EURANDOM PREPRINT SERIES

2020-010

September 4, 2020

**Large deviation principle for the maximal eigenvalue of inhomogenous Erdös-Rényi random graphs**

A. Chakrabarty, R. Hazra, F. den Hollander, M. Sfragara

# LARGE DEVIATION PRINCIPLE FOR THE MAXIMAL EIGENVALUE OF INHOMOGENEOUS ERDŐS-RÉNYI RANDOM GRAPHS

ARIJIT CHAKRABARTY, RAJAT SUBHRA HAZRA, FRANK DEN HOLLANDER,
AND MATTEO SFRAGARA

ABSTRACT. We consider an inhomogeneous Erdős-Rényi random graph $G_N$ with vertex set $[N] = \{1, \ldots, N\}$ for which the pair of vertices $i, j \in [N]$, $i \neq j$, is connected by an edge with probability $r(\frac{i}{N}, \frac{j}{N})$, independently of other pairs of vertices. Here, $r \colon [0,1]^2 \to (0,1)$ is a symmetric function that plays the role of a reference graphon. Let $\lambda_N$ be the maximal eigenvalue of the adjacency matrix of $G_N$. It is known that $\lambda_N/N$ satisfies a large deviation principle as $N \to \infty$. The associated rate function $\psi_r$ is given by a variational formula that involves the rate function $I_r$ of a large deviation principle on graphon space. We analyse this variational formula in order to identify the properties of $\psi_r$, specially when the reference graphon is of rank 1.

## 1. INTRODUCTION AND MAIN RESULTS

1.1. **Motivation.** In the past 20 years, many properties have been derived about spectra of random matrices associated with *random graphs*, like the adjacency matrix and the Laplacian matrix [3, 7, 9, 15, 18, 20–25, 30, 31]. The focus of the present paper is on *inhomogeneous* Erdős-Rényi random graphs, which are rooted in the theory of complex networks. We consider the *dense regime*, where the average degree of the vertices are proportional to the size of the graph, and analyse the rate function of the *large deviation principle for the maximal eigenvalue of the adjacency matrix* derived in [17]. In [10] the *non-dense non-sparse regime* was considered, where the degrees diverge but sublinearly in the size of the graph, and identified the scaling limit of the empirical spectral distribution of both the adjacency matrix and the Laplacian matrix. Recent results on the maximal eigenvalue in the *sparse regime*, where the degrees are stochastically bounded, can be found in [6].

Large deviations of Erdős-Rényi random graphs were studied in [12, 13, 27] with the help of the theory of graphons, in particular, subgraph densities and maximal eigenvalues. We refer to [12] for a comprehensive review of the literature. Large deviation theory for random matrices started in [4], with the study of large deviations of the empirical spectral distribution of $\beta$-ensembles with a quadratic potential. The rate was shown to be the square of the number of vertices, and the rate function was shown to be given by a non-commutative notion of entropy. The maximal eigenvalue for such ensembles was studied in [5]. Large deviations of the empirical spectral distribution of random matrices with non-Gaussian tails were derived in [8]. More recently, the maximal eigenvalue in that setting was studied in [1, 2]. The adjacency matrix of an inhomogeneous Erdős-Rényi random graph does not fall in this regime, and hence different techniques are needed in the present paper.

1.2. **LDP for inhomogeneous Erdős-Rényi random graphs.** Let

$$\mathcal{W} = \big\{ h \colon [0,1]^2 \to [0,1] \colon h(x,y) = h(y,x) \ \forall \, x, y \in [0,1] \big\} \tag{1.1}$$

1

denote the set of *graphons*. Let $\mathcal{M}$ be the set of Lebesgue measure-preserving bijective maps $\phi\colon [0,1] \mapsto [0,1]$. For two graphons $h_1, h_2 \in \mathcal{W}$, the *cut-distance* is defined by

$$d_\square(h_1, h_2) = \sup_{S,T \subset [0,1]} \left| \int_{S \times T} \mathrm{d}x \, \mathrm{d}y \left[ h_1(x,y) - h_2(x,y) \right] \right|, \tag{1.2}$$

and the *cut-metric* by

$$\delta_\square(h_1, h_2) = \inf_{\phi \in \mathcal{M}} d_\square(h_1, h_2^\phi), \tag{1.3}$$

where $h_2^\phi(x,y) = h_2(\phi(x), \phi(y))$. The cut-metric defines an equivalence relation $\sim$ on $\mathcal{W}$ by declaring $h_1 \sim h_2$ if and only if $\delta_\square(h_1, h_2) = 0$, and leads to the quotient space $\widetilde{\mathcal{W}} = \mathcal{W}/_\sim$. For $h \in \mathcal{W}$, we write $\widetilde{h}$ to denote the equivalence class of $h$ in $\widetilde{\mathcal{W}}$. The equivalence classes correspond to relabelings of the vertices of the graph. The pair $(\widetilde{\mathcal{W}}, \delta_\square)$ is a *compact metric space* [26].

Let $r \in \mathcal{W}$ be a *reference graphon* satisfying

$$\exists \eta > 0\colon \qquad \eta \leq r(x,y) \leq 1 - \eta \quad \forall\, x, y \in [0,1]^2. \tag{1.4}$$

Fix $N \in \mathbb{N}$ and consider the random graph $G_N$ with vertex set $[N] = \{1, \ldots, N\}$ where the pair of vertices $i, j \in [N]$, $i \neq j$, is connected by an edge with probability $r(\frac{i}{N}, \frac{j}{N})$, independently of other pairs of vertices. Write $\mathbb{P}_N$ to denote the law of $G_N$. Use the same symbol for the law on $\mathcal{W}$ induced by the map that associates with the graph $G_N$ its graphon $h^{G_N}$, defined by

$$h^{G_N}(x,y) = \left\{ \begin{array}{ll} 1, & \text{if there is an edge between vertex } \lceil Nx \rceil \text{ and vertex } \lceil Ny \rceil, \\ 0, & \text{otherwise.} \end{array} \right. \tag{1.5}$$

Write $\widetilde{\mathbb{P}}_N$ to denote the law of $\widetilde{h}^{G_N}$.

The following LDP is proved in [17] and is an extension of the celebrated LDP for homogeneous ERRG derived in [13] and further properties of the rate functions were derived in [27].

**Theorem 1.1. [LDP for inhomogeneous ERRG]** *Subject to* (1.4), *the sequence* $(\widetilde{\mathbb{P}}_N)_{N \in \mathbb{N}}$ *satisfies the large deviation principle on* $(\widetilde{\mathcal{W}}, \delta_\square)$ *with rate* $\binom{N}{2}$, *i.e.,*

$$\begin{aligned} \limsup_{N \to \infty} \frac{1}{\binom{N}{2}} \log \widetilde{\mathbb{P}}_N(\mathcal{C}) &\leq -\inf_{\widetilde{h} \in \mathcal{C}} J_r(\widetilde{h}) \quad \forall\, \mathcal{C} \subset \widetilde{\mathcal{W}} \text{ closed,} \\ \liminf_{N \to \infty} \frac{1}{\binom{N}{2}} \log \widetilde{\mathbb{P}}_N(\mathcal{O}) &\geq -\inf_{\widetilde{h} \in \mathcal{O}} J_r(\widetilde{h}) \quad \forall\, \mathcal{O} \subset \widetilde{\mathcal{W}} \text{ open,} \end{aligned} \tag{1.6}$$

*where the rate function* $J_r\colon \widetilde{\mathcal{W}} \to \mathbb{R}$ *is given by*

$$J_r(\widetilde{h}) = \inf_{\phi \in \mathcal{M}} I_r(h^\phi), \tag{1.7}$$

*where* $h$ *is any representative of* $\widetilde{h}$ *and*

$$I_r(h) = \int_{[0,1]^2} \mathrm{d}x \, \mathrm{d}y \, \mathcal{R}\big(h(x,y) \mid r(x,y)\big), \quad h \in \mathcal{W}, \tag{1.8}$$

*with*

$$\mathcal{R}\big(a \mid b\big) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b} \tag{1.9}$$

*the relative entropy of two Bernoulli distributions with success probabilities* $a \in [0,1]$, $b \in (0,1)$ *(with the convention* $0 \log 0 = 0$).

It is clear that $J_r$ is a good rate function, i.e., $J_r \not\equiv \infty$ and $J_r$ has compact level sets. Note that (1.7) differs from the expression in [17], where the rate function is the lower semi-continuous envelope of $I_r(h)$. However, it was shown in [28] that, under the integrability conditions

$\log r, \log(1 - r) \in L^1([0,1]^2)$, the two rate functions are equivalent, since $J_r(\widetilde{h})$ is lower semi-continuous on $\widetilde{\mathcal{W}}$. Clearly, these integrability conditions are implied by (1.4).

1.3. **Graphon operators.** With $h \in \mathcal{W}$ we associate a graphon operator acting on $L^2([0,1])$, defined as the linear integral operator

$$(T_h u)(x) = \int_{[0,1]} \mathrm{d}y\, h(x,y)u(y), \qquad x \in [0,1], \tag{1.10}$$

with $u \in L^2([0,1])$. The operator norm of $T_h$ is defined as

$$\|T_h\| = \sup_{\substack{u \in L^2([0,1]) \\ \|u\|_2 = 1}} \|T_h u\|_2, \tag{1.11}$$

where $\|\cdot\|_2$ denotes the $L^2$-norm. Given a graphon $h \in \mathcal{W}$, we have $\|T_h\| \le \|h\|_2$. Hence, a graphon sequence converging in the $L^2$-norm also converges in the operator norm.

The product of two graphons $h_1, h_2 \in \mathcal{W}$ is defined as

$$(h_1 h_2)(x,y) = \int_{[0,1]} \mathrm{d}z\, h_1(x,z)h_2(z,y), \qquad (x,y) \in [0,1]^2, \tag{1.12}$$

and the $n$-th power of a graphon $h \in \mathcal{W}$ as

$$h^n(x,y) = \int_{[0,1]^{n-1}} \mathrm{d}z_1 \cdots \mathrm{d}z_{n-1}\, h(x,z_1) \times \cdots \times h(z_{n-1},y), \qquad (x,y) \in [0,1]^2,\, n \in \mathbb{N}. \tag{1.13}$$

**Definition 1.2. [Eigenvalues and eigenfunctions]** *$\mu \in \mathbb{R}$ is said to be an eigenvalue of the graphon operator $T_h$ if there exists a non-zero function $u \in L^2([0,1])$ such that*

$$(T_h u)(x) = \mu u(x), \qquad x \in [0,1]. \tag{1.14}$$

*The function $u$ is said to be an eigenfunction associated with $\mu$.*

**Proposition 1.3. [Properties of the graphon operator]** *For any $h \in \mathcal{W}$:*
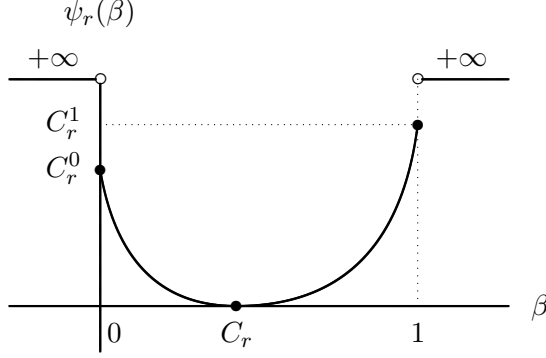*(i) The graphon operator $T_h$ is self-adjoint, bounded and continuous.*
*(ii) The graphon operator $T_h$ is diagonalisable and has countably many eigenvalues, all of which are real and can be ordered as $\mu_1 \ge \mu_2 \ge \cdots \ge 0$. Moreover, there exists a collection of eigenfunctions which form an orthonormal basis of $L^2([0,1])$.*
*(iii) The maximal eigenvalue $\mu_1$ of the graphon operator $h$ is strictly positive and has an associated eigenfunction $u_1$ satisfying $u_1(x) > 0$ for all $x \in [0,1]$. Moreover, $\mu_1 = \|T_h\|$, i.e., the maximal eigenvalue equals the operator norm.*

*Proof.* The claim is a special case of [29, Theorem 7.3] (when the compact Hermitian operators considered there are taken to be the graphon operators). See also [14, Theorem 19.2] and [19, Appendix A]. $\qquad\square$

1.4. **Main theorems.** Let $\lambda_N$ be the *maximal eigenvalue* of the adjacency matrix $A_N$ of $G_N$. Write $\mathbb{P}_N^*$ to denote the law of $\lambda_N/N$.

**Theorem 1.4. [LDP for the maximal eigenvalue]** *Subject to (1.4), the sequence $(\mathbb{P}_N^*)_{N \in \mathbb{N}}$ satisfies the large deviation principle on $\mathbb{R}$ with rate $\binom{N}{2}$ and with rate function*

$$\psi_r(\beta) = \inf_{\substack{\widetilde{h} \in \widetilde{\mathcal{W}} \\ \|T_h\| = \beta}} J_r(\widetilde{h}) = \inf_{\substack{h \in \mathcal{W} \\ \|T_h\| = \beta}} I_r(h), \qquad \beta \in \mathbb{R}. \tag{1.15}$$

FIGURE 1.  Graph of $\beta \mapsto \psi_r(\beta)$.

*Proof.* Note that $\lambda_N/N = \|T_{h^{G_N}}\|$, where $h$ is any representative of $\widetilde{h}$ (because $\|T_{\widetilde{h}}\| = \|T_{h^\phi}\|$ for all $\phi \in \mathcal{M}$). Also note that $\widetilde{h} \mapsto \|T_{\widetilde{h}}\|$ is a bounded and continuous function on $\widetilde{\mathcal{W}}$ [12, Exercises 6.1–6.2, Lemma 6.2]. Hence the claim follows from Theorem 1.1 via the contraction principle [16, Chapter 3]. □

Put

$$C_r = \|T_r\|. \tag{1.16}$$

When $\beta = C_r$, the graphon $h$ that minimizes $I_r(h)$ such that $\|T_h\| = C_r$ is the reference graphon $h = r$ almost everywhere, for which $I_r(r) = 0$ and no large deviation occurs. When $\beta > C_r$, we are looking for graphons $h$ with a larger operator norm. The large deviation cannot go above 1, which is represented by the constant graphon $h \equiv 1$, for which $I_r(1) = C_r^1$. Similarly, when $\beta < C_r$, we are looking for graphons $h$ with a smaller operator norm. The large deviation cannot go below 0, which is represented by the constant graphon $h \equiv 0$, for which $I_r(0) = C_r^0$ (see Fig. 1).

**Theorem 1.5. [Properties of the rate function]** *Subject to* (1.4)*:*
*(i) $\psi_r$ is continuous and unimodal on $[0,1]$, with a unique zero at $C_r$.*
*(ii) $\psi_r$ is strictly decreasing on $[0, C_r]$ and strictly increasing on $[C_r, 1]$.*
*(iii) For every $\beta \in [0,1]$, the set of minimisers of the variational formula for $\psi_r(\beta)$ in* (1.15) *is non-empty and compact in $\widetilde{\mathcal{W}}$.*

If the reference graphon $r$ is of rank 1, i.e.,

$$r(x,y) = \nu(x)\,\nu(y), \quad (x,y) \in [0,1]^2, \tag{1.17}$$

for some $\nu \colon [0,1] \to [0,1]$ that is bounded away from 0 and 1, then we are able to say more. Define

$$m_k = \int_{[0,1]} \nu^k, \qquad k \in \mathbb{N}. \tag{1.18}$$

Note that $C_r = m_2$. Abbreviate $B_r = \int_{[0,1]^2} r^3(1-r)$, and note that $B_r = m_3^2 - m_4^2$. Further abbreviate

$$N_r^1 = \int_{[0,1]^2} \frac{1-r}{r}, \qquad N_r^0 = \int_{[0,1]^2} \frac{r}{1-r}. \tag{1.19}$$

Recall that $\mathcal{M}$ is the set of Lebesgue measure-preserving bijective maps $\phi \colon [0,1] \to [0,1]$.

**Theorem 1.6. [Scaling of the rate function]** *Let $\psi_r$ be the rate function in* (1.15)*.*
*(i) Subject to* (1.4) *and* (1.17)*,*

$$\psi_r(\beta) = [1 + o(1)]\, K_r\, (\beta - C_r)^2, \qquad \beta \to C_r, \tag{1.20}$$

*with*

$$K_r = \frac{C_r^2}{2B_r} = \frac{m_2^2}{2(m_3^2 - m_4^2)}. \tag{1.21}$$

*(ii) Subject to* (1.4),

$$C_r^1 - \psi_r(\beta) = (1 - \beta) \left[ \log \left( \frac{N_r^1}{1 - \beta} \right) + 1 + \mathrm{o}(1) \right], \qquad \beta \uparrow 1. \tag{1.22}$$

*(iii) Subject to* (1.4),

$$C_r^0 - \psi_r(\beta) = \beta \left[ \log \left( \frac{N_r^0}{\beta} \right) + 1 + \mathrm{o}(1) \right], \qquad \beta \downarrow 0. \tag{1.23}$$

**Theorem 1.7. [Scaling of the minimisers]** *Let $h_\beta \in \mathcal{W}$ be any minimiser of the second infimum in* (1.15).
*(i) Subject to* (1.4) *and* (1.17),

$$\lim_{\beta \to C_r} (\beta - C_r)^{-1} \| h_\beta - r - (\beta - C_r)\Delta \|_2 = 0, \tag{1.24}$$

*with*

$$\Delta(x, y) = \frac{C_r}{B_r} r(x, y)^2 [1 - r(x, y)], \qquad (x, y) \in [0, 1]^2. \tag{1.25}$$

*(ii) Subject to* (1.4),

$$\lim_{\beta \uparrow 1} (1 - \beta)^{-1} \| 1 - h_\beta - (1 - \beta)\Delta \|_2 = 0, \tag{1.26}$$

*with*

$$\Delta(x, y) = \frac{1}{N_r^1} \frac{1 - r(x, y)}{r(x, y)}, \qquad (x, y) \in [0, 1]^2. \tag{1.27}$$

*(iii) Subject to* (1.4),

$$\lim_{\beta \downarrow 0} \beta^{-1} \| h_\beta - \beta\Delta \|_2 = 0, \tag{1.28}$$

*with*

$$\Delta(x, y) = \frac{1}{N_r^0} \frac{r(x, y)}{1 - r(x, y)}, \qquad (x, y) \in [0, 1]^2. \tag{1.29}$$

### 1.5. **Discussion and outline.**

**1.** Theorem 1.5 confirms the picture of $\psi_r$ drawn in Fig. 1. It remains open whether or not $\psi_r$ is convex. We do *not* expect $\psi_r$ to be analytic, because bifurcations may occur in the set of minimisers of $\psi_r$ as $\beta$ is varied.

**2.** Theorem 1.6 identifies the scaling of $\psi_r$ near its zero and near its end points, provided $r$ is of rank 1. Theorem 1.7 identifies the corresponding scaling of the minimiser $h_\beta$ of $\psi_r$. Interestingly, the scaling corrections are *not* rank 1. It remains open to determine what happens near $C_r$ when $r$ is not of rank 1 (see the Appendix).

**3.** The inverse curvature $1/K_r$ equals the variance in the central limit theorem derived in [11]. This is in line with the standard folklore of large deviation theory.

**4.** It would be interesting to investigate to what extent the condition on the reference graphon in (1.4) can be weakened to some form of integrability condition. Especially for the upper bound in the LDP this is delicate, because the proof in [17] is based on *block-graphon approximation* (see [28]).

**Outline.** The proof of Theorems 1.5–1.7 is given in Section 3 and relies on the variational formula in (1.15). Since the maximal eigenvalue is invariant under relabeling of the vertices, we can work directly with $I_r$ in (1.8) without worrying about the equivalence classes. In Section 2

we derive an expansion for the operator norm of a graphon around *any* graphon of rank 1. This expansion will be needed in Section 3.

## 2. EXPANSION AROUND GRAPHONS OF RANK 1

In order to prepare for the proof of Theorem 1.6, we show how we can expand the operator norm of a graphon around *any* graphon of rank 1.

**Lemma 2.1. [Rank 1 expansion]** *Consider a graphon $\bar{h} \in \mathcal{W}$ of rank 1 such that $\bar{h}(x, y) = \bar{\nu}(x)\bar{\nu}(y)$, $x, y \in [0, 1]$. For any $h \in \mathcal{W}$ such that $\|T_{h-\bar{h}}\| < \|T_h\|$, the operator norm $\mu = \|T_h\|$ is a solution of the equation*

$$\mu = \sum_{n \in \mathbb{N}_0} \frac{1}{\mu^n} \mathcal{F}_n(h, \bar{h}), \tag{2.1}$$

*where*

$$\mathcal{F}_n(h, \bar{h}) = \int_{[0,1]^2} dx \, dy \, \bar{\nu}(x)(h - \bar{h})^n(x, y)\bar{\nu}(y). \tag{2.2}$$

*Proof.* By Proposition 1.3, we have

$$T_h u = \mu u, \tag{2.3}$$

where $\mu$ equals both the norm and the maximal eigenvalue of $T_h$, and $u$ is the eigenfunction of $h$ corresponding to $u$. Put $g = h - \bar{h}$ and we have $(\mu - T_g)u = T_{\bar{h}}u$. This gives

$$u = (\mu - T_g)^{-1}\bar{\nu}\langle\bar{\nu}, u\rangle. \tag{2.4}$$

where we use that $\mu - T_g$ is invertible because $\|T_g\| = \|T_{h-\bar{h}}\| < \|T_h\|$. Hence, taking the inner product of $u$ with $\bar{\nu}$ and observing that $\langle\bar{\nu}, u\rangle \neq 0$, we get

$$\langle\bar{\nu}, u\rangle = \langle\bar{\nu}, u\rangle\langle\bar{\nu}, (\mu - T_g)^{-1}\bar{\nu}\rangle \tag{2.5}$$

which gives

$$\mu = \langle\bar{\nu}, (1 - T_g/\mu)^{-1}\bar{\nu}\rangle. \tag{2.6}$$

We can expand the above to get

$$\begin{aligned}
\mu &= \left\langle \bar{\nu}, \sum_{n \in \mathbb{N}_0} \left(\frac{T_g}{\mu}\right)^n \bar{\nu} \right\rangle \\
&= \sum_{n \in \mathbb{N}_0} \frac{1}{\mu^n} \int_{[0,1]^{n+1}} dx_0 \, dx_1 \cdots dx_n \, \bar{\nu}(x_0)g(x_0, x_1) \times \cdots \times g(x_{n-1}, x_n)\bar{\nu}(x_n) \qquad (2.7) \\
&= \sum_{n \in \mathbb{N}_0} \frac{1}{\mu^n} \mathcal{F}_n(h, \bar{h}).
\end{aligned}$$

Since $\|T_h\| = \mu$, we get (2.1). $\qquad\square$

Subject to (1.17), it follows from Lemma 2.1 with $h = \bar{h} = r$ that

$$C_r = \|T_r\| = m_2, \tag{2.8}$$

because only the term with $n = 0$ survives in the expansion.

**Remark 2.2. [Higher rank]** The expansion around reference graphons of rank 1 can be extended to finite rank. We provide the details in the Appendix. In this paper we focus on rank 1, for which Lemma 2.1 allows us to analyse the behaviour of $\psi_r(\beta)$ near the values $\beta = C_r$, $\beta = 1$ and $\beta = 0$. This corresponds to an expansion around the graphons $h = \nu \times \nu$, $h \equiv 1$ and $h \equiv 0$, which are all of rank 1. $\qquad\blacksquare$

## 3. PROOFS OF MAIN THEOREMS

Theorem 1.5 is proved in Section 3.1, Theorems 1.6–1.7 are proved in Sections 3.2–3.4.

### 3.1. **Continuity, unimodality and unique minimisers.**

*Proof.* We follow [12, Chapter 6]. Even though this monograph deals with constant reference graphons only, most arguments carry over to $r$ satisfying (1.4).

(i), (iii) Define

$$\psi_r^+(\beta) = \inf_{\substack{h \in \mathcal{W} \\ \|T_h\| \geq \beta}} I_r(h), \quad \psi_r^-(\beta) = \inf_{\substack{h \in \mathcal{W} \\ \|T_h\| \leq \beta}} I_r(h), \qquad \beta \in \mathbb{R}. \tag{3.1}$$

Because $h \mapsto \|T_h\|$ is a nice graph parameter, in the sense of [12, Definition 6.1], it follows that $\beta \mapsto \psi_r^+(\beta)$ is non-decreasing and continuous, while $\beta \mapsto \psi_r^-(\beta)$ is non-increasing and continuous [12, Proposition 6.1]. (The proof requires the fact that $\|f_n - f\|_2 \to 0$ implies $I_r(f_n) \to I_r(f)$ and that $I_r(f)$ is lower semi-continuous on $\mathcal{W}$.)

The variational formulas in (3.1) achieve minimisers. In fact, the sets of minimiser are non-empty compact subsets of $\widetilde{\mathcal{W}}$ [12, Theorem 6.2]. In addition, all minimisers $h$ of $\phi_r^+(h)$ satisfy $h \geq r$ almost everywhere, while all minimisers $h$ of $\phi_r^-$ satisfy $h \leq r$ almost everywhere [12, Lemma 6.3]. Moreover, because

$$\begin{aligned} h_1 \geq h_2 \geq r &\implies \|T_{h_1}\| \geq \|T_{h_2}\|, \ I_r(h_1) \geq I_r(h_2), \\ h_1 \leq h_2 \leq r &\implies \|T_{h_1}\| \leq \|T_{h_2}\|, \ I_r(h_1) \leq I_r(h_2), \end{aligned} \tag{3.2}$$

(use that $a \mapsto \mathcal{R}(a \mid b)$ is unimodal on $[0, 1]$ with unique zero at $b$), it follows that both variational formulas achieve minimisers with norm equal to $\beta$, and so

$$\psi_r(\beta) = \begin{cases} \psi_r^+(\beta), & \beta \geq C_r, \\ \psi_r^-(\beta), & \beta \leq C_r. \end{cases} \tag{3.3}$$

Hence, $\psi_r$ is continuous and unimodal on $[0, 1]$. Since $I_r(h) = 0$ if and only if $h = r$ almost everywhere, it is immediate that $C_r$ is the unique zero of $\psi_r$.

(ii) The proof is by contradiction. Suppose that $\beta \mapsto \psi_r^+(\beta)$ is not strictly increasing on $[C_r, 1]$. Then there exist $\beta_1, \beta_2 \in [C_r, 1]$ with $\beta_1 < \beta_2$ such that $\psi_r^+$ is constant on $[\beta_1, \beta_2]$. Consequently, there exist minimisers $h_{\beta_1}^{\phi_1}, h_{\beta_2}^{\phi_2}$ with $\phi_1, \phi_2 \in \mathcal{M}$ satisfying $r \leq h_{\beta_1}^{\phi_1} \leq h_{\beta_2}^{\phi_2}$ such that

$$I_r(h_{\beta_1}^{\phi_1}) = I_r(h_{\beta_2}^{\phi_2}), \qquad \|T_{h_{\beta_1}^{\phi_1}}\| = \beta_1 < \beta_2 = \|T_{h_{\beta_2}^{\phi_2}}\|. \tag{3.4}$$

However, since $a \mapsto \mathcal{R}(a \mid b)$ is strictly increasing on $[b, 1]$ (recall (1.8)), it follows that $h_{\beta_1}^{\phi_1} = h_{\beta_2}^{\phi_2}$ almost everywhere. This in turn implies that $\|T_{h_{\beta_1}^{\phi_1}}\| = \|T_{h_{\beta_2}^{\phi_2}}\|$, which is a contradiction. A similar argument shows that $\beta \mapsto \psi_r^-(\beta)$ cannot have a flat piece on $[0, C_r]$. $\square$

### 3.2. **Perturbation around the minimum.** Note that when $\beta = C_r$, the infimum in (1.15) is attained at $h = r$ and $\psi_r(C_r) = 0$. Take $\beta = C_r + \epsilon$ with $\epsilon > 0$ small, and assume that the infimum is attained by a graphon of the form $h = r + \Delta_\epsilon$, where $\Delta_\epsilon \colon [0, 1]^2 \to \mathbb{R}$ represents a perturbation of the graphon $r$. Note that $r + \Delta_\epsilon \in \mathcal{W}$, and so we are dealing with a perturbation $\Delta_\epsilon$ that is symmetric and bounded. We compare

$$\psi_r(C_r + \epsilon) = \inf_{\substack{\Delta_\epsilon \colon [0,1]^2 \to \mathbb{R} \\ r + \Delta_\epsilon \in \mathcal{W} \\ \|T_{r+\Delta_\epsilon}\| = C_r + \epsilon}} I_r(r + \Delta_\epsilon) \tag{3.5}$$

with $\psi_r(C_r) = 0$ by computing the difference

$$\delta_r(\epsilon) = \psi_r(C_r + \epsilon) - \psi_r(C_r) = \psi_r(C_r + \epsilon) \tag{3.6}$$

and studying its behaviour as $\epsilon \to 0$. Since $r(x,y) = \nu(x)\nu(y)$, $x,y \in [0,1]$, we can use Lemma 2.1 to control the norm of $T_h = T_{r+\Delta_\epsilon}$. Pick $\bar{h} = r$ and $h = r + \Delta_\epsilon$ in (2.1) such that $\|\Delta_\epsilon\|_2 \to 0$ as $\epsilon \to 0$. Note that $\|T_{\Delta_\epsilon}\| \leq \|\Delta_\epsilon\|_2 < C_r$ for $\epsilon$ small enough. Hence, writing out the expansion for the norm, we get

$$\|T_{r+\Delta_\epsilon}\| = C_r + \sum_{n \in \mathbb{N}} \frac{1}{\|T_{r+\Delta_\epsilon}\|^n} \mathcal{F}_n(r + \Delta_\epsilon, r). \tag{3.7}$$

Since $\|T_{r+\Delta_\epsilon}\| = C_r + \epsilon$, we have

$$C_r + \epsilon = C_r + \frac{\langle \nu, \Delta_\epsilon \nu \rangle}{C_r + \epsilon} + \sum_{n \in \mathbb{N} \setminus \{1\}} \frac{1}{(C_r + \epsilon)^n} \langle \nu, \Delta_\epsilon^n \nu \rangle \tag{3.8}$$

with $\langle \nu, \Delta_\epsilon \nu \rangle = \int_{[0,1]^2} r\Delta_\epsilon$. So

$$\epsilon(C_r + \epsilon) = \int_{[0,1]^2} r\Delta_\epsilon + \sum_{n \in \mathbb{N} \setminus \{1\}} \frac{1}{(C_r + \epsilon)^{n-1}} \langle \nu, \Delta_\epsilon^n \nu \rangle. \tag{3.9}$$

Since $\nu$ is bounded, using the generalized Hölder's inequality [27, Theorem 3.1] we get

$$|\langle \nu, \Delta_\epsilon^n \nu \rangle| \leq \|\Delta_\epsilon\|_2^n. \tag{3.10}$$

Since $\|\Delta_\epsilon\|_2 \to 0$ as $\epsilon \to 0$, we can choose $\epsilon$ small enough such that $\|\Delta_\epsilon\|_2 < \frac{1}{2}(C_r + \epsilon)$, which gives

$$\sum_{n \in \mathbb{N} \setminus \{1\}} \frac{1}{(C_r + \epsilon)^{n-1}} \langle \nu, \Delta_\epsilon^n \nu \rangle = \mathrm{O}\left(\|\Delta_\epsilon\|_2^2\right). \tag{3.11}$$

The constraint $\|r + \Delta_\epsilon\| = C_r + \epsilon$ therefore reads

$$\int_{[0,1]^2} r\Delta_\epsilon = \epsilon C_r + \epsilon^2 + \mathrm{O}\left(\|\Delta_\epsilon\|_2^2\right). \tag{3.12}$$

Observe that if $\Delta_\epsilon = \epsilon\Delta$ for some function $\Delta \in L^2([0,1]^2)$, then

$$\int_{[0,1]^2} r\Delta = C_r \left[1 + \mathrm{o}(1)\right]. \tag{3.13}$$

3.2.1. *Small perturbation on a given region.* In what follows we use the standard notation $\mathrm{o}(\cdot)$, $\mathrm{O}(\cdot)$, $\asymp$ to describe the asymptotic behaviour in the limit as $\epsilon \to 0$. We start by considering different types of small perturbations in a given region and computing their total cost.

We are interested in the finding the asymptotic behaviour of (3.5). In the next lemma we show that it is enough to consider $\Delta_\epsilon$ of the form $\epsilon\Delta$ for some $\Delta \in L^2([0,1]^2)$, because these perturbations contribute to the minimum cost.

**Lemma 3.1. [Order of minimal cost]** *Let $\Delta_\epsilon : [0,1]^2 \to \mathbb{R}$ be such that $r + \Delta_\epsilon \in \mathcal{W}$ and $\|T_{r+\Delta_\epsilon}\| = C_r + \epsilon$. Then*

$$I_r(r + \Delta_\epsilon) \geq 2\epsilon^2. \tag{3.14}$$

*Moreover, if $\Delta_\epsilon = \epsilon\Delta$, then*

$$I_r(r + \epsilon\Delta) = [1 + \mathrm{o}(1)]\, 2\epsilon^2 \int_{[0,1]^2} \frac{\Delta^2}{4r(1-r)}, \qquad \epsilon \to 0. \tag{3.15}$$

*Proof.* Fix $b \in [0,1]$ and abbreviate (recall (1.9))

$$\chi(a) = \mathcal{R}(a \mid b) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}, \qquad a \in [0,1]. \tag{3.16}$$

Note that

$$\chi(b) = \chi'(b) = 0, \quad \chi''(a) \geq 4, \quad a \in [0,1]. \tag{3.17}$$

Consequently,

$$\chi(a) \geq 2(a-b)^2, \qquad a \in [0,1], \tag{3.18}$$

and hence

$$I_r(r + \Delta_\epsilon) \geq 2 \int_{[0,1]^2} \Delta_\epsilon^2 = 2\|\Delta_\epsilon\|_2^2. \tag{3.19}$$

Next observe that

$$C_r + \epsilon = \|T_{r+\Delta_\epsilon}\| = \|T_r + T_{\Delta_\epsilon}\| \leq \|T_r\| + \|T_{\Delta_\epsilon}\| \leq C_r + \|\Delta_\epsilon\|_2, \tag{3.20}$$

which gives $\|\Delta_\epsilon\|_2 \geq \epsilon$. Inserting this lower bound into (3.19), we get (3.14). To get (3.15), we need a higher-order expansion of $\chi$, namely, $\chi(x) = \frac{1}{2}\chi''(b)(x-b)^2 + O((x-b)^3)$, $x \to b$. Since $r$ is bounded away from 0 and 1, and the constraint $r + \Delta_\epsilon \in \mathcal{W}$ implies that $\Delta_\epsilon(x,y) \in [-1,1]$, we see that the third-order term is smaller than the second-order term when $\Delta_\epsilon = \epsilon\Delta$. Hence (3.15) follows. $\square$

**Lemma 3.2. [Cost of small perturbations]** *Let $B \subseteq [0,1]^2$ be a measurable region with area $|B|$. Suppose that $\Delta_\epsilon = \epsilon^\alpha \Delta$ on $B$, with $\epsilon > 0$, $\alpha > 0$ and $\Delta\colon [0,1]^2 \to \mathbb{R}$. Then the contribution of $B$ to the cost $I_r(h)$ is*

$$\int_B \mathrm{d}x\,\mathrm{d}y\,\mathcal{R}(h(x,y) \mid r(x,y)) = [1 + \mathrm{o}(1)]\,\epsilon^{2\alpha} \int_B \frac{\Delta^2}{2r(1-r)}, \qquad \epsilon \to 0. \tag{3.21}$$

*If the integral diverges, then the contribution decays slower than $\epsilon^{2\alpha}$.*

*Proof.* The proof is similar to that of Lemma 3.1. $\square$

3.2.2. *Approximation by block graphons.* We next introduce block graphons, which will be useful for our perturbation analysis. It follows from Lemma 3.1 that optimal perturbations with $\Delta_\epsilon$ must satisfy $\|\Delta_\epsilon\|_2 \asymp \epsilon$, and hence it is desirable to have $\Delta_\epsilon = \epsilon\Delta$. We argue through block graphon approximations that this is indeed the case.

**Definition 3.3. [Block graphons]** *Let $\mathcal{W}_N \subset \mathcal{W}$ be the space of graphons with $N$ blocks having a constant value on each of the blocks, i.e., $f \in \mathcal{W}_N$ is of the form*

$$f(x,y) = \begin{cases} f_{i,j}, & (x,y) \in B_i \times B_j, \\ 0, & \text{otherwise}, \end{cases} \tag{3.22}$$

*where $B_i = [\frac{i-1}{N}, \frac{i}{N})$, $1 \leq i \leq N-1$ and $B_N = [\frac{N-1}{N}, 1]$ and $f_{i,j} \in [0,1]$. Write $B_{i,j} = B_i \times B_j$. With each $f \in \mathcal{W}$ associate the block graphon $f_N \in \mathcal{W}_N$ given by*

$$f_N(x,y) = N^2 \int_{B_{i,j}} \mathrm{d}x'\,\mathrm{d}y'\,f(x',y') = \bar{f}_{N,ij}, \quad (x,y) \in B_{i,j}. \tag{3.23}$$

Observe that if $f_N$ is the block graphon associated with a graphon $f$, then

$$\|T_{f_N} - T_f\| = \|T_{f_N - f}\| \leq \|f_N - f\|_2. \tag{3.24}$$

We know from [12, Proposition 2.6] that $\|f_N - f\|_2 \to 0$, and hence $\lim_{N\to\infty} \|T_{f_N}\| = \|T_f\|$ for any $f \in \mathcal{W}$ and its associated sequence of block graphons $(f_N)_{N\in\mathbb{N}}$. The following lemma shows that the cost function associated with the graphons $r$ and $f$ is well approximated by the cost function associated with the block graphons $r_N$ and $f_N$.

**Lemma 3.4.** [**Convergence of the cost function**] $\lim_{N\to\infty} I_{r_N}(f_N) = I_r(f)$ *for any* $f \in \mathcal{W}$.

*Proof.* Since $f \in L^2([0,1]^2)$, $f_N$ is bounded. The assumption in (1.4) implies that $\eta \leq r_N \leq 1-\eta$ for all $N \in \mathbb{N}$. We know from [17, Lemma 2.3] that there exists a constant $c > 0$ independent of $f$ such that

$$|I_{r_N}(f) - I_r(f)| \leq c\,\|r_N - r\|_1 \leq c\,\|r_N - r\|_2. \tag{3.25}$$

Hence

$$\begin{aligned}
|I_{r_N}(f_N) - I_r(f)| &\leq |I_{r_N}(f_N) - I_r(f_N)| + |I_r(f_N) - I_r(f)| \\
&\leq c\,\|r_N - r\|_2 + |I_r(f_N) - I_r(f)|.
\end{aligned} \tag{3.26}$$

But we know from [12, Proposition 2.6] that $\lim_{N\to\infty} \|r_N - r\|_2 = 0$ and $\lim_{N\to\infty} \|f_N - f\|_2 = 0$, while we know from [28, Lemma 3.4] that $I_r$ is continuous in the $L^2$-topology on $\mathcal{W}$. $\square$

3.2.3. *Block graphon perturbations.* In what follows we fix $N \in \mathbb{N}$, analyse different types of perturbation and identify which one is optimal. For each $N \in \mathbb{N}$, we associate with the perturbed graphon $h = r + \Delta_\epsilon$ the block graphon $h_N \in \mathcal{W}_N$ given by

$$\overline{h}_{N,ij}(x,y) = \overline{r}_{N,ij}(x,y) + \overline{\Delta_\epsilon}_{N,ij}(x,y), \quad (x,y) \in B_{i,j}, \tag{3.27}$$

with

$$\overline{r}_{N,ij} = N^2 \int_{B_{i,j}} \mathrm{d}x'\,\mathrm{d}y'\, r(x',y'), \qquad \overline{\Delta_\epsilon}_{N,ij} = N^2 \int_{B_{i,j}} \mathrm{d}x'\,\mathrm{d}y'\, \Delta_\epsilon(x',y'). \tag{3.28}$$

Observe that optimal perturbations must have $\|\Delta_\epsilon\|_2 = \mathrm{O}(\epsilon)$, and hence the constraint in (3.12) becomes

$$\sum_{i,j=1}^{N} \int_{B_{i,j}} r\Delta_\epsilon = \sum_{i,j=1}^{N} \frac{1}{N^2} \overline{r\Delta_\epsilon}_{N,ij} = [1 + \mathrm{o}(1)]\, C_r \epsilon, \qquad \epsilon \to 0. \tag{3.29}$$

The block constraint in (3.29) implies that the sum over each block must be of order $\epsilon$. We therefore must have that

$$\overline{r\Delta_\epsilon}_{N,ij} = \mathrm{O}(\epsilon), \qquad \epsilon \to 0 \qquad \forall\,(i,j), \tag{3.30}$$

which means that

$$\overline{\Delta_\epsilon}_{N,ij} = \mathrm{O}(\epsilon), \qquad \epsilon \to 0 \qquad \forall\,(i,j), \tag{3.31}$$

since (1.4) implies that $\overline{r\Delta_\epsilon}_{N,ij} \asymp \overline{\Delta_\epsilon}_{N,ij}$. There are two possible cases:

    (I) All blocks contribute to the constraint with a term of order $\epsilon$ (balanced perturbation).
    (II) Some blocks contribute to the constraint with a term of order $\epsilon$ and some with $\mathrm{o}(\epsilon)$ (unbalanced perturbation).

Perturbations of type (I) consist of a small perturbation on each block, i.e., $\overline{\Delta_\epsilon}_{N,ij} \asymp \epsilon$ for each block $B_{i,j}$. By Lemma 3.2, this contributes a term of order $\epsilon^2$ to the total cost. Since all blocks have the same type of perturbation, they all contribute in the same way, and so we get $I_{r_N}(h_N) \asymp \epsilon^2$. We will see in Corollary 3.6 that perturbations of type (II) are worse than perturbations of type (I). Let $1 \leq k \leq N^2 - 1$ be the number of blocks that contribute a term of order $\mathrm{o}(\epsilon)$ to the constraint, i.e., $\overline{\Delta_\epsilon}_{N,ij} = \mathrm{o}(\epsilon)$. By Lemma 3.2, these blocks contribute order $\mathrm{o}(\epsilon^2)$ to the total cost. The remaining blocks must fall in the class of blocks of type (I), with a perturbation of order $\epsilon$ on each of them. Lemma 3.6 below shows that the cost function attains its infimum when the perturbation of order $\epsilon$ is uniform on $[0,1]^2$.

3.2.4. *Optimal perturbation.* We have shown that perturbations of type (I) lead to the minimal total cost. They consist of small perturbations of order $\epsilon$ on all blocks, and hence on $[0,1]^2$. A sequence of such perturbations $(\Delta_{\epsilon,N})_{N\in\mathbb{N}}$ converges to a perturbation $\Delta_\epsilon$ as $N \to \infty$. We can identify the cost of $\Delta_\epsilon = \epsilon\Delta$ with $\Delta \colon [0,1]^2 \to \mathbb{R}$, which we refer to as a balanced perturbation.

**Lemma 3.5. [Balanced perturbations]** *Suppose that $\Delta_\epsilon = \epsilon\Delta$ with $\Delta \colon [0,1]^2 \to \mathbb{R}$. Let $\mathcal{M}$ be the set of Lebesgue measure-preserving bijective maps. Then*

$$\delta_r(\epsilon) = [1 + \mathrm{o}(1)]\, K_r \epsilon^2, \qquad \epsilon \to 0, \tag{3.32}$$

*with*

$$K_r = \frac{1}{2} C_r^2 \inf_{\phi \in \mathcal{M}} \frac{D_r^\phi}{(B_r^\phi)^2}, \tag{3.33}$$

*where $B_r^\phi = \int_{[0,1]^2} r^\phi r^2(1-r)$ and $D_r^\phi = \int_{[0,1]^2} (r^\phi)^2 r(1-r)$.*

*Proof.* The constraint in (3.12) becomes

$$\int_{[0,1]^2} r\Delta = [1 + \mathrm{o}(1)]\, C_r, \qquad \epsilon \to 0, \tag{3.34}$$

and we get

$$\begin{aligned}
\delta_r(\epsilon) &= \inf_{\substack{\Delta \colon [0,1]^2 \to \mathbb{R} \\ r+\epsilon\Delta \in \mathcal{W} \\ \int_{[0,1]^2} r\Delta = [1+\mathrm{o}(1)]\, C_r}} I_r(r + \epsilon\Delta) \\
&= \inf_{\substack{\Delta \colon [0,1]^2 \to \mathbb{R} \\ r+\epsilon\Delta \in \mathcal{W} \\ \int_{[0,1]^2} r\Delta = [1+\mathrm{o}(1)]\, C_r}} \int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\, \mathcal{R}\big((r + \epsilon\Delta)(x,y) \mid r(x,y)\big).
\end{aligned} \tag{3.35}$$

By Lemma 3.2 (with $\alpha = 1$), we have

$$\delta_r(\epsilon) = [1 + \mathrm{o}(1)]\, K_r \epsilon^2, \qquad \epsilon \to 0, \tag{3.36}$$

with

$$K_r = \inf_{\substack{\Delta \colon [0,1]^2 \to \mathbb{R} \\ \int_{[0,1]^2} r\Delta = C_r}} \int_{[0,1]^2} \frac{\Delta^2}{2r(1-r)}. \tag{3.37}$$

The prefactor $1+\mathrm{o}(1)$ in (3.36) arises after we scale $\Delta$ by $1+\mathrm{o}(1)$ in order to force $\int_{[0,1]^2} r\Delta = C_r$. Note that the optimisation problem in (3.37) no longer depends on $\epsilon$.

We can apply the method of Lagrange multipliers to solve this constrained optimisation problem. To that end we define the Lagrangian

$$\mathcal{L}_{A_r}(\Delta) = \int_{[0,1]^2} \frac{\Delta^2}{2\,r(1-r)} + A_r \int_{[0,1]^2} r\Delta, \tag{3.38}$$

where $A_r$ is a Langrange multiplier. Since $\int_{[0,1]^2} r = \int_{[0,1]^2} r^\phi$ for any Lebesgue measure-preserving bijective map $\phi \in \mathcal{M}$, we get that the minimizer (in the space of functions from $[0,1]^2 \to \mathbb{R}$) is of the form

$$\Delta^\phi = -A_r\, r^\phi r(1-r), \quad \phi \in \mathcal{M}. \tag{3.39}$$

We pick $A_r$ such that the constraint is satisfied, i.e.,

$$-A_r B_r^\phi = [1 + \mathrm{o}(1)]\, C_r \tag{3.40}$$

with

$$B_r^\phi = \int_{[0,1]^2} r^\phi r^2 (1-r). \tag{3.41}$$

We get

$$\Delta^\phi = \frac{C_r}{B_r^\phi} \, r^\phi r(1-r), \quad \phi \in \mathcal{M}, \tag{3.42}$$

and

$$K_r = \inf_{\phi \in \mathcal{M}} \int_{[0,1]^2} \frac{(\Delta^\phi)^2}{2r(1-r)} = \frac{1}{2} C_r^2 \inf_{\phi \in \mathcal{M}} \frac{D_r^\phi}{(B_r^\phi)^2} \tag{3.43}$$

with

$$D_r^\phi = \int_{[0,1]^2} (r^\phi)^2 r(1-r). \tag{3.44}$$

$\square$

We next show that the infimum in (3.43) is uniquely attained when $\phi$ is the identity. For this we will show that $D_r^\phi/(B_r^\phi)^2 \geq 1/B_r$ with equality if and only if $\phi = \mathrm{Id}$. Indeed, write

$$\begin{aligned}
&B_r D_r^\phi - (B_r^\phi)^2 \\
&= \int_{[0,1]^2} \mathrm{d}x \, \mathrm{d}y \, r(x,y)[1 - r(x,y)] \int_{[0,1]^2} \mathrm{d}\bar{x} \, \mathrm{d}\bar{y} \, r(\bar{x},\bar{y})[1 - r(\bar{x},\bar{y})] \\
&\quad \times \left\{ r(x,y)^2 \, r^\phi(\bar{x},\bar{y})^2 - r(x,y)r^\phi(x,y) \, r(\bar{x},\bar{y})r^\phi(\bar{x},\bar{y}) \right\} \\
&= \int_{[0,1]^2} \mathrm{d}x \, \mathrm{d}y \, r(x,y)[1 - r(x,y)] \int_{[0,1]^2} \mathrm{d}\bar{x} \, \mathrm{d}\bar{y} \, r(\bar{x},\bar{y})[1 - r(\bar{x},\bar{y})] \\
&\quad \times \tfrac{1}{2} \left\{ r(x,y)^2 \, r^\phi(\bar{x},\bar{y})^2 + r^\phi(x,y)^2 \, r(\bar{x},\bar{y})^2 - 2r(x,y)r^\phi(x,y) \, r(\bar{x},\bar{y})r^\phi(\bar{x},\bar{y}) \right\} \\
&= \int_{[0,1]^2} \mathrm{d}x \, \mathrm{d}y \, r(x,y)[1 - r(x,y)] \int_{[0,1]^2} \mathrm{d}\bar{x} \, \mathrm{d}\bar{y} \, r(\bar{x},\bar{y})[1 - r(\bar{x},\bar{y})] \\
&\quad \times \tfrac{1}{2} \Big( r(x,y)r^\phi(\bar{x},\bar{y}) - r^\phi(x,y)r(\bar{x},\bar{y}) \Big)^2,
\end{aligned} \tag{3.45}$$

where the second equality uses the symmetry between the integrals. Hence $B_r D_r^\phi - (B_r^\phi)^2 \geq 0$, with equality if and only if $r(x,y)/r^\phi(x,y) = C$ for almost every $x,y \in [0,1]^2$. Clearly, this can hold only for $C = 1$, which amounts to $\phi = \mathrm{Id}$.

We conclude that the infimum in (3.43) equals $1/B_r$, and so we find that

$$K_r = \frac{C_r^2}{2B_r}. \tag{3.46}$$

Finally, note that $C_r = m_2$ by (2.8), and that $B_r = m_3^2 - m_4^2$ by (1.18). This settles the expression for $K_r$ in (1.21).

**Corollary 3.6.** [*Unbalanced perturbations*] *Perturbations of order $\epsilon$ that are not balanced, i.e., that do not cover the entire unit square $[0,1]^2$, are worse than the balanced perturbation in Lemma 3.5.*

*Proof.* The argument of the variational formula can be reduced to an integral that considers only those regions that contribute order $\epsilon^2$, which constitute a subset $A \subset [0,1]^2$. Applying the method of Lagrange multipliers as in Lemma 3.5, we obtain that the solution is given by

$$\delta_r(\epsilon) = [1 + \mathrm{o}(1)] \, K_r' \epsilon^2, \qquad \epsilon \to 0, \tag{3.47}$$

with $K_r' > K_r$. The strict inequality comes from the fact that the optimal balanced perturbation $\Delta^{\mathrm{Id}}$ found in (3.39) is non-zero everywhere. $\qquad\square$

In conclusion, we have shown that a balanced perturbation is optimal and we have identified in (3.42) the form of the optimal balanced perturbation. Lemma 3.5 settles the claim in Theorem 1.6(i), while (3.42) settles the claim in Theorem 1.7(i).

3.3. **Perturbation near the right end.** Take $\beta = 1 - \epsilon$ and consider a graphon of the form $h = 1 - \Delta_\epsilon$, where $\Delta_\epsilon \colon [0,1]^2 \to [0,\infty)$ represents a symmetric and bounded perturbation of the constant graphon $h \equiv 1$. We compare

$$\psi_r(1 - \epsilon) = \inf_{\substack{\Delta_\epsilon \colon [0,1]^2 \to [0,\infty) \\ 1 - \Delta_\epsilon \in \mathcal{W} \\ \|T_{1-\Delta_\epsilon}\| = 1-\epsilon}} I_r(1 - \Delta_\epsilon) \tag{3.48}$$

with

$$C_r^1 = I_r(1) \tag{3.49}$$

by computing the difference

$$\delta_r(\epsilon) = \psi_r(1) - \psi_r(1 - \epsilon) \tag{3.50}$$

and studying its behaviour as $\epsilon \downarrow 0$. Since $I_r(1)$ is a constant, we can write

$$\delta_r(\epsilon) = \sup_{\substack{\Delta_\epsilon \colon [0,1]^2 \to [0,\infty) \\ 1 - \Delta_\epsilon \in \mathcal{W} \\ \|T_{1-\Delta_\epsilon}\| = 1-\epsilon}} [I_r(1) - I_r(1 - \Delta_\epsilon)]. \tag{3.51}$$

We again use the expansion in Lemma 2.1. Pick $\bar{h} = 1$ and $h = 1 - \Delta_\epsilon$ in (2.1), to get

$$\|T_{1-\Delta_\epsilon}\| = 1 + \sum_{n \in \mathbb{N}} \frac{1}{\|T_{1-\Delta_\epsilon}\|^n} \, \mathcal{F}_n(1 - \Delta_\epsilon, 1). \tag{3.52}$$

Since $\|T_{1-\Delta_\epsilon}\| = 1 - \epsilon$, this gives

$$1 - \epsilon = 1 + \frac{\langle 1, (-\Delta_\epsilon)1 \rangle}{1 - \epsilon} + \frac{\langle 1, (-\Delta_\epsilon)^2 1 \rangle}{(1 - \epsilon)^2} + \sum_{n \in \mathbb{N} \setminus \{1,2\}} \frac{\langle 1, (-\Delta_\epsilon)^n 1 \rangle}{(1 - \epsilon)^n}. \tag{3.53}$$

For $\epsilon \downarrow 0$ we have $\|\Delta_\epsilon\|_2 \downarrow 0$ and $|\langle 1, (-\Delta_\epsilon)^n 1 \rangle| = \mathrm{O}(\|\Delta_\epsilon\|_2^n)$. Therefore

$$\epsilon(1 - \epsilon) = \int_{[0,1]^2} \Delta_\epsilon - \frac{\langle 1, \Delta_\epsilon^2 1 \rangle}{(1 - \epsilon)} + \mathrm{O}\left(\|\Delta_\epsilon\|_2^3\right). \tag{3.54}$$

The restriction $1 - \Delta_\epsilon \in \mathcal{W}$ implies that $\Delta_\epsilon \in [0,1]$. Hence $\|\Delta_\epsilon\|_2^2 \leq \|\Delta_\epsilon\|_1$. Moreover,

$$1 - \epsilon = \|T_{1-\Delta_\epsilon}\| \leq \|1 - \Delta_\epsilon\|_2 \leq \sqrt{\|1 - \Delta_\epsilon\|_1}. \tag{3.55}$$

Since $\|1 - \Delta_\epsilon\|_1 = 1 - \|\Delta_\epsilon\|_1$, we have

$$\|\Delta_\epsilon\|_1 \leq 1 - (1 - \epsilon)^2 = \epsilon(2 - \epsilon). \tag{3.56}$$

Since $\|\Delta_\epsilon\|_2^3 = \mathrm{O}(\epsilon^{3/2})$, (3.54) reads

$$\frac{1}{1 - \epsilon} \int_{[0,1]^3} \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, \Delta_\epsilon(x,y)[1 - \Delta_\epsilon(y,z)] - \frac{\epsilon}{1 - \epsilon} \|\Delta_\epsilon\|_1 = \epsilon(1 - \epsilon) + \mathrm{O}(\epsilon^{3/2}), \tag{3.57}$$

which, because $\|\Delta_\epsilon\|_1 = \mathrm{O}(\epsilon)$, further reduces to

$$\int_{[0,1]^3} \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \, \Delta_\epsilon(x,y)[1 - \Delta_\epsilon(y,z)] = \epsilon \left[1 + \mathrm{O}(\epsilon^{1/2})\right]. \tag{3.58}$$

Note that when $\Delta_\epsilon = \epsilon\Delta$, the constraint reads

$$\int_{[0,1]^2} \Delta = 1 + O(\epsilon^{1/2}), \qquad \epsilon \downarrow 0. \tag{3.59}$$

The following lemma gives an upper bound for $I_r(1) - I_r(1 - \Delta_\epsilon)$.

**Lemma 3.7. [Order of minimal cost]** *Let* $\Delta_\epsilon : [0,1]^2 \to [0,1]$ *be such that* $1 - \Delta_\epsilon \in \mathcal{W}$ *and* $\|T_{1-\Delta_\epsilon}\| = 1 - \epsilon$. *Then, for* $\epsilon$ *small enough,*

$$I_r(1) - I_r(1 - \Delta_\epsilon) \le \|\Delta_\epsilon\|_1 \log \frac{1}{\|\Delta_\epsilon\|_1} + O(\|\Delta_\epsilon\|_1). \tag{3.60}$$

*Moreover,* $\delta_r(\epsilon) \le \epsilon \log \frac{1}{\epsilon} + O(\epsilon)$.

*Proof.* Abbreviate (recall (1.9))

$$\chi(a) = \mathcal{R}(a \mid r) = a \log \frac{a}{r} + (1-a) \log \frac{1-a}{1-r}, \quad a \in [0,1]. \tag{3.61}$$

Then

$$\begin{aligned}
&\chi(1) - \chi(1 - \Delta_\epsilon(x,y)) \\
&= \Delta_\epsilon(x,y) \log \left( \frac{1 - \Delta_\epsilon(x,y)}{\Delta_\epsilon(x,y)} \frac{1 - r(x,y)}{r(x,y)} \right) - \log(1 - \Delta_\epsilon(x,y)),
\end{aligned} \tag{3.62}$$

and so

$$I_r(1) - I_r(1 - \Delta_\epsilon) = \int_{[0,1]^2} \left[ \Delta_\epsilon \log \left( \frac{1 - \Delta_\epsilon}{\Delta_\epsilon} \frac{1 - r}{r} \right) - \log(1 - \Delta_\epsilon) \right]. \tag{3.63}$$

Let $\mu_\epsilon$ be the probability measure on $[0,1]^2$ whose density with respect to the Lebesgue measure is $Z_\epsilon^{-1}(1 - \Delta_\epsilon(x,y))$, where $Z_\epsilon = \int_{[0,1]^2}(1 - \Delta_\epsilon) = 1 - O(\epsilon)$. Since $u \mapsto \bar{s}(u) = u \log(1/u)$ is strictly concave, by Jensen's inequality we have

$$\int_{[0,1]^2} \Delta_\epsilon \log \left( \frac{1 - \Delta_\epsilon}{\Delta_\epsilon} \right) = Z_\epsilon \int_{[0,1]^2} \mu_\epsilon \bar{s} \left( \frac{\Delta_\epsilon}{1 - \Delta_\epsilon} \right) \le Z_\epsilon \bar{s} \left( Z_\epsilon^{-1} \|\Delta_\epsilon\|_1 \right) = \|\Delta_\epsilon\|_1 \log \left( \frac{Z_\epsilon}{\|\Delta_\epsilon\|_1} \right). \tag{3.64}$$

Moreover,

$$\int_{[0,1]^2} \Delta_\epsilon \log \left( \frac{1 - r}{r} \right) = O(\|\Delta_\epsilon\|_1), \qquad -\int_{[0,1]^2} \log(1 - \Delta_\epsilon) = O(\|\Delta_\epsilon\|_1). \tag{3.65}$$

Hence

$$I_r(1) - I_r(1 - \Delta_\epsilon) \le \|\Delta_\epsilon\|_1 \log \frac{1}{\|\Delta_\epsilon\|_1} + O(\|\Delta_\epsilon\|_1), \qquad \epsilon \downarrow 0, \tag{3.66}$$

and since $\|\Delta_\epsilon\|_1 = O(\epsilon)$ also $\delta_r(\epsilon) \le \epsilon \log \frac{1}{\epsilon} + O(\epsilon)$. $\qquad\qquad \square$

The following is the analogue of Lemma 3.2 for perturbations near the right end.

**Lemma 3.8. [Cost of small perturbations]** *Let* $B \subseteq [0,1]^2$ *be a measurable region of area* $|B|$. *Suppose that* $\Delta_\epsilon = \epsilon^\alpha \Delta$ *on* $B$ *with* $\epsilon > 0$, $\alpha \in (0,1]$ *and* $\Delta : [0,1]^2 \to [0,\infty)$. *Then the contribution of* $B$ *to the cost* $I_r(h)$ *is*

$$\int_B dx\, dy \left[ \mathcal{R}(1 \mid r(x,y)) - \mathcal{R}(1 - \epsilon^\alpha \Delta(x,y) \mid r(x,y)) \right] = [1 + o(1)] \int_B \epsilon^\alpha \Delta \log \left( \frac{1-r}{\epsilon^\alpha \Delta r} \right), \quad \epsilon \downarrow 0. \tag{3.67}$$

*Proof.* The proof is the same as that of Lemma 3.7, with the observation that

$$\mathcal{R}(1 \mid r) - \mathcal{R}(1 - \epsilon^\alpha \Delta \mid r) = [1 + o(1)] \,\epsilon^\alpha \Delta \log\left(\frac{1-r}{\epsilon^\alpha \Delta r}\right), \qquad \epsilon \downarrow 0. \tag{3.68}$$

$\square$

Following the argument in Section 3.2, we can approximate the cost function by using block graphons. We see

$$\sum_{i,j=1}^{N} \int_{B_{i,j}} \mathrm{d}x \, \mathrm{d}y \, \Delta_{\epsilon,N}(x,y) = \sum_{i,j=1}^{N} \frac{1}{N^2} \overline{\Delta}_{\epsilon N, ij} = \epsilon \, [1 + o(1)], \qquad \epsilon \downarrow 0. \tag{3.69}$$

The block constraint in (3.69) implies that the sum over each block must be of order $\epsilon$. Hence

$$\overline{\Delta}_{\epsilon N, ij} = \mathrm{O}(\epsilon), \qquad \epsilon \downarrow 0 \qquad \forall \, (i,j). \tag{3.70}$$

There are two cases to distinguish: all blocks contribute to the constraint with a term of order $\epsilon$ (balanced perturbation), or some of the blocks contribute to the constraint with a term of order $\epsilon$ and some with $o(\epsilon)$. Analogously to the analysis in Section 3.2, by using Lemma 3.8 we can compute the total cost that different types of block perturbations produce. This again shows that the optimal perturbations are the balanced perturbations, consisting of perturbations of order $\epsilon$ on every block. As $N \to \infty$, a sequence of such perturbations converges to a perturbation $\Delta_\epsilon = \epsilon \Delta$ with $\Delta \colon [0,1]^2 \to [0,\infty)$, which we analyse next.

**Lemma 3.9. [Balanced perturbations]** *Suppose that* $\Delta_\epsilon = \epsilon \Delta$ *with* $\Delta \colon [0,1]^2 \to [0,\infty)$. *Then*

$$\delta_r(\epsilon) = [1 + \mathrm{O}(\epsilon^{1/2})] \left\{ \epsilon + \epsilon \log\left(\frac{N_r^1}{\epsilon}\right) \right\} + \mathrm{O}(\epsilon^2), \qquad \epsilon \downarrow 0. \tag{3.71}$$

*Proof.* By (3.59) and (3.63),

$$
\begin{aligned}
\delta_r(\epsilon) &= \sup_{\substack{\Delta \colon [0,1]^2 \to [0,\infty) \\ 1 - \epsilon \Delta \in \mathcal{W} \\ \int_{[0,1]^2} \Delta = 1 + \mathrm{O}(\epsilon^{1/2})}} \left[ I_r(1) - I_r(1 - \epsilon\Delta) \right] \\
&= \sup_{\substack{\Delta \colon [0,1]^2 \to [0,\infty) \\ 1 - \epsilon \Delta \in \mathcal{W} \\ \int_{[0,1]^2} \Delta = 1 + \mathrm{O}(\epsilon^{1/2})}} \int_{[0,1]^2} \left[ \epsilon \Delta \log\left(\frac{1 - \epsilon\Delta}{\epsilon\Delta} \frac{1-r}{r}\right) - \log(1 - \epsilon\Delta) \right].
\end{aligned}
\tag{3.72}
$$

The integral in (3.72) equals

$$\int_{[0,1]^2} \left[ \epsilon \Delta \log\left(\frac{1-r}{\epsilon \Delta r}\right) - (1 - \epsilon\Delta) \log(1 - \epsilon\Delta) \right] = \int_{[0,1]^2} \epsilon \Delta \log\left(\frac{1-r}{\epsilon \Delta r}\right) + \epsilon \int_{[0,1]^2} \Delta + \mathrm{O}(\epsilon^2). \tag{3.73}$$

Hence

$$\delta_r(\epsilon) = [1 + \mathrm{O}(\epsilon^{1/2})] \left\{ \epsilon + \sup_{\substack{\Delta \colon [0,1]^2 \to [0,\infty) \\ \int_{[0,1]^2} \Delta = 1}} \int_{[0,1]^2} \epsilon \Delta \log\left(\frac{1-r}{\epsilon \Delta r}\right) \right\} + \mathrm{O}(\epsilon^2), \tag{3.74}$$

where the prefactor arises after we scale $\Delta$ by $1 + \mathrm{O}(\epsilon^{1/2})$ in order to force $\int_{[0,1]^2} \Delta = 1$. Note that the constraint under the supremum no longer depends on $\epsilon$.

We can solve the optimisation problem by applying the method of Lagrange multipliers. To that end we define the Lagrangian

$$\mathcal{L}_{A_r}(\Delta) = \int_{[0,1]^2} \epsilon\Delta \log\left(\frac{1-r}{\epsilon\Delta r}\right) + A_r \int_{[0,1]^2} \Delta, \tag{3.75}$$

where $A_r$ is a Langrange multiplier. Since $\int_{[0,1]^2} \log\frac{1-r}{r} = \int_{[0,1]^2} \log\frac{1-r^\phi}{r^\phi}$ for any Lebesgue measure-preserving bijective map $\phi \in \mathcal{M}$, we get that the minimizer (in the space of functions from $[0,1]^2 \to \mathbb{R}$) is of the form

$$\Delta^\phi = \mathrm{e}^{-\frac{\epsilon-A_r}{\epsilon}}\frac{1}{\epsilon}\frac{1-r^\phi}{r^\phi}, \quad \phi \in \mathcal{M}. \tag{3.76}$$

We pick $A_r$ such that the constraint $\int_{[0,1]^2} \Delta = 1$ is satisfied. This gives

$$\Delta^\phi = \frac{1}{N_r^1}\frac{1-r^\phi}{r^\phi}, \quad \phi \in \mathcal{M}, \tag{3.77}$$

with $N_r^1 = \int_{[0,1]^2} \frac{1-r}{r}$. Hence the supremum in (3.74) becomes

$$\sup_{\phi\in\mathcal{M}} \int_{[0,1]^2} \epsilon\Delta^\phi \log\left(\frac{1-r}{\epsilon\Delta^\phi r}\right). \tag{3.78}$$

We have

$$\int_{[0,1]^2} \epsilon\Delta^\phi \log\left(\frac{1-r}{\epsilon\Delta^\phi r}\right) = \epsilon\log\left(\frac{N_r^1}{\epsilon}\right) - \epsilon\int_{[0,1]^2} \Delta^\phi \log\left(\frac{\Delta^\phi}{\Delta}\right), \tag{3.79}$$

where we use that $\int_{[0,1]^2} \Delta^\phi = 1$. Since the function $u \mapsto s(u) = u\log u$ is strictly convex on $[0,\infty)$, Jensen's inequality gives

$$\int_{[0,1]^2} \Delta^\phi \log\left(\frac{\Delta^\phi}{\Delta}\right) = \int_{[0,1]^2} \Delta\, s\left(\frac{\Delta^\phi}{\Delta}\right) \geq s\left(\int_{[0,1]^2} \Delta\frac{\Delta^\phi}{\Delta}\right) = s\left(\int_{[0,1]^2} \Delta^\phi\right) = s(1) = 0, \tag{3.80}$$

where we use that $\int_{[0,1]^2} \Delta = 1$. Equality holds if and only if $\Delta = \Delta^\phi$ almost everywhere on $[0,1]^2$, which amounts to $\phi = \mathrm{Id}$. Hence the supremum in (3.78) is uniquely attained at $\phi = \mathrm{Id}$ and equals

$$\int_{[0,1]^2} \epsilon\frac{1}{N_r^1}\frac{(1-r)}{r}\log\left(\frac{N_r^1}{\epsilon}\right) = \epsilon\log\left(\frac{N_r^1}{\epsilon}\right). \tag{3.81}$$

Consequently, (3.74) gives (3.71). □

Lemma 3.9 settles the claim in Theorem 1.6(ii). Since we have shown that a balanced perturbation is optimal, (3.77) settles the claim in Theorem 1.7(ii).

3.4. **Perturbation near the left end.** Take $\beta = \epsilon$ and consider a graphon of the form $h = \Delta_\epsilon$, where $\Delta_\epsilon\colon [0,1]^2 \to [0,\infty)$ represents a symmetric and bounded perturbation of the constant graphon $h \equiv 0$. We compare

$$\psi_r(\epsilon) = \inf_{\substack{\Delta_\epsilon\colon [0,1]^2\to[0,\infty)\\ \Delta_\epsilon\in\mathcal{W}\\ \|\Delta_\epsilon\|=\epsilon}} I_r(\Delta_\epsilon) \tag{3.82}$$

with

$$\psi_r(0) = I_r(0) \tag{3.83}$$

by computing the difference

$$\delta_r(\epsilon) = \psi_r(\epsilon) - \psi_r(0) \tag{3.84}$$

and studying its behaviour as $\epsilon \to 0$.

We claim that analysing (3.84) is equivalent to analysing

$$\delta_{\hat{r}}(\epsilon) = \phi_{\hat{r}}(1) - \phi_{\hat{r}}(1 - \epsilon), \tag{3.85}$$

where $\hat{r}$ is the *reflection* of $r$ defined as

$$\hat{r}(x, y) = 1 - r(x, y), \qquad x, y \in [0, 1]. \tag{3.86}$$

Indeed,

$$I_r(0) = \int_{[0,1]^2} \mathcal{R}(0 \mid r) = \int_{[0,1]^2} \log\left(\frac{1}{1-r}\right) = \int_{[0,1]^2} \mathcal{R}(1 \mid \hat{r}) = I_{\hat{r}}(1) \tag{3.87}$$

and

$$\begin{aligned}
I_r(\Delta_\epsilon) &= \int_{[0,1]^2} \mathcal{R}(\Delta_\epsilon \mid r) = \int_{[0,1]^2} \left[\Delta_\epsilon \log\left(\frac{\Delta_\epsilon}{r}\right) + (1 - \Delta_\epsilon) \log\left(\frac{1 - \Delta_\epsilon}{1 - r}\right)\right] \\
&= \int_{[0,1]^2} \mathcal{R}(1 - \Delta_\epsilon \mid \hat{r}) = I_{\hat{r}}(1 - \Delta_\epsilon).
\end{aligned} \tag{3.88}$$

We can therefore use the results in Section 3.3. From Lemma 3.9 we know that

$$\delta_{\hat{r}}(\epsilon) = [1 + \mathrm{O}(\epsilon^{1/2})] \left\{\epsilon + \epsilon \log\left(\frac{N_r^1}{\epsilon}\right)\right\} + \mathrm{O}(\epsilon^2), \qquad \epsilon \downarrow 0, \tag{3.89}$$

and hence we obtain

$$\delta_r(\epsilon) = [1 + \mathrm{O}(\epsilon^{1/2})] \left\{\epsilon + \epsilon \log\left(\frac{N_r^0}{\epsilon}\right)\right\} + \mathrm{O}(\epsilon^2), \qquad \epsilon \downarrow 0, \tag{3.90}$$

Consequently, the optimal perturbation is given by the balanced perturbation $\Delta_\epsilon = \epsilon \Delta$ with

$$\Delta = \frac{1}{N_r^0} \frac{r}{1 - r}, \tag{3.91}$$

with $N_r^0 = \int_{[0,1]^2} \frac{r}{1-r}$.

The scaling in (3.90) settles the claim in Theorem 1.6(iii). Since we have shown that a balanced perturbation is optimal, (3.91) settles the claim in Theorem 1.7(iii).

## Appendix A. Appendix

**Lemma A.1. [Finite-rank expansion]** *Consider a graphon $\bar{h} \in \mathcal{W}$ such that*

$$\bar{h}(x, y) = \sum_{i=1}^{k} \theta_i \bar{\nu}_i(x) \bar{\nu}_i(y), \qquad x, y \in [0, 1], \tag{A.1}$$

*for some $k \in \mathbb{N}$, where $\theta_1 > \theta_2 \geq \ldots \geq \theta_k \geq 0$ and $\{\bar{\nu}_1, \bar{\nu}_2, \ldots, \bar{\nu}_k\}$ is an orthonormal set in $L^2[0, 1]$. Then there exists an $\varepsilon > 0$ such that, for any $h \in \mathcal{W}$ satisfying $\|T_{h-\bar{h}}\| < \min(\varepsilon, \|T_h\|)$, the operator norm $\|T_h\|$ solves the equation*

$$\|T_h\| = \lambda_k\left(\sum_{n \in \mathbb{N}_0} \|T_h\|^{-n} \mathcal{F}_n(h, \bar{h})\right), \tag{A.2}$$

*where $\lambda_k(M)$ is the largest eigenvalue of a $k \times k$ Hermitian matrix $M$, and $\mathcal{F}_n(h, \bar{h})$ is a $k \times k$ matrix whose $(i, j)$-th entry is*

$$\sqrt{\theta_i \theta_j} \int_{[0,1]^2} \mathrm{d}x \, \mathrm{d}y \, \bar{\nu}_i(x) (h - \bar{h})^n(x, y) \bar{\nu}_j(y) \tag{A.3}$$

*for $1 \leq i, j \leq k$ and $n \in \mathbb{N}_0$.*

*Proof.* Put $\mu = \|T_h\|$, and let $u$ be the eigenfunction of $h$ corresponding to $\mu$, i.e.,

$$T_h u = \mu u. \tag{A.4}$$

Put $g = h - \bar{h}$ and rewrite the above as

$$(\mu - T_g)u = T_{\bar{h}}u. \tag{A.5}$$

The assumption $\|T_{h-\bar{h}}\| < \|T_h\|$ implies that $\mu - T_g$ is invertible, which allows us to write

$$u = (\mu - T_g)^{-1}T_{\bar{h}}u = \sum_{j=1}^{k} \theta_j \langle \bar{\nu}_j, u \rangle (\mu - T_g)^{-1}\bar{\nu}_j. \tag{A.6}$$

For fixed $1 \leq i \leq k$, it follows that

$$\langle \bar{\nu}_i, u \rangle = \sum_{j=1}^{k} \theta_j \langle \bar{\nu}_j, u \rangle \langle \bar{\nu}_i, (\mu - T_g)^{-1}\bar{\nu}_j \rangle. \tag{A.7}$$

Multiplying both sides by $\mu\sqrt{\theta_i}$, we get

$$Mv = \mu v, \tag{A.8}$$

where $M = (M_{ij})_{1 \leq i,j \leq k}$ is the $k \times k$ real symmetric matrix with elements

$$M_{ij} = \sqrt{\theta_i \theta_j} \left\langle \bar{\nu}_i, \left(1 - \frac{T_g}{\mu}\right)^{-1} \bar{\nu}_j \right\rangle, \qquad 1 \leq i, j \leq k, \tag{A.9}$$

and

$$v = \left[\sqrt{\theta_1}\langle \bar{\nu}_1, u \rangle, \ldots, \sqrt{\theta_k}\langle \bar{\nu}_k, u \rangle \right]'. \tag{A.10}$$

The first entry of $v$ is non-zero for $\varepsilon$ small with $\|T_g\| < \varepsilon$. Thus, (A.8) means that $\mu$ is an eigenvalue of $M$. By studying the diagonal entries of $M$, we can shown with the help of the Gershgorin circle theorem that, for small $\|T_g\|$,

$$\mu = \lambda_k(M). \tag{A.11}$$

With the help of the observation

$$M_{ij} = \sqrt{\theta_i \theta_j} \sum_{n \in \mathbb{N}_0} \mu^{-n} \langle \bar{\nu}_i, g^n \bar{\nu}_j \rangle, \qquad 1 \leq i, j \leq k, \tag{A.12}$$

i.e.,

$$M = \sum_{n \in \mathbb{N}_0} \mu^{-n} \mathcal{F}_n(h, \bar{h}), \tag{A.13}$$

this completes the proof. $\qquad\square$

## ACKNOWLEDGMENT

## References

[1] F. Augeri. Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails. *Electron. J. Probab.*, 21:Paper No. 32, 49, 2016. doi: 10.1214/16-EJP4146. URL https://doi.org/10.1214/16-EJP4146.

[2] F. Augeri, A. Guionnet, and J. Husson. Large deviations for the largest eigenvalue of sub-Gaussian matrices. *arXiv preprint arXiv:1911.10591*, 2019.

[3] M. Bauer and O. Golinelli. Random incidence matrices: moments of the spectral density. *J. Statist. Phys.*, 103(1-2):301–337, 2001. ISSN 0022-4715. doi: 10.1023/A:1004879905284. URL https://doi.org/10.1023/A:1004879905284.

[4] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields*, 108(4):517–542, 1997. ISSN 0178-8051. doi: 10.1007/s004400050119. URL https://doi.org/10.1007/s004400050119.

[5] G. Ben Arous, A. Dembo, and A. Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1):1–67, 2001. ISSN 0178-8051. doi: 10.1007/PL00008774. URL https://doi.org/10.1007/PL00008774.

[6] F. Benaych-Georges, C. Bordenave, and A. Knowles. Largest eigenvalues of sparse inhomogeneous Erdős-Rényi random graphs. *Ann. Probab.*, 47(3):1653–1676, 2019. ISSN 0091-1798. doi: 10.1214/18-AOP1293. URL https://doi.org/10.1214/18-AOP1293.

[7] S. Bhamidi, S. N. Evans, and A. Sen. Spectra of large random trees. *J. Theoret. Probab.*, 25(3):613–654, 2012. ISSN 0894-9840. doi: 10.1007/s10959-011-0360-9. URL https://doi.org/10.1007/s10959-011-0360-9.

[8] C. Bordenave and P. Caputo. A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.*, 42(6):2454–2496, 2014. ISSN 0091-1798. doi: 10.1214/13-AOP866. URL https://doi.org/10.1214/13-AOP866.

[9] C. Bordenave and M. Lelarge. Resolvent of large random graphs. *Random Structures Algorithms*, 37(3):332–352, 2010. ISSN 1042-9832. doi: 10.1002/rsa.20313. URL https://doi.org/10.1002/rsa.20313.

[10] A. Chakrabarty, R. S. Hazra, F. den Hollander, and M. Sfragara. Spectra of adjacency and Laplacian matrices of inhomogeneous Erdős-Rényi random graphs. *Random Matrices: Theory and Applications (to appear)*, page 2150009. URL https://doi.org/10.1142/S201032632150009X.

[11] A. Chakrabarty, S. Chakraborty, and R. S. Hazra. Eigenvalues outside the bulk of inhomogeneous Erdős-Rényi random graphs. *arXiv preprint arXiv:1911.08244*, 2019.

[12] S. Chatterjee. *Large Deviations for Random Graphs*, volume 2197 of *Lecture Notes in Mathematics*. Springer, Cham, 2017. ISBN 978-3-319-65815-5; 978-3-319-65816-2. doi: 10.1007/978-3-319-65816-2. URL https://doi.org/10.1007/978-3-319-65816-2. Lecture notes from the 45th Probability Summer School held in Saint-Flour, June 2015, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

[13] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011. ISSN 0195-6698. doi: 10.1016/j.ejc.2011.03.014. URL https://doi.org/10.1016/j.ejc.2011.03.014.

[14] K. Deimling. *Nonlinear Functional Analysis*. Springer-Verlag, Berlin, 1985. ISBN 3-540-13928-1. doi: 10.1007/978-3-662-00547-7. URL https://doi.org/10.1007/978-3-662-00547-7.

[15] A. Dembo and E. Lubetzky. Empirical spectral distributions of sparse random graphs. *arXiv preprint arXiv:1610.05186*, 2016.

[16] F. den Hollander. *Large Deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000. ISBN 0-8218-1989-5.

[17] S. Dhara and S. Sen. Large deviation for uniform graphs with given degrees. *arXiv preprint arXiv:1904.07666*, 2019.

[18] X. Ding and T. Jiang. Spectral distributions of adjacency and Laplacian matrices of random graphs. *Ann. Appl. Probab.*, 20(6):2086–2117, 2010. ISSN 1050-5164. doi: 10.1214/10-AAP677. URL https://doi.org/10.1214/10-AAP677.

[19] M. Disertori, F. Merkl, and S. W. W. Rolles. Localization for a nonlinear sigma model in a strip related to vertex reinforced jump processes. *Comm. Math. Phys.*, 332(2):783–825, 2014. ISSN 0010-3616. doi: 10.1007/s00220-014-2102-1. URL https://doi.org/10.1007/s00220-014-2102-1.

[20] I. Dumitriu and S. Pal. Sparse regular random graphs: spectral density and eigenvectors. *Ann. Probab.*, 40(5):2197–2235, 2012. ISSN 0091-1798. doi: 10.1214/11-AOP673. URL https://doi.org/10.1214/11-AOP673.

[21] I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek. Spectra of ?real-world? graphs: Beyond the semicircle law. *Physical Review E*, 64(2):026704, 2001.

[22] T. Jiang. Empirical distributions of Laplacian matrices of large dilute random graphs. *Random Matrices Theory Appl.*, 1(3):1250004, 20, 2012. ISSN 2010-3263. doi: 10.1142/S2010326312500049. URL https://doi.org/10.1142/S2010326312500049.

[23] T. Jiang. Low eigenvalues of Laplacian matrices of large random graphs. *Probab. Theory Related Fields*, 153(3-4):671–690, 2012. ISSN 0178-8051. doi: 10.1007/s00440-011-0357-4. URL https://doi.org/10.1007/s00440-011-0357-4.

[24] O. Khorunzhy, M. Shcherbina, and V. Vengerovsky. Eigenvalue distribution of large weighted random graphs. *J. Math. Phys.*, 45(4):1648–1672, 2004. ISSN 0022-2488. doi: 10.1063/1.1667610. URL https://doi.org/10.1063/1.1667610.

[25] J. O. Lee and K. Schnelli. Local law and Tracy-Widom limit for sparse random matrices. *Probab. Theory Related Fields*, 171(1-2):543–616, 2018. ISSN 0178-8051. doi: 10.1007/s00440-017-0787-8.

[26] L. Lovász. *Large Networks and Graph Limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012. ISBN 978-0-8218-9085-1. doi: 10.1090/coll/060. URL https://doi.org/10.1090/coll/060.

[27] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. *Random Structures Algorithms*, 47(1):109–146, 2015. ISSN 1042-9832. doi: 10.1002/rsa.20536. URL https://doi.org/10.1002/rsa.20536.

[28] M. J. R. Markering. The large deviation principle for inhomogeneous Erdős-Rényi random graphs. *Bachelor thesis Leiden University*, May 2020.

[29] F. Sauvigny. *Partial Differential Equations. 2*. Universitext. Springer-Verlag London, Ltd., London, 2012. ISBN 978-1-4471-2983-7. doi: 10.1007/978-1-4471-2984-4. URL https://doi.org/10.1007/978-1-4471-2984-4. Functional analytic methods, With consideration of lectures by E. Heinz, Second revised and enlarged edition of the 2006 translation.

[30] L. V. Tran, V. H. Vu, and K. Wang. Sparse random graphs: eigenvalues and eigenvectors. *Random Structures Algorithms*, 42(1):110–134, 2013. ISSN 1042-9832. doi: 10.1002/rsa.20406. URL https://doi.org/10.1002/rsa.20406.

[31] Y. Zhu. A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures Algorithms*, 56(1):251–279, 2020. ISSN 1042-9832. doi: 10.1002/rsa.20894. URL https://doi.org/10.1002/rsa.20894.

Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India
*Email address*: arijit.isi@gmail.com
*Email address*: rajatmaths@gmail.com

Mathematisch instituut, Universiteit Leiden, The Netherlands
*Email address*: denholla@math.leidenuniv.nl
*Email address*: m.sfragara@math.leidenuniv.nl