}u(x) &= (\gamma - \alpha_1 - \alpha_4)\left(\alpha_1\frac{\sqrt{1-\gamma}}{\gamma}r_1^2 + \sqrt{1-\gamma}r^2 + \alpha_4\frac{\sqrt{1-\gamma}}{\gamma}r_2^2\right)\\
&= \gamma(1-\varepsilon^2-\varepsilon^3)\left(\frac{\sqrt{1-\gamma}}{\gamma}\left(\alpha_1 r_1^2 + \alpha_4 r_2^2\right) + \sqrt{1-\gamma}r^2\right)\\
&= \gamma(1-\varepsilon^2-\varepsilon^3)\left(\frac{\sqrt{1-\gamma}}{\gamma}\left(\gamma s\varepsilon^2 r_1^2 + \gamma\left((1-s)\varepsilon^2+\varepsilon^3\right)r_2^2\right) + \sqrt{1-\gamma}r^2\right)\\
&= \gamma\sqrt{1-\gamma}(1-\varepsilon^2-\varepsilon^3)\left(s\varepsilon^2 r_1^2 + \left((1-s)\varepsilon^2+\varepsilon^3\right)r_2^2 + r^2\right)\\
&= \gamma\sqrt{1-\gamma}(1-\varepsilon^2-\varepsilon^3)\left(\varepsilon^2 r^2 + \varepsilon^3 r_2^2 + r^2\right) = \gamma\sqrt{1-\gamma}\left(r^2 + (r_2^2 - r^2)\varepsilon^3 + O(\varepsilon^4)\right)\\
&> \gamma\sqrt{1-\gamma}r^2 = r\sqrt{\gamma(1-\gamma)}\sqrt{\gamma}r = r\sqrt{\gamma(1-\gamma)}u(x),
\end{aligned}$$

where again the inequality holds for sufficiently small $\varepsilon > 0$.

We have shown that $T_{g^\varepsilon}u(x) > r\sqrt{\gamma(1-\gamma)}u(x)$ for all $x \in [0,1]$, for $\varepsilon > 0$ sufficiently small. Therefore, there exists $\varepsilon > 0$ such that $\|g^\varepsilon\|_{\mathrm{op}} > r\sqrt{\gamma(1-\gamma)}$. □

7. Open questions

We collect here some questions arising naturally from our investigations.

1. Our results establish a 'reentrant phase transition' in upper tail large deviations for homomorphism densities in specific block model random graphs. Note that our results on the symmetric regime are quite general, and applicable for arbitrary block graphons. In contrast, our proof for the existence of a symmetry breaking regime is case-specific, and does not generalise directly. It is natural to believe that this reentrant phase transition phenomenon should hold for a much wider family of block graphons, and it would be interesting to investigate this further.
2. A natural follow up question concerns the precise boundary between the symmetric and non-symmetric regimes. So far, this boundary has been identified for very homogeneous graphs – the Erdős–Rényi random graph in [34] and the Erdős–Rényi bipartite graph in this article. We expect the general case to be significantly more challenging and is beyond the scope of this paper.
3. Theorems 19 and 20 together identify a symmetric regime for the homomorphism density of a regular subgraph. In this regime, is the solution to the variational problem unique?
4. Another natural direction of inquiry concerns the behaviour of the minimiser(s) in the symmetry breaking regime. In fact, we do not even know whether the upper tail variational problem (6) has a unique minimiser in the symmetry breaking phase. Any tangible progress on this uniqueness question would be a promising start in this direction. Moreover, it would be of interest to identify the structure of the minimiser(s) in the non-symmetric regime. These questions remain open even for Erdős–Rényi graphs, and were already raised in [14] and [34].
5. Finally, we note that our analysis of the upper tail variational problem (6) is restricted to regular subgraphs. Non-trivial extensions to non-regular graphs will likely require new ideas, and will provide new insights on the upper tail problem.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

8. Appendix

8.1 Weak topology LDP upper bound

To prove the upper bound LDP in the weak topology, Theorem 48, we will use a general LDP upper bound given in [11, Section 4.3] which we restate as Lemma 96 below.

We need some notation. Let $\mathscr{H}$ be a real topological vector space whose topology satisfies the Hausdorff property. Let $\mathscr{H}^*$ denote the dual space of continuous linear functionals on $\mathscr{H}$. Let $\mathscr{B}$ denote the Borel sigma-algebra of $\mathscr{H}$ and let $\{\mu_n\}_{n\geq 1}$ be a sequence of probability measures on $(\mathscr{H}, \mathscr{B})$. Define the logarithmic moment generating function $\Lambda_n : \mathscr{H}^* \to (-\infty, \infty]$ of μ_n as

$$\Lambda_n(\lambda) = \log \int_{\mathscr{H}} \exp(\lambda(x)) d\mu_n(x).$$

Given a 'rate' $\{\varepsilon_n\}_{n\geq 1}$, i.e., a sequence of positive real numbers ε_n tending to 0, we define $\bar{\Lambda} : \mathscr{H}^* \to [-\infty, \infty]$ and its Fenchel-Legendre transform $\bar{\Lambda}^* : \mathscr{H} \to [-\infty, \infty]$ as

$$\bar{\Lambda}(\lambda) = \limsup_{n\to\infty} \varepsilon_n \Lambda_n(\lambda/\varepsilon_n)$$

$$\bar{\Lambda}^*(x) = \sup_{\lambda\in\mathscr{H}^*} (\lambda(x) - \bar{\Lambda}(\lambda)).$$

Lemma 96 (Theorem 4.1 of [11]). *For any compact set $\Gamma \subseteq \mathscr{H}$,*

$$\limsup_{n\to\infty} \varepsilon_n \log \mu_n(\Gamma) \leq - \inf_{x\in\Gamma} \bar{\Lambda}^*(x).$$

Proof of Theorem 48. We closely follow the proof of Theorem 5.1 in [11]. Let $\mathscr{H}$ be the vector space $L^2([0,1]^2)$ with the weak topology. For each $a, f \in \mathscr{H}$, define λ_a in the dual space $\mathscr{H}^*$ as

$$\lambda_a(f) = \int_{[0,1]^2} a(x,y) f(x,y) dx dy.$$

Setting $\mu_n = \mathbb{P}_{kn,W_0}$, define $\Lambda_n : \mathscr{H}^* \to \mathbb{R}$ as

$$\Lambda_n(\lambda) = \log \int_{\mathscr{H}} \exp(\lambda(f)) d\mathbb{P}_{kn,W_0}(f) = \log \left(\mathbb{E}_{f\sim\mathbb{P}_{kn,W_0}} \left[\exp(\lambda(f)) \right] \right).$$

Set $\varepsilon_n = \frac{2}{(kn)^2}$, and let

$$\bar{\Lambda}(\lambda) = \limsup_{n\to\infty} \frac{2\Lambda_n((kn)^2\lambda/2)}{(kn)^2}.$$

Let f^G be the empirical graphon on kn vertices drawn from W_0. For $i, j \in [kn]$, let X_{ij} be the indicator for the event that $\{i, j\}$ is an edge in G. Since G is a simple undirected graph $X_{ij} = X_{ji}$ and $X_{ii} = 0$. For ease of notation let $W_0^{ij} = W_0(i/(kn), j/(kn))$. Note $X_{ij} \sim Bern(W_0^{ij})$. Let $I_{kn,1} = [0, 1/kn]$, $I_{kn,i} = ((i-1)/kn, i/kn]$ for $i = 2, \ldots, kn$ and let $B_{i,j,n}$ be the square $I_{kn,i} \times I_{kn,j}$. Let S be the set of symmetric L^2 functions. For $a \in S$, let $\hat{a}_n$ denote the level kn approximant, i.e.

$$\hat{a}_n(x,y) = (kn)^2 \int_B a(w,z) dw\, dz$$

where $B = B_{i,j,n}$ is such that $(x,y) \in B_{i,j,n}$.

Observe that for an empirical graphon f^G,

$$\lambda_a(f^G) = \sum_{1\leq i,j\leq kn, i\neq j} X_{ij} \int_{B(i,j,n)} a(x,y) dx\, dy = \sum_{1\leq i<j\leq kn} X_{ij} \int_{B(i,j,n)\cup B(j,i,n)} a(x,y) dx\, dy.$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Recall $X_{ij} \sim Bern(W_0^{ij})$, and so for any θ

$$\mathbb{E}\left[\exp\left(\theta X_{ij}\right)\right] = W_0^{ij}\exp(\theta) + 1 - W_0^{ij}.$$

Since the events $\{X_{ij}\}_{i<j}$ are independent, it follows that for any $a \in S$

$$\begin{aligned}
\Lambda_n((kn)^2\lambda_a/2) &= \log\left(\mathbb{E}_{f\sim\mathbb{P}_{kn,W_0}}\left[\exp\left((kn)^2\lambda_a(f)/2\right)\right]\right)\\
&= \log\left(\mathbb{E}_{f\sim\mathbb{P}_{kn,W_0}}\left[\exp\left(\frac{(kn)^2}{2}\sum_{1\le i<j\le kn} X_{ij}\int_{B(i,j,n)\cup B(j,i,n)} a(x,y)dxdy\right)\right]\right)\\
&= \log\prod_{1\le i<j\le kn}\left(W_0^{ij}\exp\left(\frac{(kn)^2}{2}\int_{B(i,j,n)\cup B(j,i,n)} a(x,y)dxdy\right) + 1 - W_0^{ij}\right)\\
&= \sum_{1\le i<j\le kn}\log\left(W_0^{ij}\exp\left(\frac{(kn)^2}{2}\int_{B(i,j,n)\cup B(j,i,n)} a(x,y)dxdy\right) + 1 - W_0^{ij}\right)\\
&= (kn)^2\sum_{1\le i<j\le kn}\int_{B(i,j,n)}\log\left(W_0^{ij}\exp\left(\hat{a}_n(x,y)\right) + 1 - W_0^{ij}\right)dxdy\\
&= \frac{(kn)^2}{2}\int_{[0,1]^2\setminus B_n}\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right) + 1 - W_0(x,y)\right)dxdy \qquad (47)
\end{aligned}$$

where $B_n \triangleq \bigcup_{i=1}^n B(i,i,n)$.

Next we consider $u_p(x) \triangleq \log\left(pe^x + 1 - p\right)$, in order to reason about the limit of the above integral as $n \to \infty$. For $p \in [0,1]$, $u_p'(x) = pe^x/(pe^x + 1 - p)$, and so $|u_p'(x)| \le 1$ everywhere. Thus $|u_p(x) - u_p(y)| \le |x - y|$ for all $x, y \in \mathbb{R}$. It follows that

$$\begin{aligned}
&|\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right) + 1 - W_0(x,y)\right) - \log\left(W_0(x,y)\exp\left(a(x,y)\right) + 1 - W_0(x,y)\right)|\\
&\quad\le |\hat{a}_n(x,y) - a(x,y)|.
\end{aligned}$$

By Proposition 2.6 of [11], $\hat{a}_n \to a$ in L^2, and therefore using the above inequality and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
&\int_{[0,1]^2}\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right) + 1 - W_0(x,y)\right)\\
&-\left(\log\left(W_0(x,y)\exp\left(a(x,y)\right) + 1 - W_0(x,y)\right)\right)dxdy\\
&\le \int_{[0,1]^2}|\hat{a}_n(x,y) - a(x,y)|dxdy\\
&\le \left(\int_{[0,1]^2}\left(\hat{a}_n(x,y) - a(x,y)\right)^2 dxdy\right)^{\frac{1}{2}} \to 0.
\end{aligned}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Therefore,

$$\lim_{n\to\infty}\int_{[0,1]^2}\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right)+1-W_0(x,y)\right)dx\,dy$$

$$=\int_{[0,1]^2}\log\left(W_0(x,y)\exp\left(a(x,y)\right)+1-W_0(x,y)\right)dx\,dy. \tag{48}$$

Next we consider the limit of the above integral over the set B_n. Since $|u_p(x)-u_p(y)|\le|x-y|$, taking $y=0$, we obtain $|u_p(x)|\le|x|$. The Cauchy–Schwarz inequality then implies that

$$\left|\int_{B_n}\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right)+1-W_0(x,y)\right)dxdy\right|\le\left|\int_{B_n}|\hat{a}_n(x,y)|dxdy\right|$$

$$\le\left(|B_n|\int\hat{a}_n(x,y)^2dxdy\right)^{1/2}\le\frac{\|\hat{a}_n\|_2}{\sqrt{kn}}.$$

Since $\hat{a}_n\to a$ in L^2,

$$\lim_{n\to\infty}\int_{B_n}\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right)+1-W_0(x,y)\right)dx\,dy=0. \tag{49}$$

Finally we use (47), (49), and (48) to compute for $a\in S$

$$\overline{\Lambda}(\lambda_a)=\limsup_{n\to\infty}\frac{2\Lambda_n\left((kn)^2\lambda_a/2\right)}{(kn)^2}$$

$$=\limsup_{n\to\infty}\int_{[0,1]^2\setminus B_n}\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right)+1-W_0(x,y)\right)dx\,dy$$

$$=\limsup_{n\to\infty}\int_{[0,1]^2}\log\left(W_0(x,y)\exp\left(\hat{a}_n(x,y)\right)+1-W_0(x,y)\right)dx\,dy$$

$$=\int_{[0,1]^2}\log\left(W_0(x,y)\exp\left(a(x,y)\right)+1-W_0(x,y)\right)dx\,dy.$$

For $f\in\mathscr{H}$, let

$$\bar{\Lambda}^*(x)\triangleq\sup_{\lambda\in\mathscr{H}^*}\left(\lambda(f)-\bar{\Lambda}(\lambda)\right).$$

By Proposition 36,

$$\bar{\Lambda}^*(f)\ge\sup_{a\in S}\left(\lambda_a(f)-\bar{\Lambda}(\lambda_a)\right)=2I_{W_0}(f).$$

Combined with the compactness of the weak topology [11, Proposition 2.8], [11, Theorem 4.1], stated here as Lemma 96, implies that

$$\limsup_{n\to\infty}\frac{2}{(kn)^2}\log\mathbb{P}_{kn,W_0}(F)\le-\inf_{f\in F}2I_{W_0}(f).$$

□

8.2 Other useful results

The following theorem appears in [23]. We include the notation given by [34].

Theorem 97. *Let $\mu_1,\dots,\mu_n$ be probability measures on $\Omega_1,\dots,\Omega_n$, respectively, and let $\mu=\prod_{i=1}^n\mu_i$ be the product measure on $\Omega=\prod_{i=1}^n\Omega_i$. Let $A_1,\dots,A_m$ be nonempty subsets of $[n]=$*

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$\{1, \dots, n\}$ and write $\Omega_A = \prod_{l \in A} \Omega_l$ and $\mu_A = \prod_{l \in A} \mu_l$. Let $f_i \in L^{p_i}(\Omega_{A_i}, \mu_{A_i})$ with $p_i \geq 1$ for each $i \in [m]$ and suppose in addition that $\sum_{i: l \in A_i} \frac{1}{p_i} \leq 1$ for each $l \in [n]$. Then

$$\int \prod_{i=1}^{m} |f_i| d\mu \leq \prod_{i=1}^{m} \left(\int |f_i|^{p_i} d\mu_{A_i} \right)^{\frac{1}{p_i}}.$$

Assume without loss of generality that the sets $\{k : i \in A_k\}$ are distinct for each $i \in [n]$, and $\int |f_j|^{p_j} d\mu_{A_j} > 0$ for all $j \in [m]$. Then equality holds if and only if there exist functions f_{ji} on Ω_i and constants $\alpha_{ji} > 0$ such that

1. *for all $j \in [m]$, $|f_j| = \prod_{i \in A_j} |f_{ji}|$ almost everywhere with respect to μ_{A_j}*
2. *for all $i \in [n]$, $r, s \in \{k : i \in A_k\}$, and $r \neq s$, $\alpha_{ri} |f_{ri}|^{p_r} = \alpha_{si} |f_{si}|^{p_s}$ almost everywhere with respect to μ_i.*

If $p_i = d$ for every $i \in [m]$, the inequality reduces to

$$\int \prod_{i=1}^{m} |f_i| d\mu \leq \prod_{i=1}^{m} \left(\int |f_i|^{d} d\mu_{A_i} \right)^{\frac{1}{d}}.$$

Proof of Proposition 45. The proof is very similar to that of Theorem 2.3 [14], and we only sketch it here. Since both f_n and g_n are block graphons, the distance $d_\square(f_n, g_n)$ can be written as a maximum over 4^{kn} pairs of sets $S, T \subset [0, 1]$, that are unions of a subset of the intervals used in the definition of g_n. But given S and T, the expectation of $\int_{S \times T} f_n$ is $\int_{S \times T} g_n$. Azuma's inequality then shows that the probability that the difference is larger than ε is bounded by $e^{-c\varepsilon^2 (kn)^2}$ for some universal constant $c > 0$. The union bound now implies the proposition. □

Proposition 98 (Lemma A.1 of [34]). *Let $d \geq 1$ and $p \in (0, 1)$. Consider $\psi_p(x) = h_p\left(x^{1/d}\right)$ with domain $[0, 1]$.*

1. *The function $\psi_p(x)$ is decreasing on $[0, p^d]$ and increasing on $[p^d, 1]$.*
2. *Let*

$$p_0 = \frac{d - 1}{d - 1 + e^{\frac{d}{d-1}}}.$$

   (a) *If $p > p_0$, then $\psi_p(x)$ is convex and $\psi_p'' > 0$ on $[0, 1]$.*
   (b) *If $p = p_0$, then $\psi_p(x)$ is convex and $\psi_p'' = 0$ at exactly one point in $(p^d, 1)$.*
   (c) *For $p < p_0$, the function $\psi_p(x)$ has exactly two inflection points r_1^d and r_2^d with $p < r_1 < r_2 \leq 1$. The function ψ_p is convex on $[0, r_1^d] \cup [r_2^d, 1]$ and concave on $[r_1^d, r_2^d]$. Moreover ψ_p'' is strictly positive on $[0, r_1^d) \cup (r_2^d, 1]$.*

8.3 Behaviour at $t = t_{\max}$

8.3.1. Proof of Theorem 21

In this subsection we prove Theorem 21.

Proof of Theorem 21. First, observe $t(H, \tilde{f}_{\max}) = t_{\max}$. Consider the optimisation problem

$$\min\{J_{W_0}(\tilde{f}) : t(H, \tilde{f}) \geq t_{\max}\}. \tag{50}$$

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

Let $\tilde{f}_*$ be a minimiser of (50). Lemma 82 states that there exists a sequence $f_n \in \mathcal{W}_\Omega$ such that $t(H, f_n) \to t_{\max}$, $f_n \geq W_0$ pointwise, $f_n = W_0$ on irrelevant blocks, $I_{W_0}(f_n) \to J_{W_0}(\tilde{f}_*)$, and $\delta_\square(f_n, \tilde{f}_*) \to 0$. For $n \geq n_0$, the f_n satisfy the ε'-neighbourhood minorant condition for some $\varepsilon' > 0$ (this follows from Lemma 84). It suffices to prove the claim for ε satisfying $0 < \varepsilon \leq \varepsilon'$; we therefore assume this inequality. Lemma 72 implies that $\tilde{f}_* \in \tilde{\mathcal{B}}^\gamma$, and $J_{W_0}(\tilde{f}_*) = I_{W_0}(g) < \infty$ for some $g \in \mathcal{B}^\gamma$. Therefore

$$I_{W_0}(g) = \min\{J_{W_0}(\tilde{f}) : t(H, \tilde{f}) \geq t_{\max}\} \leq I_{W_0}(f_{\max}) < \infty,$$

and thus $g \in \mathcal{B}^\gamma \cap \mathcal{W}_\Omega$. Further, $\delta_\square(f_*, g) = 0$ implies $t(H, g) = t_{\max}$. This implies $g = 1$ on the relevant blocks. Moreover, $I_{W_0}(g) \leq I_{W_0}(f_{\max})$ which is possible iff $g = W_0$ on the irrelevant blocks. We therefore conclude that $g = f_{\max}$. We then have $\tilde{f}_* = \tilde{f}_{\max}$ and $J_{W_0}(\tilde{f}_*) = J_{W_0}(\tilde{f}_{\max}) = I_{W_0}(f_{\max})$. We have thus established that $\tilde{f}_{\max}$ is the unique minimiser to the upper tail variational problem for $t = t_{\max}$.

Next, we have,

$$\mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}}, \tilde{f}_{\max}) \geq \varepsilon \,\middle|\, t(H, G_{kn}) \geq t_{\max}\right) = \frac{\mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}}, \tilde{f}_{\max}) \geq \varepsilon, t(H, G_{kn}) \geq t_{\max}\right)}{\mathbb{P}_{kn,W_0}\left(t(H, G_{kn}) \geq t_{\max}\right)}. \tag{51}$$

Note that the set $\{\tilde{f} : \delta_\square(\tilde{f}, \tilde{f}_{\max}) \geq \varepsilon, t(H, \tilde{f}) \geq t_{\max}\}$ is closed, and thus Theorem 11 implies

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}}, \tilde{f}_{\max}) \geq \varepsilon, t(H, G_{kn}) \geq t_{\max}\right) \\ \leq -\inf\{J_{W_0}(\tilde{f}) : \delta_\square(\tilde{f}, \tilde{f}_{\max}) \geq \varepsilon, t(H, \tilde{f}) \geq t_{\max}\} =: -C'.\end{aligned}$$

We conclude that for every $\eta > 0$, there exists $N(\eta)$ such that if $n \geq N(\eta)$, then

$$\frac{1}{(kn)^2} \log \mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}}, \tilde{f}_{\max}) \geq \varepsilon, t(H, G_{kn}) \geq t_{\max}\right) \leq -C' + \eta,$$

or equivalently

$$\mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}}, \tilde{f}_{\max}) \geq \varepsilon, t(H, G_{kn}) \geq t_{\max}\right) \leq \exp\left(-(kn)^2(C' - \eta)\right). \tag{52}$$

Next, we turn to the denominator. Let $(I_i \times I_j)_{i,j\in[k]}$ be the blocks of W_0, and let $R \subset [0,1]^2$ denote the union of the relevant blocks. Recall the definition of $\vartheta(\cdot)$ from (2), let $A = (A_{ij})_{i,j\in[kn]}$ denote the adjacency matrix of G_{kn} and let S be the set of relevant edges:

$$S = \left\{(i,j) \in [kn]^2 : I_{k\left(\frac{i}{kn}\right)} \times I_{k\left(\frac{j}{kn}\right)} \subset R\right\}.$$

Observe that

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

$$\begin{aligned}
\mathbb{P}_{kn,W_0}(t(H,G_{kn}) = t_{\max}) &= \mathbb{P}\left(\cap_{(i,j)\in S}\{A_{ij}=1\}\right)\\
&= \prod_{a:I_a\times I_a\subset R} p_{aa}^{\binom{n}{2}} \prod_{a<b:I_a\times I_b\subset R} p_{ab}^{n^2}.\\
&= \exp\left(\binom{n}{2}\sum_{a:I_a\times I_a\subset R}\log(p_{aa}) + n^2\sum_{a<b:I_a\times I_b\subset R}\log(p_{ab})\right)\\
&= \exp\left(\frac{1}{2}n^2\sum_{a,b:I_a\times I_b\subset R}\log(p_{ab}) - \frac{n}{2}\sum_{a:I_a\times I_a\subset R}\log(p_{aa})\right)\\
&= \exp\left(-\frac{1}{2}(kn)^2\sum_{a,b:I_a\times I_b\subset R}\frac{1}{k^2}\log\left(\frac{1}{p_{ab}}\right) - \frac{n}{2}\sum_{a:I_a\times I_a\subset R}\log(p_{aa})\right)\\
&= \exp\left(-(kn)^2 I_{W_0}(f_{\max})\,(1+o(1))\right).
\end{aligned}$$

Recall that $I_{W_0}(f_{\max}) = J_{W_0}(\tilde{f}_{\max})$, so that

$$\mathbb{P}_{kn,W_0}(t(H,G_{kn}) \geq t_{\max}) = \exp\left(-(kn)^2 J_{W_0}(\tilde{f}_{\max})(1+o(1))\right). \tag{53}$$

Applying (52) and (53) to (51), we obtain for $n \geq N(\eta)$

$$\begin{aligned}
\mathbb{P}_{kn,W_0}\left(\delta_\square(f^{G_{kn}},\tilde{f}_{\max}) \geq \varepsilon \,\middle|\, t(H,G_{kn}) \geq t_{\max}\right) &\leq \exp\left(-(kn)^2(C'-\eta) + (kn)^2 J_{W_0}(\tilde{f}_{\max})(1+o(1))\right)\\
&= \exp\left(-(kn)^2\left(C' - \eta - J_{W_0}(\tilde{f}_{\max})(1+o(1))\right)\right).
\end{aligned}$$

Recalling that $\tilde{f}_{\max}$ is the unique minimiser of (50), the proof is complete by observing that

$$\begin{aligned}
C' &= \inf\{J_{W_0}(\tilde{f}) : \delta_\square(\tilde{f},\tilde{f}_{\max}) \geq \delta, t(H,\tilde{f}) \geq t_{\max}\}\\
&> \inf\{J_{W_0}(\tilde{f}) : t(H,\tilde{f}) \geq t_{\max}\} = J_{W_0}(\tilde{f}_{\max}).
\end{aligned}$$

□

8.3.2. Elaboration on Remark 22

Let W_0 be a uniform k-block graphon, and $\tau = t(H,\cdot)$, where H is a finite d-regular graph. By Lemma 73, $t(H,f) \leq t(H,f^*) \leq t_{\max}$, with equality if and only if $f^* = \mathbf{1}_{W_0>0}$, as all non-trivial blocks of W_0 are relevant. Thus $f^* = \mathbf{1}_{W_0>0}$ is the unique solution to $t(H,f) = t_{\max}$ in this setting.

Note that if $W_0 = f_p^\gamma$, and $\tau(\tilde{f}) = \|f\|_{\mathrm{op}}$, $t_{\max} = \sqrt{\gamma(1-\gamma)}$. Further, using Lemma 94, we conclude that $\|\tilde{f}^G\|_{\mathrm{op}} \geq t_{\max}$ implies that $\|f^G\|_2^2 \geq 2\gamma(1-\gamma)$. This is possible if and only if $\tilde{f}^G = \tilde{f}_1^\gamma$. This establishes the desired claim.

Acknowledgments

This work was initiated when CB, JC, JG, and SP were affiliated with Microsoft Research New England (MSR NE). The authors all thank MSR NE for the vibrant research environment. SP was supported by the NSF Graduate Research Fellowship DGE-1650044. JG was supported by a Microsoft Research PhD fellowship. The work was continued while SP was at the School of Mathematics at the Georgia Institute of Technology and then at the NSF-Simons Center for

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

the Mathematical and Statistical Analysis of Biology at Harvard, and JG was at the Operations Research Center and then the Mathematics Department, Massachusetts Institute of Technology. The authors thank Yufei Zhao and Souvik Dhara for helpful discussions during the early part of the project. They also thank Oleg Pikhurko and Jan Grebik for generously sharing their results on LDPs for W-random graphs. Finally, the authors thank the anonymous referee for pointing out an error regarding the uniqueness of the optimiser in the symmetric phase in an earlier version of this manuscript.

References

[1] Augeri, F. (2020) Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős–Rényi graphs. *Ann. Probab.* **48**(5) 2404–2448.
[2] Austin, T. (2019) The structure of low-complexity Gibbs measures on product spaces. *Ann. Probab.* **47**(6) 4002–4023.
[3] Basak, A. and Basu, R. (2023) Upper tail large deviations of the cycle counts in Erdős–Rényi graphs in the full localized regime. *Commun. Pure Appl. Math.* **76**(1) 3–72.
[4] Bhattacharya, B. B. and Ganguly , S. (2020) Upper tails for edge eigenvalues of random graphs. *SIAM J. Discrete Math.* **34**(2) 1069–1083.
[5] Bhattacharya, B. B., Ganguly, S., Lubetzky, E. and Zhao, Y. (2017) Upper tails and independence polynomials in random graphs. *Adv. Math.* **319** 313–347.
[6] Bhattacharya, B. B., Ganguly, S., Shao, X. and Zhao, Y. (2016) Upper tails for arithmetic progressions in a random set. *Int. Math. Res. Not.* **2020**(1) 167–213.
[7] Bhattacharya, S. and Dembo, A. (2021) Upper tail for homomorphism counts in constrained sparse random graphs. *Random Struct Algor.* **59**(3) 315–338.
[8] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T. and Vesztergombi, K. (2008) Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.* **219**(6) 1801–1851.
[9] Borgs, C., Chayes, J., Lovász, L., Sós, V. and Vesztergombi, K. (2012) Convergent sequences of dense graphs II: Multiway cuts and statistical physics. *Ann. Math.* **179**(1) 151–219.
[10] Chatterjee, S. (2012) The missing log in large deviations for triangle counts. *Random Struct Algor.* **40**(4) 437–451.
[11] Chatterjee, S. (2017) Large deviations for random graphs. *Lecture Notes in Math.* Vol. **2197**, Saint-Flour Probability Summer School. Cham: Springer, xi+167 pp.
[12] Chatterjee, S. and Dembo, A. (2016) Nonlinear large deviations. *Adv. Math.* **299** 396–450.
[13] Chatterjee, S. and Dey, P. S. (2010) Applications of Stein's method for concentration inequalities. *Ann. Probab.* **38**(6) 2443–2485.
[14] Chatterjee, S. and Varadhan, S. R. S. (2011) The large deviation principle for the Erdős–Rényi random graph. *Eur. J. Comb.* **32**(7) 1000–1017.
[15] Cook, N. and Dembo, A. (2020) Large deviations of subgraph counts for sparse Erdős–Rényi graphs. *Adv. Math.* **373** 107289.
[16] DeMarco, B. and Kahn, J. (2012) Upper tails for triangles. *Random Struct Algor.* **40**(4) 452–459.
[17] DeMarco, R. and Kahn, J. (2012) Tight upper tail bounds for cliques. *Random Struct Algor.* **41**(4) 469–487.
[18] Dembo, A. and Lubetzky, E. (2018) A large deviation principle for the Erdős–Rényi uniform random graph. *Electron. Commun. Probab.* **23**(79) 13.
[19] Dembo, A. and Zeitouni, O. (2010) *Large Deviations Techniques and Applications*, vol. 38 Stochastic Modelling and Applied Probability. Springer-Verlag. Berlin Heidelberg, second edition.
[20] Dhara, S. and Sen, S. (2022) Large deviation for uniform graphs with given degrees. *Ann. Appl. Probab.* **32**(3) 2327–2353.
[21] Durrett, R. (2019) *Probability: Theory and Examples*. Vol. **49**. Cambridge University Press.
[22] Eldan, R. (2018) Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* **28**(6) 1548–1596.
[23] Finner, H. (1992) A generalization of Hölder's inequality and some probability inequalities, *Ann. Probab.* **20**(4) 1893–1901.
[24] Grebík, J. and Pikhurko, O. (2021) Large deviation principles for block and step graphon random graph models. arXiv preprint arXiv: 2101.07025.
[25] Harel, M., Mousset, F. and Samotij, W. (2022) Upper tails via high moments and entropic stability. *Duke Math. J.* **1**(1) 1–104.
[26] Janson, S. (2013) Graphons, cut norm and distance, couplings and rearrangements. In *New York Journal of Mathematics Monographs*, Vol. 4, University at Albany. Albany, NY.
[27] Janson, S., Oleszkiewicz, K. and Ruciński, A. (2004) Upper tails for subgraph counts in random graphs. *Isr. J. Math.* **142**(1) 61–92.
[28] Janson, S. and Ruciński, A. (2002) The infamous upper tail. *Random Struct Algor.* **20**(3) 317–342.
[29] Janson, S. and Ruciński, A. (2004) The deletion method for upper tail estimates. *Combinatorica* **24**(4) 615–640.

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press

[30] Kim, J. H. and Van, H. V. (2004) Divide and conquer martingales and the number of triangles in a random graph. *Random Struct Algor*. **24**(2) 166–174.
[31] Liu, Y. P. and Zhao, Y. (2021) On the upper tail problem for random hypergraphs. *Random Struct Algor*. **58**(2) 179–220.
[32] Lovász, L. (2012) *Large Networks and Graph Limits*. vol. 60, American Mathematical Society.
[33] Lovász, L. and Szegedy, B. (2007) Szemerédi's lemma for the analyst. *Geom. Funct. Anal.* **17**(1) 252–270.
[34] Lubetzky, E. and Zhao, Y. (2015) On replica symmetry of large deviations in random graphs. *Random Struct. Algor*. **47**(1) 109–146.
[35] Markering, M. (2023) The large deviation principle for inhomogeneous Erdős–Rényi random graphs. *J. Theor. Probab.* **36**(2) 711–727.
[36] Šileikis, M. and Warnke, L. (2019) A counterexample to the DeMarco-Kahn upper tail conjecture. *Random Struct. Algor.* **55**(4) 775–794.
[37] Van, H. V. (2001) A large deviation result on the number of small subgraphs of a random graph. *Comb. Prob. Comput.* **10**(1) 79–94.

Cite this article: Borgs C, Chayes J, Gaudio J, Petti S, and Sen S (2025). A large deviation principle for block models. *Combinatorics, Probability and Computing* **34**, 680–753. https://doi.org/10.1017/S0963548325000008

https://doi.org/10.1017/S0963548325000008 Published online by Cambridge University Press