Wireless random-access networks and spectra of random graphs

Sfragara, M.

Author: Sfragara, M.
Title: Wireless random-access networks and spectra of random-graphs
Issue Date: 2020-10-28

CHAPTER 6

Largest eigenvalue of the adjacency matrix

This chapter is based on:
A. Chakrabarty, R.S. Hazra, F. den Hollander, M. Sfragara. *Large deviation principle for the maximal eigenvalue of inhomogeneous Erdős-Rényi random graphs.* [arXiv:2008.08367], 2020.

Abstract

We consider inhomogeneous Erdős-Rényi random graphs G_N on N vertices in the dense regime. The edge between the pair of vertices $\{i,j\}$ is retained with probability $r(\frac{i}{N},\frac{j}{N})$, $1 \leq i \neq j \leq N$, independently of other edges, where $r\colon [0,1]^2 \to (0,1)$ is a symmetric function that plays the role of a reference graphon. Let λ_N be the largest eigenvalue of the adjacency matrix of G_N. It is known that λ_N/N satisfies a large deviation principle as $N \to \infty$. The associated rate function ψ_r is given by a variational formula that involves the rate function I_r of a large deviation principle on graphon space. We analyze this variational formula in order to identify the properties of ψ_r, specially when the reference graphon is of rank 1.

§6.1 Introduction and main results

In Section 6.1.1 we define the mathematical model and we state the large deviation principle (LDP) for inhomogeneous Erdős Rényi random graphs. In Section 6.1.2 we present some facts about graphon operators. In Section 6.1.3 we state the LDP for the largest eigenvalue of the adjacency matrix, together with some properties of the rate function. Moreover, under the assumption that the connection probabilities have a multiplicative structure, we identify the scaling behavior of the rate function around its minimum and its end points. In Section 6.1.4 we briefly discuss these results and give an outline of the remainder of the chapter.

§6.1.1 Setting

We refer to Section 1.2.3 for a general introduction to spectra of Erdős-Rényi random graphs. We focus on *inhomogeneous Erdős-Rényi random graphs* and consider the dense regime, where the degrees of the vertices diverge linearly with the size of the graph.

Recall Section 1.2.5 for an introduction to graphon theory. Let $r \in \mathcal{W}$ be a *reference graphon* satisfying

$$\exists\, \eta > 0\colon \qquad \eta \le r(x,y) \le 1-\eta \qquad \forall\,(x,y) \in [0,1]^2. \tag{6.1}$$

Fix $N \in \mathbb{N}$ and consider the random graph G_N with vertex set $[N] = \{1,\dots,N\}$ 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

$$h^{G_N}(x,y) = \begin{cases} 1, & \text{if there is an edge between vertex } \lceil Nx \rceil \text{ and vertex } \lceil Ny \rceil, \\ 0, & \text{otherwise.} \end{cases} \tag{6.2}$$

Recall the equivalence relation $\sim$ on $\mathcal{W}$ defined in Section 1.2.5 and write $\widetilde{\mathbb{P}}_N$ to denote the law of $\widetilde{h}^{G_N}$.

The following LDP is proved in [145] and is an extension of the celebrated LDP for homogeneous Erdős-Rényi random graphs derived in [138]. Further properties of the rate function were derived in [179].

Theorem 6.1.1 (LDP for inhomogeneous Erdős-Rényi random graphs). *Subject to* (6.1)*, the sequence* $(\widetilde{\mathbb{P}}_N)_{N\in\mathbb{N}}$ *satisfies the LDP on* $(\widetilde{\mathcal{W}}, \delta_\square)$ *with rate* $\binom{N}{2}$*, i.e.,*

$$\begin{aligned} \limsup_{N\to\infty} \frac{1}{\binom{N}{2}} \log \widetilde{\mathbb{P}}_N(\mathcal{C}) &\le -\inf_{\widetilde{h}\in\mathcal{C}} J_r(\widetilde{h}) \qquad \forall\, \mathcal{C} \subset \widetilde{\mathcal{W}} \textit{ closed}, \\ \liminf_{N\to\infty} \frac{1}{\binom{N}{2}} \log \widetilde{\mathbb{P}}_N(\mathcal{O}) &\ge -\inf_{\widetilde{h}\in\mathcal{O}} J_r(\widetilde{h}) \qquad \forall\, \mathcal{O} \subset \widetilde{\mathcal{W}} \textit{ open}, \end{aligned} \tag{6.3}$$

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

$$J_r(\widetilde{h}) = \inf_{\phi\in\mathcal{M}} I_r(h^\phi), \tag{6.4}$$

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

$$I_r(h) = \int_{[0,1]^2} \mathcal{R}\big(h(x,y) \mid r(x,y)\big)\, dx\, dy, \qquad h \in \mathcal{W}, \tag{6.5}$$

with

$$\mathcal{R}\big(a \mid b\big) = a \log \tfrac{a}{b} + (1-a) \log \tfrac{1-a}{1-b} \tag{6.6}$$

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 (6.4) differs from the expression in [145], where the rate function is the lower semi-continuous envelope of $I_r(h)$. However, it was shown in [180] 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 (6.1).

§6.1.2 Graphon operators

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

$$(T_h u)(x) = \int_{[0,1]} h(x,y) u(y)\, dy, \qquad x \in [0,1], \tag{6.7}$$

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

$$\|T_h\| = \sup_{\substack{u \in L^2([0,1]) \\ \|u\|_2 = 1}} \|T_h u\|_2, \tag{6.8}$$

where $\|\cdot\|_2$ denotes the L^2-norm. Given a graphon $h \in \mathcal{W}$, we have that $\|T_h\| \leq \|h\|_2$. 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

$$(h_1 h_2)(x,y) = \int_{[0,1]} h_1(x,z) h_2(z,y)\, dz, \qquad (x,y) \in [0,1]^2, \tag{6.9}$$

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

$$h^n(x,y) = \int_{[0,1]^{n-1}} h(x,z_1) \cdots h(z_{n-1},y)\, dz_1 \cdots dz_{n-1}, \qquad (x,y) \in [0,1]^2,\ n \in \mathbb{N}. \tag{6.10}$$

Definition 6.1.2 (Eigenvalues and eigenfunctions).
The number $\mu \in \mathbb{R}$ 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

$$(T_h u)(x) = \mu u(x), \qquad x \in [0,1]. \tag{6.11}$$

The function u is said to be an *eigenfunction* associated with μ.

Proposition 6.1.3 (Properties of the graphon operator).
For any $h \in \mathcal{W}$ the following statements hold.

(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 $\mu_1 \geq \mu_2 \geq \ldots \geq 0$. Moreover, there exists a collection of eigenfunctions which form an orthonormal basis of $L^2([0,1])$.

(iii) The largest eigenvalue μ_1 of the graphon operator T_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 largest eigenvalue equals the operator norm.

Proof. The claim is a special case of [184, Theorem 7.3] (when the compact Hermitian operators considered there are taken to be the graphon operators). See also [143, Theorem 19.2] and [147, Appendix A]. □

§6.1.3 Main theorems

Let λ_N be the largest eigenvalue of the adjacency matrix A_N of G_N. Write $\mathbb{P}_N^*$ to denote the law of λ_N/N.

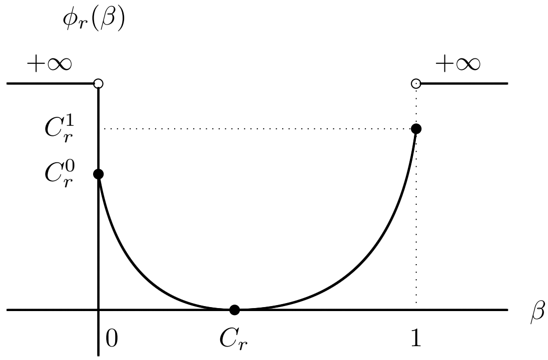

Figure 6.1: Graph of $\beta \mapsto \psi_r(\beta)$.

Theorem 6.1.4 (LDP for the largest eigenvalue).
Subject to (6.1)*, the sequence $(\mathbb{P}_N^*)_{N\in\mathbb{N}}$ satisfies the LDP on $\mathbb{R}$ with rate $\binom{N}{2}$ and with rate function*

$$\psi_r(\beta) = \inf_{\substack{\widetilde{h}\in\widetilde{\mathcal{W}} \\ \|T_{\widetilde{h}}\|=\beta}} J_r(\widetilde{h}) = \inf_{\substack{h\in\mathcal{W} \\ \|T_h\|=\beta}} I_r(h), \qquad \beta \in \mathbb{R}. \tag{6.12}$$

Proof. Note that $\lambda_N/N = \|T_{h^{G_N}}\|$, where h is any representative of $\widetilde{h}$ (we use the fact that $\|T_{\widetilde{h}}\| = \|T_{h^\phi}\|$ for all $\phi \in \mathcal{M}$). Also note that $\widetilde{h} \mapsto \|T_{\widetilde{h}}\|$ is a bounded and continuous function on $\widetilde{\mathcal{W}}$ [137, Exercises 6.1–6.2, Lemma 6.2]. Hence the claim follows from Theorem 6.1.1 via the contraction principle (see [167, Chapter 3]). □

Put

$$C_r = \|T_r\|. \tag{6.13}$$

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 Figure 6.1).

Theorem 6.1.5 (Properties of the rate function).
Subject to (6.1)*, the rate function in* (6.12) *satisfies the following.*

(i) The rate function ψ_r is continuous and unimodal on $[0,1]$*, with a unique zero at C_r.*

(ii) The rate function ψ_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 $\psi_r(\beta)$ is non-empty and compact in $\widetilde{\mathcal{W}}$.

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

$$r(x,y) = \nu(x)\,\nu(y), \qquad (x,y) \in [0,1]^2, \tag{6.14}$$

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

$$m_k = \int_{[0,1]} \nu(x)^k\,dx, \qquad k \in \mathbb{N}. \tag{6.15}$$

Note that $C_r = m_2$. Abbreviate

$$B_r = \int_{[0,1]^2} r(x,y)^3(1 - r(x,y))\,dx\,dy, \tag{6.16}$$

and note that $B_r = m_3^2 - m_4^2$. Further abbreviate

$$N_r^1 = \int_{[0,1]^2} \frac{1 - r(x,y)}{r(x,y)}\,dx\,dy, \qquad N_r^0 = \int_{[0,1]^2} \frac{r(x,y)}{1 - r(x,y)}\,dx\,dy. \tag{6.17}$$

Recall from Section 1.2.5 that $\mathcal{M}$ is the set of Lebesgue measure-preserving bijective maps $\phi\colon\ [0,1] \to [0,1]$.

Theorem 6.1.6 (Scaling of the rate function).
Let ψ_r be the rate function in (6.12).

(i) Subject to (6.1) *and* (6.14),

$$\psi_r(\beta) = K_r\,(\beta - C_r)^2\,[1 + o(1)], \qquad \beta \to C_r, \tag{6.18}$$

with

$$K_r = \frac{C_r^2}{2B_r} = \frac{m_2^2}{2(m_3^2 - m_4^2)}. \tag{6.19}$$

(ii) Subject to (6.1),

$$C_r^1 - \psi_r(\beta) = (1-\beta)\bigg(\log\frac{N_r^1}{1-\beta} + 1 + o(1)\bigg), \qquad \beta \to 1. \tag{6.20}$$

(iii) Subject to (6.1),

$$C_r^0 - \psi_r(\beta) = \beta\bigg(\log\frac{N_r^0}{\beta} + 1 + o(1)\bigg), \qquad \beta \to 0. \tag{6.21}$$

Theorem 6.1.7 (Scaling of the minimisers).
Let $h_\beta \in \mathcal{W}$ be any minimiser of the second infimum in (6.12).

(i) Subject to (6.1) *and* (6.14),

$$\lim_{\beta \to C_r} (\beta - C_r)^{-1} \| h_\beta - r - (\beta - C_r)\Delta\|_2 = 0, \tag{6.22}$$

with

$$\Delta(x,y) = \frac{C_r}{B_r}\, r(x,y)^2(1-r(x,y)), \qquad (x,y) \in [0,1]^2. \tag{6.23}$$

(ii) Subject to (6.1),

$$\lim_{\beta \to 1} (1-\beta)^{-1} \| 1 - h_\beta - (1-\beta)\Delta\|_2 = 0, \tag{6.24}$$

with

$$\Delta(x,y) = \frac{1}{N_r^1}\,\frac{1-r(x,y)}{r(x,y)}, \qquad (x,y) \in [0,1]^2. \tag{6.25}$$

(iii) Subject to (6.1),

$$\lim_{\beta \to 0} \beta^{-1} \| h_\beta - \beta\Delta\|_2 = 0, \tag{6.26}$$

with

$$\Delta(x,y) = \frac{1}{N_r^0}\,\frac{r(x,y)}{1-r(x,y)}, \qquad (x,y) \in [0,1]^2. \tag{6.27}$$

§6.1.4 Discussion and outline

Theorems. Theorem 6.1.5 confirms the picture of ψ_r drawn in Figure 6.1. It remains open whether or not ψ_r is convex. We do not expect ψ_r to be analytic, because bifurcations may occur in the set of minimisers of ψ_r as β is varied. Theorem 6.1.6 identifies the scaling of ψ_r around its minimum and near its end points, provided r is of rank 1. The inverse curvature $1/K_r$ equals the variance in the central limit theorem derived in [130]. This is in line with the standard folkore of large deviation theory. Theorem 6.1.7 identifies the corresponding scaling of the minimiser h_β of ψ_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 Appendix F).

Conditions. It would be interesting to investigate to what extent the condition on the reference graphon in (6.1) can be weakened to some form of integrability condition.

Especially for the upper bound in the LDP this is delicate, because the proof in [145] is based on block-graphon approximation (see [180]).

Outline of the chapter. The remainder of this chapter is organized as follows. In Section 6.2 we derive an expansion for the operator norm of a graphon around any graphon of rank 1. In Section 6.3 we prove our main theorems. In Appendix F we show how the expansion around reference graphons can be extended to finite rank.

§6.2 Expansion around rank-one graphons

In this section we show how we can expand the operator norm of a graphon around any graphon of rank 1. This prepares for the perturbation analysis in Sections 6.3.2–6.3.4.

Lemma 6.2.1 (Rank-one expansion).
Consider a graphon $\bar{h} \in \mathcal{W}$ of rank 1 such that $\bar{h}(x,y) = \bar{\nu}(x)\bar{\nu}(y)$, $(x,y) \in [0,1]^2$. 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

$$\mu = \sum_{n\in\mathbb{N}_0} \frac{1}{\mu^n}\mathcal{F}_n(h,\bar{h}), \tag{6.28}$$

where

$$\mathcal{F}_n(h,\bar{h}) = \int_{[0,1]^2} \bar{\nu}(x)(h-\bar{h})^n(x,y)\bar{\nu}(y)\,dx\,dy. \tag{6.29}$$

Proof. By Proposition 6.1.3, we have

$$T_h u = \mu u, \tag{6.30}$$

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

$$u = (\mu - T_g)^{-1}\bar{\nu}\langle\bar{\nu}, u\rangle, \tag{6.31}$$

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

$$\langle\bar{\nu},u\rangle = \langle\bar{\nu},u\rangle\langle\bar{\nu},(\mu - T_g)^{-1}\bar{\nu}\rangle, \tag{6.32}$$

which gives

$$\mu = \langle\bar{\nu},(1 - T_g/\mu)^{-1}\bar{\nu}\rangle. \tag{6.33}$$

We can expand the above to get

$$\begin{aligned}
\mu &= \left\langle \bar{\nu}, \sum_{n\in\mathbb{N}_0}\left(\frac{T_g}{\mu}\right)^n \bar{\nu}\right\rangle \\
&= \sum_{n\in\mathbb{N}_0}\frac{1}{\mu^n}\int_{[0,1]^{n+1}} \bar{\nu}(x_0)g(x_0,x_1)\cdots g(x_{n-1},x_n)\bar{\nu}(x_n)\,dx_0\;dx_1\cdots dx_n \\
&= \sum_{n\in\mathbb{N}_0}\frac{1}{\mu^n}\mathcal{F}_n(h,\bar{h}),
\end{aligned} \tag{6.34}$$

and this completes the proof. □

CHAPTER 6

Subject to (6.14), it follows from Lemma 6.2.1 with $h = \bar{h} = r$ that

$$C_r = \|T_r\| = m_2, \tag{6.35}$$

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

Remark 6.2.2 (Higher rank).
The expansion around reference graphons of rank 1 can be extended to finite rank. We provide the details in Appendix F. In this chapter we focus on rank 1, for which Lemma 6.2.1 allows us to analyse the behavior of $\psi_r(\beta)$ near the values $\beta = C_r$, $\beta = 1$ and $\beta = 0$. Note that both the graphons $h = r$ and $h \equiv 1$ are of rank 1.

§6.3 Proofs of the main results

In this section we prove the theorems in Section 6.1.3. In Section 6.3.1 we prove Theorem 6.1.5. In the last three sections we prove Theorems 6.1.6–6.1.7 by analyzing graphon perturbations around the minimum of the rate function and near its end points. The proofs of the theorems rely on the variational formula in (6.12). Since the largest eigenvalue is invariant under relabeling of the vertices, we can work directly with I_r in (6.5) without worrying about the equivalence classes.

§6.3.1 Proof: properties of the rate function

Proof of Theorem 6.1.5. We follow [137, Chapter 6]. Even though this monograph deals with constant reference graphons only, most arguments carry over to r satisfying (6.1). Define

$$\psi_r^+(\beta) = \inf_{\substack{h\in\mathcal{W}\\ \|T_h\|\geq\beta}} I_r(h), \quad \psi_r^-(\beta) = \inf_{\substack{h\in\mathcal{W}\\ \|T_h\|\leq\beta}} I_r(h), \qquad \beta\in\mathbb{R}. \tag{6.36}$$

(i) Because $h \mapsto \|T_h\|$ is a nice graph parameter, in the sense of [137, 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 (see [137, 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 continuity and unimodality of ψ_r follow from the proof of (iii). Moreover, since $I_r(h) = 0$ if and only if $h = r$ almost everywhere, it is immediate that C_r is the unique zero of ψ_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 $\beta_1, \beta_2 \in [C_r, 1]$, $\beta_1 < \beta_2$, such that ψ_r^+ is constant on $[\beta_1, \beta_2]$. Consequently, there exist minimisers $h_{\beta_1}^{\phi_1}, h_{\beta_2}^{\phi_2}$, $\phi_1, \phi_2 \in \mathcal{M}$, satisfying $r \leq h_{\beta_1}^{\phi_1} \leq h_{\beta_2}^{\phi_2}$, such that $I_r(h_{\beta_1}^{\phi_1}) = I_r(h_{\beta_2}^{\phi_2})$ and $\|T_{h_{\beta_1}^{\phi_1}}\| = \beta_1 < \beta_2 = \|T_{h_{\beta_2}^{\phi_2}}\|$. However, since $a \mapsto \mathcal{R}(a \mid b)$ is strictly increasing on $[b, 1]$ (recall (6.5)), 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 $[0, C_r]$.

(iii) The variational formulas in (6.36) achieve minimisers. In fact, the sets of minimiser are non-empty compact subsets of $\widetilde{\mathcal{W}}$ (see [137, Theorem 6.2]). In addition, all minimisers h of $\phi_r^+(h)$ satisfy $h \geq r$ almost everywhere, while all minimisers h of ϕ_r^- satisfy $h \leq r$ almost everywhere (see [137, Lemma 6.3]). Moreover, because

$$\begin{aligned} h_1 \geq h_2 \geq r \quad &\Longrightarrow \quad \|T_{h_1}\| \geq \|T_{h_2}\|,\ I_r(h_1) \geq I_r(h_2), \\ h_1 \leq h_2 \leq r \quad &\Longrightarrow \quad \|T_{h_1}\| \leq \|T_{h_2}\|,\ I_r(h_1) \leq I_r(h_2), \end{aligned} \tag{6.37}$$

(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 β, and so

$$\psi_r(\beta) = \begin{cases} \psi_r^+(\beta), & \beta \geq C_r, \\ \psi_r^-(\beta), & \beta \leq C_r. \end{cases} \tag{6.38}$$

□

§6.3.2 Proof: perturbation around the minimum

Note that when $\beta = C_r$, the infimum in (6.12) 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 $r + \Delta_\epsilon \in \mathcal{W}$, hence we are dealing with a perturbation Δ_ϵ which is symmetric and bounded. We compare

$$\psi_r(C_r + \epsilon) = \inf_{\substack{\Delta_\epsilon\colon\ [0,1]^2 \to \mathbb{R} \\ r+\Delta_\epsilon \in \mathcal{W} \\ \|T_{r+\Delta_\epsilon}\| = C_r + \epsilon}} I_r(r + \Delta_\epsilon) \tag{6.39}$$

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

$$\delta_r(\epsilon) = \psi_r(C_r + \epsilon) - \psi_r(C_r) = \psi_r(C_r + \epsilon) \tag{6.40}$$

and studying its behavior as $\epsilon \to 0$. Since $r(x,y) = \nu(x)\nu(y)$, $(x,y) \in [0,1]^2$, we can use Lemma 6.2.1 to control the norm of $T_h = T_{r+\Delta_\epsilon}$. Pick $\bar{h} = r$ and $h = r + \Delta_\epsilon$ in (6.28) such that $\|\Delta_\epsilon\|_2 \to 0$ as $\epsilon \to 0$. Note that $\|T_{\Delta_\epsilon}\| \leq \|\Delta_\epsilon\|_2 < C_r$ for ϵ small enough. Hence, writing out the expansion for the norm, we get

$$\|T_{r+\Delta_\epsilon}\| = C_r + \sum_{n\in\mathbb{N}} \frac{1}{\|T_{r+\Delta_\epsilon}\|^n}\, \mathcal{F}_n(r + \Delta_\epsilon, r). \tag{6.41}$$

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

$$C_r + \epsilon = C_r + \frac{\langle \nu, \Delta_\epsilon \nu\rangle}{C_r + \epsilon} + \sum_{n\in\mathbb{N}\setminus\{1\}} \frac{1}{(C_r + \epsilon)^n} \langle \nu, \Delta_\epsilon^n \nu\rangle \tag{6.42}$$

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

$$\epsilon(C_r + \epsilon) = \int_{[0,1]^2} r\Delta_\epsilon + \sum_{n\in\mathbb{N}\setminus\{1\}} \frac{1}{(C_r + \epsilon)^{n-1}} \langle \nu, \Delta_\epsilon^n \nu\rangle. \tag{6.43}$$

Chapter 6

Since ν is bounded, using the generalized Hölder's inequality (see [179, Theorem 3.1]) we get

$$|\langle \nu, \Delta_\epsilon^n \nu\rangle| \le \|\Delta_\epsilon\|_2^n. \tag{6.44}$$

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

$$\sum_{n\in\mathbb{N}\setminus\{1\}} \frac{1}{(C_r+\epsilon)^{n-1}} \langle \nu, \Delta_\epsilon^n \nu\rangle = O\big(\|\Delta_\epsilon\|_2^2\big). \tag{6.45}$$

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

$$\int_{[0,1]^2} r\Delta_\epsilon = \epsilon C_r + \epsilon^2 + O\big(\|\Delta_\epsilon\|_2^2\big). \tag{6.46}$$

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

$$\int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]. \tag{6.47}$$

Small perturbation on a given region. In what follows we use the standard notation $o(\cdot)$, $O(\cdot)$, $\asymp$ to describe the asymptotic behavior in the limit as $\epsilon \to 0$. We first show that it is enough to consider Δ_ϵ of the form $\epsilon\Delta$ for some $\Delta \in L^2([0,1]^2)$, because these perturbations contribute to the minimum cost.

Lemma 6.3.1 (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

$$I_r(r+\Delta_\epsilon) \ge 2\epsilon^2. \tag{6.48}$$

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

$$I_r(r+\epsilon\Delta) = 2\epsilon^2 \int_{[0,1]^2} \frac{\Delta(x,y)^2}{4r(x,y)(1-r(x,y))}\,dx\,dy\,[1+o(1)], \qquad \epsilon \to 0. \tag{6.49}$$

Proof. Fix $b \in [0,1]$ and abbreviate (recall (6.6))

$$\chi(a) = \mathcal{R}(a \mid b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}, \qquad a \in [0,1]. \tag{6.50}$$

Note that

$$\chi(b) = \chi'(b) = 0, \quad \chi''(a) \ge 4, \qquad a \in [0,1]. \tag{6.51}$$

Consequently,

$$\chi(a) \ge 2(a-b)^2, \qquad a \in [0,1], \tag{6.52}$$

and hence

$$I_r(r+\Delta_\epsilon) \ge 2\int_{[0,1]^2} \Delta_\epsilon^2 = 2\|\Delta_\epsilon\|_2^2. \tag{6.53}$$

Next observe that

$$C_r+\epsilon = \|T_{r+\Delta_\epsilon}\| = \|T_r + T_{\Delta_\epsilon}\| \le \|T_r\| + \|T_{\Delta_\epsilon}\| \le C_r + \|\Delta_\epsilon\|_2, \tag{6.54}$$

which gives $\|\Delta_\epsilon\|_2 \geq \epsilon$. Inserting this lower bound into (6.53), we get (6.48). To get (6.49), we need a higher-order expansion of χ, namely,

$$\chi(x) = \tfrac{1}{2}\chi''(b)(x-b)^2 + O((x-b)^3), \qquad x \to b. \tag{6.55}$$

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 (6.49) follows. □

Next we consider different types of small perturbations in a given region and compute their total cost.

Lemma 6.3.2 (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

$$\int_B \mathcal{R}(h \mid r) = \epsilon^{2\alpha} \int_B \frac{\Delta^2}{2r(1-r)}[1+o(1)], \qquad \epsilon \to 0. \tag{6.56}$$

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

Proof. The proof is similar to that of Lemma 6.3.1. □

Approximation by block graphons. We next introduce block graphons, which will be useful for our perturbation analysis. It follows from Lemma 6.3.1 that optimal perturbations with Δ_ϵ 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 6.3.3 (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

$$f(x,y) = \begin{cases} f_{i,j}, & \text{if } (x,y) \in B_i \times B_j, \\ 0, & \text{otherwise}, \end{cases} \tag{6.57}$$

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

$$f_N(x,y) = N^2 \int_{B_{i,j}} f(x',y')\, dx'\, dy' = \bar{f}_{N,ij}, \qquad (x,y) \in B_{i,j}. \tag{6.58}$$

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

$$\|T_{f_N} - T_f\| = \|T_{f_N - f}\| \leq \|f_N - f\|_2. \tag{6.59}$$

We know from [137, Proposition 2.6] that, for any $f \in \mathcal{W}$ and its associated sequence of block graphons $(f_N)_{N\in\mathbb{N}}$, $\|f_N - f\|_2 \to 0$ and hence $\lim_{N\to\infty} \|T_{f_N}\| = \|T_f\|$. 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 6.3.4 (Convergence of the cost function).
For any $f \in \mathcal{W}$

$$\lim_{N\to\infty} I_{r_N}(f_N) = I_r(f). \tag{6.60}$$

Proof. Since $f \in L^2([0,1]^2)$, f_N is bounded. The assumption in (6.1) implies that $\eta \le r_N \le 1-\eta$ for all $N \in \mathbb{N}$. We know from [145, Lemma 2.3] that there exists a constant $c > 0$ independent of f such that

$$|I_{r_N}(f) - I_r(f)| \le c\,\|r_N - r\|_1 \le c\,\|r_N - r\|_2. \tag{6.61}$$

Hence

$$\begin{aligned} |I_{r_N}(f_N) - I_r(f)| &\le |I_{r_N}(f_N) - I_r(f_N)| + |I_r(f_N) - I_r(f)| \\ &\le c\,\|r_N - r\|_2 + |I_r(f_N) - I_r(f)|. \end{aligned} \tag{6.62}$$

Since $\lim_{N\to\infty}\|r_N - r\|_2 = 0$, the first term tends to zero. Since $\lim_{N\to\infty}\|f_N - f\|_2 = 0$ and I_r is continuous in the L^2-topology on $\mathcal{W}$ (see [180, Lemma 3.4]), also the second term tends to zero and the claim follows. □

Block graphon perturbations. In what follows we fix $N \in \mathbb{N}$, analyze 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

$$\overline{h}_{N,ij}(x,y) = \overline{r}_{N,ij}(x,y) + \overline{\Delta_\epsilon}_{N,ij}(x,y), \qquad (x,y) \in B_{i,j}, \tag{6.63}$$

with

$$\overline{r}_{N,ij} = N^2 \int_{B_{i,j}} r(x',y')\,dx'\,dy', \quad \overline{\Delta_\epsilon}_{N,ij} = N^2 \int_{B_{i,j}} \Delta_\epsilon(x',y')\,dx'\,dy'. \tag{6.64}$$

Observe that optimal perturbations must have $\|\Delta_\epsilon\|_2 = O(\epsilon)$, and hence the constraint in (6.46) becomes

$$\sum_{i,j=1}^{N} \int_{B_{i,j}} r(x,y)\Delta_\epsilon(x,y)\,dx\,dy = \sum_{i,j=1}^{N} \frac{1}{N^2}\,\overline{r\Delta_\epsilon}_{N,ij} = C_r\epsilon\,[1+o(1)], \qquad \epsilon \to 0. \tag{6.65}$$

The block constraint in (6.65) implies that the sum over each block must be of order ϵ. We therefore must have that

$$\overline{r\Delta_\epsilon}_{N,ij} = O(\epsilon), \qquad \epsilon \to 0, \qquad \forall (i,j), \tag{6.66}$$

which means that

$$\overline{\Delta_\epsilon}_{N,ij} = O(\epsilon), \qquad \epsilon \to 0, \qquad \forall (i,j), \tag{6.67}$$

since (6.1) implies that $\overline{r\Delta_\epsilon}_{N,ij} \asymp \overline{\Delta_\epsilon}_{N,ij}$. There are the following two possible cases.

(I) All blocks contribute to the constraint with a term of order ϵ (*balanced perturbation*).

CHAPTER 6

(II) Some blocks contribute to the constraint with a term of order ϵ and some with $o(\epsilon)$ (*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 6.3.2, this contributes a term of order ϵ^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 6.3.6 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 $o(\epsilon)$ to the constraint, i.e., $\overline{\Delta_{\epsilon N,ij}} = o(\epsilon)$. By Lemma 6.3.2, these blocks contribute order $o(\epsilon^2)$ to the total cost. The remaining blocks must fall in the class of blocks of type (I), with a perturbation of order ϵ on each of them. Corollary 6.3.6 below shows that the cost function attains its infimum when the small perturbation of order ϵ is uniform on $[0,1]^2$.

Optimal perturbation. We have shown that perturbations of type (I) lead to the minimal total cost. They consist of perturbations of order ϵ 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 Δ_ϵ 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 balanced perturbation.

Lemma 6.3.5 (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

$$\delta_r(\epsilon) = K_r\epsilon^2\,[1+o(1)], \qquad \epsilon \to 0, \tag{6.68}$$

with

$$K_r = \frac{1}{2}C_r^2 \inf_{\phi\in\mathcal{M}} \frac{D_r^\phi}{(B_r^\phi)^2}, \tag{6.69}$$

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 (6.46) becomes

$$\int_{[0,1]^2} r\Delta = C_r\,[1+o(1)], \qquad \epsilon \to 0, \tag{6.70}$$

and we get

$$\begin{aligned} \delta_r(\epsilon) &= \inf_{\substack{\Delta\colon\ [0,1]^2\to\mathbb{R} \\ r+\epsilon\Delta\in\mathcal{W} \\ \int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]}} I_r(r+\epsilon\Delta) \\ &= \inf_{\substack{\Delta\colon\ [0,1]^2\to\mathbb{R} \\ r+\epsilon\Delta\in\mathcal{W} \\ \int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]}} \int_{[0,1]^2} \mathcal{R}\big((r+\epsilon\Delta)(x,y) \mid r(x,y)\big)\,dx\,dy. \end{aligned} \tag{6.71}$$

By Lemma 6.3.2 (with $\alpha = 1$), we have

$$\delta_r(\epsilon) = K_r\epsilon^2\,[1+o(1)], \qquad \epsilon \to 0, \tag{6.72}$$

with

$$K_r = \inf_{\substack{\Delta\colon [0,1]^2 \to \mathbb{R} \\ r+\epsilon\Delta\in\mathcal{W} \\ \int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]}} \int_{[0,1]^2} \frac{\Delta(x,y)^2}{2r(x,y)(1-r(x,y))}\,dx\,dy. \tag{6.73}$$

The term $1+o(1)$ in (6.72) arises after we scale Δ by $1+o(1)$ in order to force $\int_{[0,1]^2} r\Delta = C_r$. Note that the optimization problem in (6.73) no longer depends on ϵ.

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

$$\mathcal{L}_{A_r}(\Delta) = \int_{[0,1]^2} \frac{\Delta^2}{2\,r(1-r)} + A_r \int_{[0,1]^2} r\Delta, \tag{6.74}$$

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

$$\Delta^\phi(x,y) = -A_r\, r^\phi(x,y) r(x,y)(1-r(x,y)), \qquad (x,y)\in[0,1]^2, \qquad \phi\in\mathcal{M}. \tag{6.75}$$

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

$$-A_r B_r^\phi = C_r\,[1+o(1)] \tag{6.76}$$

with

$$B_r^\phi = \int_{[0,1]^2} r(x,y)^\phi r(x,y)^2(1-r(x,y))\,dx\,dy. \tag{6.77}$$

We get

$$\Delta^\phi(x,y) = \frac{C_r}{B_r^\phi}\, r^\phi(x,y) r(x,y)(1-r(x,y)), \qquad (x,y)\in[0,1]^2, \qquad \phi\in\mathcal{M}, \tag{6.78}$$

and

$$K_r = \inf_{\phi\in\mathcal{M}} \int_{[0,1]^2} \frac{(\Delta^\phi)^2}{2r(1-r)} = \frac{1}{2} C_r^2 \inf_{\phi\in\mathcal{M}} \frac{D_r^\phi}{(B_r^\phi)^2} \tag{6.79}$$

with

$$D_r^\phi = \int_{[0,1]^2} (r^\phi)^2 r(1-r). \tag{6.80}$$

This completes the proof. □

We next show that the infimum in (6.79) is uniquely attained when ϕ is the identity. For this we show that $D_r^\phi/(B_r^\phi)^2 \geq 1/B_r$ with equality if and only if $\phi = \mathrm{Id}$.

Indeed, write

$$
\begin{aligned}
&B_r D_r^\phi - (B_r^\phi)^2 \\
&\quad = \int_{[0,1]^2} r(x,y)(1-r(x,y))\,dx\,dy \int_{[0,1]^2} r(\bar{x},\bar{y})(1-r(\bar{x},\bar{y}))\,d\bar{x}\,d\bar{y} \\
&\qquad \times \Big(r(x,y)^2 r^\phi(\bar{x},\bar{y})^2 - r(x,y) r^\phi(x,y) r(\bar{x},\bar{y}) r^\phi(\bar{x},\bar{y}) \Big) \\
&\quad = \int_{[0,1]^2} r(x,y)(1-r(x,y))\,dx\,dy \int_{[0,1]^2} r(\bar{x},\bar{y})(1-r(\bar{x},\bar{y}))\,d\bar{x}\,d\bar{y} \\
&\qquad \times \frac{1}{2}\Big(r(x,y)^2 r^\phi(\bar{x},\bar{y})^2 + r^\phi(x,y)^2 r(\bar{x},\bar{y})^2 - 2r(x,y) r^\phi(x,y) r(\bar{x},\bar{y}) r^\phi(\bar{x},\bar{y}) \Big) \\
&\quad = \int_{[0,1]^2} r(x,y)(1-r(x,y))\,dx\,dy \int_{[0,1]^2} r(\bar{x},\bar{y})(1-r(\bar{x},\bar{y}))\,d\bar{x}\,d\bar{y} \\
&\qquad \times \frac{1}{2}\Big(r(x,y) r^\phi(\bar{x},\bar{y}) - r^\phi(x,y) r(\bar{x},\bar{y}) \Big)^2,
\end{aligned}
\tag{6.81}
$$

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

We conclude that the infimum in (6.79) equals $1/B_r$, and so we find that

$$
K_r = \frac{C_r^2}{2B_r}. \tag{6.82}
$$

Finally, note that $C_r = m_2$ by (6.35), and that $B_r = m_3^2 - m_4^2$ by (6.15). This settles the expression for K_r in (6.19).

Corollary 6.3.6 (Unbalanced perturbations).
Perturbations of order ϵ 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 6.3.5.

Proof. The argument of the variational formula can be reduced to an integral that considers only those regions that contribute order ϵ^2, which constitute a subset of $[0,1]^2$. Applying the method of Lagrange multipliers as in Lemma 6.3.5, we obtain that the solution is given by

$$
\delta_r(\epsilon) = [1 + o(1)]\, K_r' \epsilon^2, \qquad \epsilon \to 0, \tag{6.83}
$$

with $K_r' > K_r$. The strict inequality comes from the fact that the optimal balanced perturbation Δ^{Id} found in (6.75) is non-zero everywhere. □

Proof of Theorems 6.1.6(i)–6.1.7(i). We have shown that a balanced perturbation is optimal and we have identified in (6.78) the form of the optimal balanced perturbation. The claim in Theorem 6.1.6(i) is settled by Lemma 6.3.5 and (6.82), while (6.78) settles the claim in Theorem 6.1.7(i). □

§6.3.3 Proof: perturbation near the right end

For $\beta = 1-\epsilon$ 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

$$\psi_r(1-\epsilon) = \inf_{\substack{\Delta_\epsilon\colon\ [0,1]^2\to[0,\infty)\\ 1-\Delta_\epsilon\in\mathcal{W}\\ \|T_{1-\Delta_\epsilon}\|=1-\epsilon}} I_r(1-\Delta_\epsilon) \tag{6.84}$$

with

$$C_r^1 = I_r(1) \tag{6.85}$$

by computing the difference

$$\delta_r(\epsilon) = \psi_r(1) - \psi_r(1-\epsilon) \tag{6.86}$$

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

$$\delta_r(\epsilon) = \sup_{\substack{\Delta_\epsilon\colon\ [0,1]^2\to[0,\infty)\\ 1-\Delta_\epsilon\in\mathcal{W}\\ \|T_{1-\Delta_\epsilon}\|=1-\epsilon}} \big(I_r(1) - I_r(1-\Delta_\epsilon)\big). \tag{6.87}$$

We again use the expansion in Lemma 6.2.1. Pick $\bar{h} = 1$ and $h = 1-\Delta_\epsilon$ in (6.28), to get

$$\|T_{1-\Delta_\epsilon}\| = 1 + \sum_{n\in\mathbb{N}} \frac{1}{\|T_{1-\Delta_\epsilon}\|^n}\, \mathcal{F}_n(1-\Delta_\epsilon, 1). \tag{6.88}$$

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

$$1-\epsilon = 1 + \frac{\langle 1, (-\Delta_\epsilon)1\rangle}{1-\epsilon} + \frac{\langle 1, (-\Delta_\epsilon)^2 1\rangle}{(1-\epsilon)^2} + \sum_{n\in\mathbb{N}\setminus\{1,2\}} \frac{\langle 1, (-\Delta_\epsilon)^n 1\rangle}{(1-\epsilon)^n}. \tag{6.89}$$

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

$$\epsilon(1-\epsilon) = \int_{[0,1]^2} \Delta_\epsilon - \frac{\langle 1, \Delta_\epsilon^2 1\rangle}{(1-\epsilon)} + O\big(\|\Delta_\epsilon\|_2^3\big). \tag{6.90}$$

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

$$1-\epsilon = \|T_{1-\Delta_\epsilon}\| \le \|1-\Delta_\epsilon\|_2 \le \sqrt{\|1-\Delta_\epsilon\|_1}. \tag{6.91}$$

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

$$\|\Delta_\epsilon\|_1 \le 1-(1-\epsilon)^2 = \epsilon(2-\epsilon). \tag{6.92}$$

Since $\|\Delta_\epsilon\|_2^3 = O(\epsilon^{3/2})$, (6.90) reads

$$\frac{1}{1-\epsilon}\int_{[0,1]^3} \Delta_\epsilon(x,y)(1-\Delta_\epsilon(y,z))\,dx\,dy\,dz - \frac{\epsilon}{1-\epsilon}\,\|\Delta_\epsilon\|_1 = \epsilon(1-\epsilon) + O(\epsilon^{3/2}), \tag{6.93}$$

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

$$\int_{[0,1]^3} \Delta_\epsilon(x,y)(1-\Delta_\epsilon(y,z))\,dx\,dy\,dz = \epsilon\,[1+O(\epsilon^{1/2})]. \tag{6.94}$$

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

$$\int_{[0,1]^2} \Delta = 1 + O(\epsilon^{1/2}), \qquad \epsilon \to 0. \tag{6.95}$$

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

Lemma 6.3.7 (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 ϵ small enough,

$$I_r(1) - I_r(1-\Delta_\epsilon) \le \|\Delta_\epsilon\|_1 \log \frac{1}{\|\Delta_\epsilon\|_1} + O(\|\Delta_\epsilon\|_1). \tag{6.96}$$

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

Proof. Abbreviate (recall (6.6))

$$\chi(a) = \mathcal{R}(a \mid r) = a \log \frac{a}{r} + (1-a)\log\frac{1-a}{1-r}, \qquad a \in [0,1]. \tag{6.97}$$

Then

$$\chi(1) - \chi(1-\Delta_\epsilon(x,y)) = \Delta_\epsilon(x,y) \log\left(\frac{1-\Delta_\epsilon(x,y)}{\Delta_\epsilon(x,y)}\,\frac{1-r(x,y)}{r(x,y)}\right) - \log(1-\Delta_\epsilon(x,y)), \tag{6.98}$$

and so

$$I_r(1) - I_r(1-\Delta_\epsilon) = \int_{[0,1]^2} \left(\Delta_\epsilon \log\left(\frac{1-\Delta_\epsilon}{\Delta_\epsilon}\,\frac{1-r}{r}\right) - \log(1-\Delta_\epsilon)\right). \tag{6.99}$$

Let μ_ϵ 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

$$\begin{aligned}
\int_{[0,1]^2} \Delta_\epsilon \log\left(\frac{1-\Delta_\epsilon}{\Delta_\epsilon}\right) &= Z_\epsilon \int_{[0,1]^2} \mu_\epsilon \bar{s}\Big(\frac{\Delta_\epsilon}{1-\Delta_\epsilon}\Big) \\
&\le Z_\epsilon \bar{s}\big(Z_\epsilon^{-1}\|\Delta_\epsilon\|_1\big) \\
&= \|\Delta_\epsilon\|_1 \log\left(\frac{Z_\epsilon}{\|\Delta_\epsilon\|_1}\right).
\end{aligned} \tag{6.100}$$

Moreover,

$$\int_{[0,1]^2} \Delta_\epsilon \log\left(\frac{1-r}{r}\right) = O(\|\Delta_\epsilon\|_1), \quad -\int_{[0,1]^2} \log(1-\Delta_\epsilon) = O(\|\Delta_\epsilon\|_1). \tag{6.101}$$

Hence

$$I_r(1) - I_r(1-\Delta_\epsilon) \le \|\Delta_\epsilon\|_1 \log \frac{1}{\|\Delta_\epsilon\|_1} + O(\|\Delta_\epsilon\|_1), \qquad \epsilon \to 0, \tag{6.102}$$

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

The following is the analogue of Lemma 6.3.2 for perturbations near the right end.

Lemma 6.3.8 (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 $\epsilon > 0$, $\alpha > 0$ and $\Delta\colon [0,1]^2 \to [0,\infty)$. Then the contribution of B to the cost $I_r(h)$ is

$$\int_B \mathcal{R}(1 \mid r) - \mathcal{R}(h \mid r) = \int_B \epsilon^\alpha \Delta \log\left(\frac{1-r}{\epsilon^\alpha \Delta r}\right)[1+o(1)], \qquad \epsilon \to 0. \tag{6.103}$$

Proof. Observe that

$$\mathcal{R}(1 \mid r) - \mathcal{R}(1-\epsilon^\alpha\Delta \mid r) = \epsilon^\alpha \Delta \log\left(\frac{1-r}{\epsilon^\alpha \Delta r}\right)[1+o(1)], \qquad \epsilon \to 0. \tag{6.104}$$

The proof is analogous to that of Lemma 6.3.2. □

Following the argument in Section 6.3.2, we can approximate the cost function by using block graphons. The constraint becomes

$$\sum_{i,j=1}^{N} \int_{B_{i,j}} \Delta_{\epsilon,N}(x,y)\,dx\,dy = \sum_{i,j=1}^{N} \frac{1}{N^2}\overline{\Delta_{\epsilon}}_{N,ij} = \epsilon\,[1+o(1)], \qquad \epsilon \to 0. \tag{6.105}$$

The block constraint in (6.105) implies that the sum over each block must be of order ϵ. Hence

$$\overline{\Delta_{\epsilon}}_{N,ij} = O(\epsilon), \qquad \epsilon \to 0, \qquad \forall\,(i,j). \tag{6.106}$$

There are two cases to distinguish: all blocks contribute to the constraint with a term of order ϵ (balanced perturbation), or some of the blocks contribute to the constraint with a term of order ϵ and some with $o(\epsilon)$. Analogously to the analysis in Section 6.3.2, using Lemma 6.3.8, 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 small perturbations of order ϵ 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 analyze next.

Lemma 6.3.9 (Balanced perturbations).
Suppose that $\Delta_\epsilon = \epsilon\Delta$ with $\Delta\colon [0,1]^2 \to [0,\infty)$. Then

$$\delta_r(\epsilon) = \left(\epsilon + \epsilon\log\left(\frac{N_r^1}{\epsilon}\right)\right)[1+O(\epsilon^{1/2})] + O(\epsilon^2), \qquad \epsilon \to 0. \tag{6.107}$$

Proof. By (6.95) and (6.99),

$$\begin{aligned}\delta_r(\epsilon) &= \sup_{\substack{\Delta\colon [0,1]^2 \to [0,\infty)\\ 1-\epsilon\Delta \in \mathcal{W}\\ \int_{[0,1]^2}\Delta = 1+O(\epsilon^{1/2})}} (I_r(1) - I_r(1-\epsilon\Delta))\\ &= \sup_{\substack{\Delta\colon [0,1]^2 \to [0,\infty)\\ 1-\epsilon\Delta \in \mathcal{W}\\ \int_{[0,1]^2}\Delta = 1+O(\epsilon^{1/2})}} \int_{[0,1]^2} \left(\epsilon\Delta\log\left(\frac{1-\epsilon\Delta}{\epsilon\Delta}\frac{1-r}{r}\right) - \log(1-\epsilon\Delta)\right).\end{aligned} \tag{6.108}$$

The integral in (6.108) equals

$$\int_{[0,1]^2}\left(\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)-(1-\epsilon\Delta)\log(1-\epsilon\Delta)\right)=\int_{[0,1]^2}\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)+\epsilon\int_{[0,1]^2}\Delta+O(\epsilon^2). \tag{6.109}$$

Hence

$$\delta_r(\epsilon)=\left(\epsilon+\sup_{\substack{\Delta:\ [0,1]^2\to[0,\infty)\\ \int_{[0,1]^2}\Delta=1}}\int_{[0,1]^2}\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)\right)[1+\mathrm{O}(\epsilon^{1/2})]+O(\epsilon^2), \tag{6.110}$$

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

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

$$\mathcal{L}_{A_r}(\Delta)=\int_{[0,1]^2}\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)+A_r\int_{[0,1]^2}\Delta, \tag{6.111}$$

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

$$\Delta^\phi(x,y)=e^{-\frac{\epsilon-A_r}{\epsilon}}\frac{1}{\epsilon}\frac{1-r^\phi(x,y)}{r^\phi(x,y)},\qquad(x,y)\in[0,1]^2,\qquad\phi\in\mathcal{M}. \tag{6.112}$$

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

$$\Delta^\phi(x,y)=\frac{1}{N_r^1}\frac{1-r^\phi(x,y)}{r^\phi(x,y)},\qquad(x,y)\in[0,1]^2,\qquad\phi\in\mathcal{M}, \tag{6.113}$$

with $N_r^1=\int_{[0,1]^2}\frac{(1-r)}{r}$. Hence the supremum in (6.110) becomes

$$\sup_{\phi\in\mathcal{M}}\int_{[0,1]^2}\epsilon\Delta^\phi\log\left(\frac{1-r}{\epsilon\Delta^\phi r}\right). \tag{6.114}$$

We have

$$\int_{[0,1]^2}\epsilon\Delta^\phi\log\left(\frac{1-r}{\epsilon\Delta^\phi r}\right)=\epsilon\log\left(\frac{N_r^1}{\epsilon}\right)-\epsilon\int_{[0,1]^2}\Delta^\phi\log\left(\frac{\Delta^\phi}{\Delta}\right), \tag{6.115}$$

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

$$\int_{[0,1]^2}\Delta^\phi\log\left(\frac{\Delta^\phi}{\Delta}\right)=\int_{[0,1]^2}\Delta\, s\Big(\frac{\Delta^\phi}{\Delta}\Big)\geq s\Big(\int_{[0,1]^2}\Delta\frac{\Delta^\phi}{\Delta}\Big)=s\Big(\int_{[0,1]^2}\Delta^\phi\Big)=s(1)=0, \tag{6.116}$$

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 (6.114) is uniquely attained at $\phi = \mathrm{Id}$ and equals

$$\int_{[0,1]^2} \epsilon \frac{1}{N_r^1} \frac{(1-r)}{r} \log\left(\frac{N_r^1}{\epsilon}\right) = \epsilon \log\left(\frac{N_r^1}{\epsilon}\right). \tag{6.117}$$

Consequently, (6.110) gives (6.107), and this completes the proof. □

Proof of Theorems 6.1.6(ii)–6.1.7(ii). The claim in Theorem 6.1.6(ii) is settled by Lemma 6.3.9. Since we have shown that a balanced perturbation is optimal, (6.113) settles the claim in Theorem 6.1.7(ii). □

§6.3.4 Proof: perturbation near the left end

For $\beta = \epsilon$ 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

$$\psi_r(\epsilon) = \inf_{\substack{\Delta_\epsilon \colon [0,1]^2 \to [0,\infty) \\ \Delta_\epsilon \in \mathcal{W} \\ \|T_{\Delta_\epsilon}\| = \epsilon}} I_r(\Delta_\epsilon) \tag{6.118}$$

with

$$\psi_r(0) = I_r(0) \tag{6.119}$$

by computing the difference

$$\delta_r(\epsilon) = \psi_r(\epsilon) - \psi_r(0) \tag{6.120}$$

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

We claim that analyzing (6.120) is equivalent to analyzing

$$\delta_{\hat{r}}(\epsilon) = \phi_{\hat{r}}(1) - \phi_{\hat{r}}(1-\epsilon), \tag{6.121}$$

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

$$\hat{r}(x,y) = 1 - r(x,y), \qquad (x,y) \in [0,1]^2. \tag{6.122}$$

Indeed,

$$I_r(0) = \int_{[0,1]^2} \mathcal{R}(0 \mid r) = \int_{[0,1]^2} \log\left(\frac{1}{1-r}\right) = \int_{[0,1]^2} \mathcal{R}(1 \mid \hat{r}) = I_{\hat{r}}(1) \tag{6.123}$$

and

$$\begin{aligned} I_r(\Delta_\epsilon) &= \int_{[0,1]^2} \mathcal{R}(\Delta_\epsilon \mid r) = \int_{[0,1]^2} \left(\Delta_\epsilon \log\left(\frac{\Delta_\epsilon}{r}\right) + (1-\Delta_\epsilon)\log\left(\frac{1-\Delta_\epsilon}{1-r}\right)\right) \\ &= \int_{[0,1]^2} \mathcal{R}(1-\Delta_\epsilon \mid \hat{r}) = I_{\hat{r}}(1-\Delta_\epsilon). \end{aligned} \tag{6.124}$$

We can therefore use the results in Section 6.3.3. From Lemma 6.3.9 we know that

$$\delta_{\hat{r}}(\epsilon) = \left(\epsilon + \epsilon \log\left(\frac{N^1_{\hat{r}}}{\epsilon}\right)\right)[1 + O(\epsilon^{1/2})] + O(\epsilon^2), \qquad \epsilon \to 0, \tag{6.125}$$

and hence we obtain

$$\delta_{r}(\epsilon) = \left(\epsilon + \epsilon \log\left(\frac{N^0_{r}}{\epsilon}\right)\right)[1 + O(\epsilon^{1/2})] + O(\epsilon^2), \qquad \epsilon \to 0. \tag{6.126}$$

The optimal perturbation is then given by the balanced perturbation $\Delta_\epsilon = \epsilon\Delta$ with

$$\Delta(x,y) = \frac{1}{N^0_r}\,\frac{r(x,y)}{1-r(x,y)}, \qquad (x,y) \in [0,1]^2, \tag{6.127}$$

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

Proof of Theorems 6.1.6(iii)–6.1.7(iii). The claim in Theorem 6.1.6(iii) is settled by the scaling in (6.126). Since we have shown that a balanced perturbation is optimal, (6.127) settles the claim in Theorem 6.1.7(iii). □

§F Appendix: finite-rank expansion

The following lemma shows how the expansion around reference graphons of rank 1 can be extended to finite rank.

Lemma F.1 (Finite-rank expansion).
Consider a graphon $\bar{h} \in \mathcal{W}$ such that

$$\bar{h}(x,y) = \sum_{i=1}^{k} \theta_i \bar{\nu}_i(x)\bar{\nu}_i(y), \qquad (x,y) \in [0,1]^2, \tag{6.128}$$

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 $\epsilon > 0$ such that, for any $h \in \mathcal{W}$ satisfying $\|T_{h-\bar{h}}\| < \min(\epsilon, \|T_h\|)$, the operator norm $\|T_h\|$ solves the equation

$$\|T_h\| = \lambda_k\left(\sum_{n\in\mathbb{N}_0} \frac{1}{\|T_h\|^n}\mathcal{F}_n(h,\bar{h})\right), \tag{6.129}$$

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

$$\sqrt{\theta_i\theta_j}\int_{[0,1]^2} \bar{\nu}_i(x)(h-\bar{h})^n(x,y)\bar{\nu}_j(y)\,dx\,dy \tag{6.130}$$

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

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

$$T_h u = \mu u. \tag{6.131}$$

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

$$(\mu - T_g)u = T_{\bar{h}}u. \tag{6.132}$$

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

$$u = (\mu - T_g)^{-1}T_{\bar{h}}u = \sum_{j=1}^{k} \theta_j \langle \bar{\nu}_j, u\rangle (\mu - T_g)^{-1}\bar{\nu}_j. \tag{6.133}$$

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

$$\langle \bar{\nu}_i, u\rangle = \sum_{j=1}^{k} \theta_j \langle \bar{\nu}_j, u\rangle \langle \bar{\nu}_i, (\mu - T_g)^{-1}\bar{\nu}_j\rangle. \tag{6.134}$$

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

$$Mv = \mu v, \tag{6.135}$$

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

$$M_{ij} = \sqrt{\theta_i\theta_j}\left\langle \bar{\nu}_i, \left(1 - \frac{T_g}{\mu}\right)^{-1} \bar{\nu}_j \right\rangle, \qquad 1 \le i,j \le k, \tag{6.136}$$

and

$$v = \left(\sqrt{\theta}_1\langle \bar{\nu}_1, u\rangle, \ldots, \sqrt{\theta}_k\langle \bar{\nu}_k, u\rangle\right)'. \tag{6.137}$$

The first entry of v is non-zero for ϵ small with $\|T_g\| < \epsilon$. Thus, (6.135) means that μ 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 $\|T_g\|$,

$$\mu = \lambda_k(M). \tag{6.138}$$

With the help of the observation

$$M_{ij} = \sqrt{\theta_i\theta_j}\sum_{n\in\mathbb{N}_0} \frac{1}{\mu^n}\langle \bar{\nu}_i, g^n\bar{\nu}_j\rangle, \qquad 1 \le i,j \le k, \tag{6.139}$$

i.e.,

$$M = \sum_{n\in\mathbb{N}_0} \frac{1}{\mu^n}\mathcal{F}_n(h,\bar{h}), \tag{6.140}$$

this completes the proof. □