Graph sequences sampled from Robinson graphons

Mahya Ghandehari[a,*], Jeannette Janssen[b]

[a] *Department of Mathematical Sciences, University of Delaware, Newark, DE, USA, 19716*
[b] *Department of Mathematics & Statistics, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 3J5*

Abstract

The function Γ on the space of graphons, introduced in [CGH+15], aims to measure the extent to which a graphon w exhibits the *Robinson* property: for all $x<y<z$, $w(x,z)\leq \min\{w(x,y),w(y,z)\}$. Robinson graphons form a model for graphs with a natural line embedding so that most edges are local. The function Γ is compatible with the cut-norm $\|\cdot\|_\square$, in the sense that graphons close in cut-norm have similar Γ-values. In particular, any graphon close in cut-norm to the set of all Robinson graphons has small Γ-values. Here we show the converse, by proving that every graphon w can be approximated by a Robinson graphon R_w so that $\|w-R_w\|_\square$ is bounded in terms of $\Gamma(w)$. We then use classical techniques from functional analysis to show that a converging graph sequence $\{G_n\}$ converges to a Robinson graphon if and only if $\Gamma(G_n)\to 0$. Finally, using probabilistic techniques we show that the rate of convergence of Γ for graph sequences sampled from a Robinson graphon can differ substantially depending on how strongly w exhibits the Robinson property.

Keywords: graphon, Robinson, spatial network
2020 MSC: 05C50, 05C62

1. Introduction

A graphon is a symmetric measurable function from $[0,1]^2$ to $[0,1]$. Graphons were introduced in [LS06] as limit objects of convergent (dense) graph sequences. Every graphon w gives rise to a rather general random graph model whose samples are graphs of any desired size. Such so-called w-random graphs are important as "continuous" generalizations of stochastic block models, which have been prominent tools for modeling and analyzing large networks; see [Abb17] for a survey on recent developments in the field of stochastic block models and their applications in community detection.

In this paper, we focus our attention on graphons w where the associated w-random graph model exhibits a spatial, line-embedded structure. In such models, vertices can be identified with points on the line segment $[0,1]$, and the link probability between points $x<y$ increases as y moves closer to x. See for example [CHW19, HRH02] for such latent space models for social networks, and [SAC19] for an overview on continuous latent space models in general. Graphons associated with such models are called *Robinson graphons*. Namely, a graphon $w:[0,1]^2\to[0,1]$ is Robinson if, for all $0\leq x\leq y\leq z\leq 1$,

$$w(x,z)\leq \min\{w(x,y),w(y,z)\}. \tag{1}$$

Robinson graphons were introduced in [CGH+15] under the name of diagonally increasing graphons. The new terminology is compatible with the concept of Robinson matrices (also known as R-matrices) which appear in the study of the well-known and challenging problem of *seriation*. We refer the reader to [Lii10] for a historical overview of seriation and its applications, and to [CS11, PF14, FdV16,

*Corresponding author
Email address: mahya@udel.edu (Mahya Ghandehari)

LST17, FMR19] for recent advancements and new methodologies for seriation. Robinson matrices are significant from the graph theoretic point of view as well: the adjacency matrix of a graph, if labeled properly, is Robinson precisely when the graph is a unit interval graph.

When dealing with real-life networks, an interesting question is whether a graph resembles an instance of a w-random process with Robinson w. One expects random instances of Robinson graphons to be "almost Robinson" if the sample size is large. That is, the sampled graph is expected to exhibit asymptotic aggregated line-embedded behavior, in the sense that, for any two intervals $I_1 \leq I_2$ in $[0,1]$, the density of edges joining vertices with labels in I_1 to those with labels in I_2 will be not significantly larger than the density of edges amongst vertices whose labels fall in the interval $[\sup(I_1), \inf(I_2)]$. Here, we pursue a systematic approach for formalizing these concepts.

This paper builds upon the previous work of the authors and collaborators in [CGH$^+$15], where we introduced a function Γ on the space of graphons such that

1. $\Gamma(w) = 0$ precisely when w is a Robinson graphon.

2. If w is close to a Robinson graphon, then $\Gamma(w)$ is close to 0.

Intuitively, we think of Γ as a gauge of Robinson property. Here, the distance between graphons is measured by the cut-norm, denoted by $\|\cdot\|_\square$, which has close relation with the theory of (dense) graph limits. Indeed, the cut-norm gives rise to the correct metric to define convergence of growing sequences of dense graphs in the sense of Lovász-Szegedy [LS06]. This is because the cut-norm is robust under sampling. Namely, for a growing sequence of graphs $\{G_n\}$ sampled from a graphon w, almost surely, the graphs in the sequence $\{G_n\}$ can be labeled so that $\|w_{G_n} - w\|_\square \to 0$. Here w_{G_n} is the graphon associated with the adjacency matrix of G_n; for details see Section 2.

Our original motivation for introducing Γ in [CGH$^+$15] was to recognize graph sequences which are sampled from Robinson graphons. To complete this task, we need to strengthen property 2 to an "if and only if" condition. This, however, proved to be a rather challenging problem, which we resolve in the current article. To our knowledge, Γ is the first $\|\cdot\|_\square$-norm continuous function which allows the recognition of graph sequences that are eventually Robinson. It may not be possible to compute the exact value of Γ, since it involves taking a supremum over all measurable sets. However, the continuity results of [CGH$^+$15] guarantee that it can be approximated as accurately as one wants. This justifies our claim that Γ, or the graph/graphon parameter it induces, is a suitable candidate for formalizing the notion of *almost Robinson graphs*, i.e. large graphs sampled from Robinson graphons.

1.1. Contribution of this paper

In this paper we show that (1) Γ indicates closeness to a Robinson graphon, (2) Γ identifies graph sequences converging to a Robinson graphon, and (3) Γ-values of graph sequences sampled from a Robinson graphon w converge to zero at different rates depending on the strength of the Robinson property of w.

I. Robinson approximation of graphons

The main result of this article demonstrates the stability of Γ near 0, in the sense that if $\Gamma(w)$ is close to 0, then w must be close to a Robinson graphon. This resolves [CGH$^+$15, Conjecture 6.5]. Even though sections 3 and 4 deal with the particular function Γ, the results hold for any Γ-type function, i.e. any function on the collection of graphons satisfying a natural condition (as in Lemma 4.1), and attaining 0 when applied to Robinson graphons.

THEOREM 1.1. *Let $w : [0,1]^2 \to [0,1]$ be a graphon. Then there exists a Robinson graphon u so that $\|w - u\|_\square \leq 14\Gamma(w)^{1/7}$.*

In Section 3, we define a Robinson approximation R_w for a graphon w (see Definition 3.1), and in Section 4 we show that $\|w - R_w\|_\square \leq 14\Gamma(w)^{1/7}$. Even though the Robinson approximation R_w is easy to state, the proof is rather technical and involves delicate estimates. The significance of Theorem 1.1 lies in the fact that it measures the error of approximation in terms of cut-norm, which is the suitable norm when studying converging w-random graph sequences. The complicated nature of the proof is due to the facts that firstly cut-norm is not as easy to compute as ℓ^p norms; and secondly, the function Γ is defined via certain aggregated averages to ensure its continuity. Finally, suppose a given graph G is sampled from *some* Robinson graphon. The Robinson approximation of w_G provides an approximation for the underlying graphon. This is an instance of the *graphon estimation problem*, where the goal is to invert the sampling process and to recover a graphon from a sampled graph.

II. Recognition of graph sequences sampled from Robinson graphons

In Section 5, we present an important application of Theorem 1.1. Namely, we combine Theorem 1.1 with some classical techniques from functional analysis to prove the following:

THEOREM 1.2. *Let $\{G_n\}$ be a convergent sequence of (dense) graphs. If $\Gamma(G_n) \to 0$, then $\{G_n\}$ converges to a Robinson graphon.*

The limit object of a convergent graph sequence is an equivalence class of graphons defined by the cut-distance $\delta_\square$, rather than the cut-norm itself (see Section 2). The cut-distance produces a graph limit theory that applies to isomorphism classes of graphs. In [CGH+15], it was shown that if $\Gamma(G_n) \to 0$, then the limit object of $\{G_n\}$ can be represented by graphons u with arbitrary small $\Gamma(u)$. We were, however, unable to show that the limit object can be represented by an actual Robinson graphon; a gap which we close by Theorem 1.2 of this article. We show a similar result for the graph/graphon parameter induced from Γ, denoted by $\widetilde{\Gamma}$. Theorem 1.2, together with property 2, illustrates the significance of Γ by proving that it identifies almost Robinson graphs.

III. Rate of decay for samples of a Robinson graphon

Not all Robinson graphons exhibit the linearly embedded property to the same degree. Consider

$$w_1(x,y) = \begin{cases} p & \text{if } |y-x| \leq d, \\ 0 & \text{otherwise.} \end{cases} \qquad w_2(x,y) = p - c|y-x|$$

where $p \in (0,1)$, $d \in (0, 0.5)$, and $c \in (0, p]$, which are both Robinson. Intuitively, however, w_2 has a stronger line-embedded representation. In Section 6, we show that there is indeed a difference in order between 'flat' graphons like w_1, and 'steep' graphons like w_2.

THEOREM 1.3. *Let $w : [0,1]^2 \to (0,1)$ be a Robinson graphon, and let $G \sim \mathbb{G}(n, w)$.*

(i) If w has a flat *region, i.e. there exist measurable sets $S, T \subseteq [0,1]$ with positive measure such that $w|_{S\times T} = p$, then with exponential probability for a graph G sampled from w, $\widetilde{\Gamma}(G) = \Omega(n^{-1/2})$.*

(ii) If w is a steep *graphon, i.e. its partial derivatives exist and are bounded away from 0, then with exponential probability for a graph G sampled from w, $\widetilde{\Gamma}(G) = O(n^{-2/3})$.*

1.2. Similar graph parameters

In [GJ19], the authors introduced a function Γ_1 on the space of matrices, which attains 0 exactly when it is applied to a Robinson matrix. While Γ_1 is easy to compute, it fails to be continuous in cut-norm (or equivalently the graph limit topology). So Γ_1 is not a suitable Robinson measurement for growing networks, whereas Γ provides us with a tool to measure Robinson resemblance of large graphs.

Finally, we remark that this article tackles an instance of the question "given a graphon with specific properties, how can we infer properties of the graphs which converge to it?" These types of questions or their reverse versions have been studied for various classes of graphs/graphons; see for example [BJR12] for graph sequences converging to monotone graphons and [DHJ08] for random threshold graphs.

2. Definitions, notations and background

We denote by W_0 the set of all measurable functions $w:[0,1]^2 \to [0,1]$ which are symmetric, i.e. $w(x,y)=w(y,x)$ for every point (x,y) in $[0,1]^2$. Let W denote the span of W_0, i.e. the set of all real-valued bounded symmetric and measurable functions on $[0,1]^2$. Functions in W_0 are called graphons. Every $n\times n$ symmetric matrix $A=[a_{ij}]$ can be identified with a graphon, denoted by w_A, in the following manner: Partition $[0,1]$ into n equal-sized intervals I_i. For every $i,j\in\{1,\dots,n\}$, let w_A attain a_{ij} on $I_i\times I_j$. Every labeled graph G on n vertices can be identified with the graphon associated with the adjacency matrix of G. We denote this graphon by w_G.

2.1. Cut-norm, cut-distance, graph limits, and w-random graphs

The topology described by convergent (dense) graph sequences can be formalized by endowing W with the cut-norm, introduced in [FK99]. For $w\in W$, the cut-norm is defined as:

$$\|w\|_\square = \sup_{S,T\subset[0,1]} \left| \int_{S\times T} w(x,y)dxdy \right|, \tag{2}$$

where the supremum is taken over all measurable subsets S,T of $[0,1]$. To develop an unlabeled graph limit theory, the cut-distance between $u,w\in W$ is defined as follows.

$$\delta_\square(u,w) = \inf_{\phi\in\Phi} \|u^\phi - w\|_\square, \tag{3}$$

where Φ is the space of all measure preserving bijections on $[0,1]$, and $w^\phi(x,y)=w(\phi(x),\phi(y))$. This definition ensures that $\delta_\square(w,u)=0$ when the graphons w and u are associated with the same graph G with two different vertex labelings. In general, two graphons u and w are said to be $\delta_\square$-*equivalent* (or equivalent, for short), if $\delta_\square(u,w)=0$.

It is known that a graph sequence $\{G_n\}$ converges in the sense of Lovász-Szegedy whenever the corresponding sequence of graphons $\{w_{G_n}\}$ is $\delta_\square$-Cauchy. The limit object for such a convergent sequence can be represented as a graphon in W_0 (not necessarily integer-valued, or corresponding to a graph). That the graph sequence $\{G_n\}$ is convergent to a limit object $w\in W_0$ is equivalent to $\delta_\square(w_{G_n},w)\to 0$ as n tends to infinity, which in turn is equivalent to the existence of suitable labelings for vertices of G_n for which we have

$$\|w_{G_n} - w\|_\square = \sup_{S,T\subset[0,1]} \Big| \int_{S\times T} (w_{G_n}-w) \Big| \to 0. \tag{4}$$

See [BCL+11, Theorem 2.3] for the above convergence results.

The concept of *w-random graphs* was introduced in [LS06], as a tool for generating examples of convergent graph sequences. For a graphon w, we define the random process $\mathbb{G}(n,w)$ on the vertex set $\{1,2,\dots,n\}$, where edges are formed according to w in two steps. First, each vertex i receives a label x_i drawn uniformly at random from $[0,1]$. Next, for each pair of vertices $i<j$ independently, an edge $\{i,j\}$ is added with probability $w(x_i,x_j)$. Such edge-independent w-random graphs arise naturally in the theory of graph limits. In fact, almost surely the sequence $\{\mathbb{G}(n,w)\}_n$ forms a convergent graph sequence, for which the limit object is just the graphon w. See [Lov12] for a comprehensive account of dense graph limit theory.

2.2. Functions Γ and $\widetilde{\Gamma}$

We now give the definition of the function Γ, which is a non-negative valued function on W. Note that Γ is not a graphon parameter, as its value does not solely depend on the equivalence class of a given graphon, but rather on the actual representative itself.

Definition 2.1 ([CGH+15]). For a function w in W, and a measurable subset $A \subseteq [0,1]$, we define

$$\begin{aligned}\Gamma(w,A) &= \int\int_{y<z} \Big[\int_{x\in A\cap[0,y]} (w(x,z)-w(x,y))\,dx\Big]_+ dydz \\ &+ \int\int_{y<z} \Big[\int_{x\in A\cap[z,1]} (w(x,y)-w(x,z))\,dx\Big]_+ dydz,\end{aligned}$$

where $[x]_+ := \max\{x,0\}$. Moreover, $\Gamma(w)$ is defined as

$$\Gamma(w) = \sup\big\{\Gamma(w,A) : A \subseteq [0,1] \text{ measurable}\big\}.$$

The function Γ attains 0 when applied to a Robinson graphon. It was shown in [CGH+15, Proposition 4.2] that $\Gamma(w) = 0$ if and only if w is a.e. equal to a Robinson graphon. Namely, $\Gamma(w) = 0$ precisely when there exists a Robinson graphon u so that $\|u-w\|_\square = 0$. This fact is indeed a trivial case of Theorem 1.1.

One can think of Γ as a function on labeled graphs in a natural manner, namely $\Gamma(G) = \Gamma(w_G)$ for a labeled graph G. When dealing with graphs, Γ identifies unit interval graphs labeled "properly", i.e. labeled unit interval graphs whose adjacency matrices are Robinson. We denote such labeling of a graph, if exists, a *Robinson labeling.* To turn Γ into a graphon/graph parameter, we consider the following natural definition:

$$\widetilde{\Gamma}(w) = \inf\{\Gamma(u) : \delta_\square(u,w) = 0\}, \tag{5}$$

$$\widetilde{\Gamma}(G) = \min\{\Gamma(w_H) : \ H \text{ is a labeled graph isomorphic to } G\}. \tag{6}$$

It was shown in [CGH+15, Theorem 6.4] that $\widetilde{\Gamma}$ is $\delta_\square$-continuous. Note that $\delta_\square(w_G,w_H) = 0$ for any labeled graph H isomorphic to G, so $\widetilde{\Gamma}(w_G) \le \widetilde{\Gamma}(G)$. From Theorem 6.4 in [CGH+15] we also have that, for any converging graph sequence $\{G_n\}$, the numerical sequences $\{\widetilde{\Gamma}(w_{G_n})\}$ and $\{\widetilde{\Gamma}(G_n)\}$ converge to the same value. We conjecture that, in general, $\widetilde{\Gamma}(G) \le \widetilde{\Gamma}(w_G) + O(\frac{1}{|V(G)|})$, but leave this question aside since it is tangential to the main theme of the paper.

Remark 2.2. The concepts of $\Gamma(G)$ and $\widetilde{\Gamma}(G)$ were formulated using a similar but discrete approach in [CGH+15], and were denoted by $\Gamma^*(G,\prec)$ and $\Gamma^*(G)$ respectively, where $\prec$ indicates the ordering on $V(G)$. It turns out that the two approaches result in asymptotically equal functions; see Section 6 for more details. To declutter notations, we have chosen to work with the simpler definitions $\Gamma(G)$ and $\widetilde{\Gamma}(G)$ in the present article.

3. Robinson approximation of graphons

In this section, we define a Robinson approximation R_w for a graphon w. Later, we will prove that the error of approximation, namely $\|w-R_w\|_\square$, is bounded in terms of $\Gamma(w)$. The following notations will be used in the rest of this article.

Notation. Let S and T be measurable subsets of $[0,1]$.

(i) We write $S \le T$ to signify that for all $x\in S$, $y \in T$, we have $x \le y$.

(ii) The product $S \times T$ is called a *cell* or *rectangle* in $[0,1]^2$. We defined the average value of w on the rectangle $S \times T$ to be

$$\overline{w}(S,T) = \frac{1}{|S|\,|T|} \int_S \int_T w(x,y)\ dx\, dy.$$

We now proceed to define the Robinson approximation R_w for a given graphon w. First, we need some preliminary definitions. Since both w and R_w are symmetric, we restrict our attention to the region above the diagonal, which we denote by Δ. Namely, we denote

$$\Delta = \left\{(x,y) \in [0,1]^2 :\ x \leq y\right\}.$$

The value of the Robinson approximation at a given point is determined by the behavior of the graphon on the *upper left (UL)* and *lower right (LR)* regions defined by that point. Precisely, for any $(a,b) \in \Delta$, we define

$$\begin{aligned} \mathrm{UL}(a,b) &= [0,a] \times [b,1], \\ \mathrm{LR}(a,b) &= [a,b] \times [a,b] \cap \Delta. \end{aligned}$$

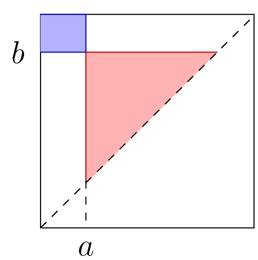

Figure 1: Regions $\mathrm{UL}(a,b)$ (blue) and $\mathrm{LR}(a,b)$ (red)

The upper left and lower right regions provide us with an alternative, and rather more geometric, definition of Robinson graphons. Indeed, a graphon $w \in W$ is Robinson if and only if, for all $(a,b) \in \Delta$ and all $(x,y) \in \mathrm{UL}(a,b)$, $w(x,y) \leq w(a,b)$. Alternatively, w is Robinson if and only if, for all $(a,b) \in \Delta$ and all $(x,y) \in \mathrm{LR}(a,b)$, $w(x,y) \geq w(a,b)$.

Definition 3.1 (Robinson approximation for graphons). Given a graphon $w \in W_0$ with $\Gamma(w) > 0$, the *Robinson approximation* R_w is defined as follows. Let $\alpha = \Gamma(w)^{2/7}$. Then for all $(x,y) \in \Delta$,

$$R_w(x,y) = R_w(y,x) = \sup\left\{\overline{w}(S,T)\ :\ S \times T \subseteq \mathrm{UL}(x,y),\ |S| = |T| = \alpha\right\}, \tag{7}$$

taking the convention that $\sup \emptyset = 0$. Moreover, we set $R_w = w$, if w is Robinson itself.

From the definition of UL and LR regions, it immediately follows that R_w is indeed Robinson. Namely, let $(a,b) \in \Delta$ and $(x,y) \in \mathrm{LR}(a,b)$. Then $\mathrm{UL}(a,b) \subseteq \mathrm{UL}(x,y)$, and thus $R_w(a,b) \leq R_w(x,y)$.

We now restate our main theorem in a more detailed form. The proof will follow in Section 4.

Theorem 3.2 (Equivalent to Theorem 1.1). *Let $w : [0,1]^2 \to [0,1]$ be a graphon, and R_w be as given in Definition 3.1. Then $\|R_w - w\|_\square \leq 14\Gamma(w)^{1/7}$.*

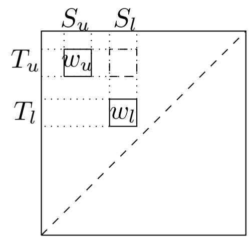

Figure 2: We let $w_u = \overline{w}(S_u \times T_u)$, $w_l = \overline{w}(S_l \times T_l)$. Lemma 4.1 bounds Γ if $w_u > w_l$.

4. Properties of R_w and proof of Theorem 3.2

The proof of Theorem 3.2 relies on a simple, yet important, lower estimate for $\Gamma(w)$, which we present in Subsection 4.1 (see Lemma 4.1). In fact, this lemma inspires the definition of the upper left (UL) and lower right (LR) regions. To obtain a bound on the error $\|w - R_w\|_\square$, we analyze points $(x, y) \in \Delta$ based on the average behavior of the graphon w on the two regions $\mathrm{UL}(x, y)$ and $\mathrm{LR}(x, y)$. Indeed, we define black regions, white regions and grey regions in Δ, each of which containing points with similar average behavior over the corresponding LR and/or UL regions. It turns out that understanding features of these regions and their interactions is the key to our final estimates; this is the content of Subsection 4.2. The proof of Theorem 3.2 is based on the idea that the total area of all the grey regions is small (see Lemma 4.10), while the local average difference between w and R_w inside either the black or white regions is controlled by $\Gamma(w)$. These facts lead to the conclusion that $\|R_w - w\|_\square$ must be small.

4.1. A lower bound on Γ in terms of the "upper left" and "lower right" regions

The lower bound for Γ, presented here, is in terms of average values of w on rectangular regions of $[0, 1]^2$ that are far from and close to the diagonal.

LEMMA 4.1. *Let $w \in W$. Let $S_u \le S_l \le T_l \le T_u$ be subsets of $[0, 1]$ with $|S_u| = |S_l| = |T_l| = |T_u| \ge \alpha$. Then*

$$\Gamma(w) \ge \alpha^3 \Big(\overline{w}(S_u, T_u) - \overline{w}(S_l, T_l)\Big).$$

Proof. Setting $A = S_l \cup T_u$, we get

$$\begin{aligned}
\Gamma(w) &\ge \Gamma(w, A) \\
&\ge \int_{x \in T_l} \int_{y \in T_u} \left[\int_{z \in S_l} w(y, z) - w(x, z) dz \right]_+ dx\, dy \\
&+ \int_{x \in S_u} \int_{y \in S_l} \left[\int_{z \in T_u} w(x, z) - w(y, z) dz \right]_+ dx\, dy \\
&\ge \left[\int_{x \in T_l} \int_{y \in T_u} \int_{z \in S_l} w(y, z) - w(x, z) dz\, dx\, dy \right]_+ \\
&+ \left[\int_{x \in S_u} \int_{y \in S_l} \int_{z \in T_u} w(x, z) - w(y, z) dz\, dx\, dy \right]_+ \\
&= |T_l||T_u||S_l| \big[\overline{w}(S_l, T_u) - \overline{w}(S_l, T_l)\big]_+ + |S_l||S_u||T_u| \big[\overline{w}(S_u, T_u) - \overline{w}(S_l, T_u)\big]_+ \\
&\ge \alpha^3 \big[\overline{w}(S_l, T_u) - \overline{w}(S_l, T_l)\big]_+ + \alpha^3 \big[\overline{w}(S_u, T_u) - \overline{w}(S_l, T_u)\big]_+ \\
&\ge \alpha^3 \big(\overline{w}(S_u, T_u) - \overline{w}(S_l, T_l)\big).
\end{aligned}$$

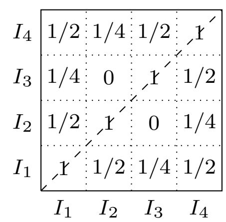

Figure 3: w as in Remark 4.2.

$\square$

REMARK 4.2. The following example shows that the above lower bound is sharp up to a factor of $\frac{5}{4}$. For $i = 1, \ldots, 4$, let $I_i = (\frac{i-1}{4}, \frac{i}{4}]$. Define $w \in W_0$ to be

$$w(x,y) = w(y,x) = \begin{cases} 0 & \text{for } (x,y) \in I_2 \times I_3, \\ \frac{1}{2} & \text{for } (x,y) \in I_1 \times I_2 \cup I_1 \times I_4 \cup I_3 \times I_4, \\ \frac{1}{4} & \text{for } (x,y) \in I_1 \times I_3 \cup I_2 \times I_4, \\ 1 & \text{otherwise,} \end{cases}$$

as shown in Figure 3.

Applying Lemma 4.1 to the sets $I_1 \leq I_2 \leq I_3 \leq I_4$, we obtain that $\Gamma(w) \geq \left(\frac{1}{4}\right)^3 \left(\frac{1}{2}\right)$. Now consider any set $A \subseteq [0,1]$, and let $A_i = A \cap I_i$ and $\alpha_i = |A_i|$ for $i = 1, \ldots, 4$. If $y, z \in I_i$ for some $i = 1, \ldots, 4$, then $w(x,z) = w(x,y)$ for all x. If $y \in I_1 \cup I_2$, then $w(x,z) \leq w(x,y)$ for all $x < y < z$. If $z \in I_3 \cup I_4$, then $w(x,y) \leq w(x,z)$ for all $x > z > y$. Therefore, only pairs $(y,z) \in I_3 \times I_4 \cup I_1 \times I_2$ can make a positive contribution to $\Gamma(w,A)$, and

$$\begin{aligned}
\Gamma(w,A) &= \int_{y\in I_3}\int_{z\in I_4}\Big[\int_{x\in A_1}(\frac{1}{2}-\frac{1}{4})dx + \int_{x\in A_2}(\frac{1}{4}-0)dx + \int_{x\in A_3\cap[0,y]}(\frac{1}{2}-1)dx\Big]_+ \\
&\quad + \int_{y\in I_1}\int_{z\in I_2}\Big[\int_{x\in A_2\cap[z,1]}(\frac{1}{2}-1)dx + \int_{x\in A_3}(\frac{1}{4}-0)dx + \int_{x\in A_4}(\frac{1}{2}-\frac{1}{4})dx\Big]_+ \\
&= \frac{1}{16}\left(\int_{y\in I_3}\Big[\alpha_1+\alpha_2-2\big|A_3\cap[\frac{1}{2},y]\big|\Big]_+ dy + \int_{z\in I_2}\Big[\alpha_3+\alpha_4-2\big|A_2\cap[z,\frac{1}{2}]\big|\Big]_+ dz\right) \\
&\leq \frac{1}{16}\left(\int_{y=\frac{1}{2}}^{\frac{3}{4}}\Big[\alpha_1+\alpha_2-2[y+\alpha_3-\frac{3}{4}]_+\Big]_+ dy + \int_{z=\frac{1}{4}}^{\frac{1}{2}}\Big[\alpha_3+\alpha_4-2[\frac{1}{4}+\alpha_2-z]_+\Big]_+ dz\right) \\
&= \frac{1}{16}\left(\int_{y=\frac{1}{2}}^{\frac{3}{4}-\alpha_3}(\alpha_1+\alpha_2)\,dy + \int_{y=\frac{3}{4}-\alpha_3}^{\frac{3}{4}}\Big[\alpha_1+\alpha_2-2(y+\alpha_3-\frac{3}{4})\Big]_+ dy\right. \\
&\quad \left. + \int_{z=\frac{1}{4}}^{\frac{1}{4}+\alpha_2}\Big[\alpha_3+\alpha_4-2(\frac{1}{4}+\alpha_2-z)\Big]_+ dz + \int_{\frac{1}{4}+\alpha_2}^{\frac{1}{2}}(\alpha_3+\alpha_4)dz\right) \\
&= \frac{1}{16}\left(\int_{y=0}^{\alpha_3}\Big[\alpha_1+\alpha_2-2y\Big]_+ dy + \int_{z=0}^{\alpha_2}\Big[\alpha_3+\alpha_4-2z\Big]_+ dz\right. \\
&\quad \left. +(\alpha_1+\alpha_2)(\frac{1}{4}-\alpha_3) + (\alpha_3+\alpha_4)(\frac{1}{4}-\alpha_2)\right),
\end{aligned}$$

where we used the fact that, for any $y \in I_3$, the measure $\big|A_3 \cap [\frac{1}{2}, y]\big|$ is minimized when $A_3 = [\frac{3}{4} - \alpha_3, \frac{3}{4}]$, in which case we have $\big|A_3 \cap [\frac{1}{2}, y]\big| = [y - \frac{3}{4} + \alpha_3]_+$ for every $y \in I_3$. Similarly, for $z \in I_2$, $\big|A_2 \cap [z, \frac{1}{2}]\big|$

is minimized when $A_2 = (\frac{1}{4}, \frac{1}{4} + \alpha_2]$, and the minimum value equals to $[\frac{1}{4} + \alpha_2 - z]_+$. The right hand side of the above inequality increases when $\alpha_1 = \alpha_4 = \frac{1}{4}$. So we have

$$\Gamma(w, A) \quad \leq \quad \frac{1}{16}\left(\int_{y=0}^{\alpha_3} \left[\frac{1}{4} + \alpha_2 - 2y\right]_+ dy + \int_{z=0}^{\alpha_2} \left[\frac{1}{4} + \alpha_3 - 2z\right]_+ dz + \frac{1}{8} - 2\alpha_2\alpha_3\right). \tag{8}$$

By symmetry, we may assume *wlog* that $\alpha_2 \leq \alpha_3$. Note also that $\alpha_3 \leq \frac{1}{8} + \frac{\alpha_3}{2}$, since $\alpha_3 \leq \frac{1}{4}$. In the case where $\alpha_3 \leq \frac{1}{8} + \frac{\alpha_2}{2}$, the right hand side of (8) reduces to

$$\frac{1}{16}\left((\frac{1}{4} + \alpha_2)\alpha_3 - \alpha_3^2 + (\frac{1}{4} + \alpha_3)\alpha_2 - \alpha_2^2 + \frac{1}{8} - 2\alpha_2\alpha_3\right).$$

This is maximized when $\alpha_2 = \alpha_3 = \frac{1}{8}$, and the above expression has value $(\frac{1}{16})(\frac{5}{32})$.

Now consider the case $\alpha_3 \geq \frac{1}{8} + \frac{\alpha_2}{2}$. Then the right hand side of (8) becomes

$$\frac{1}{16}\left((\frac{1}{4} + \alpha_2)(\frac{1}{8} + \frac{\alpha_2}{2}) - (\frac{1}{8} + \frac{\alpha_2}{2})^2 + (\frac{1}{4} + \alpha_3)\alpha_2 - \alpha_2^2 + \frac{1}{8} - 2\alpha_2\alpha_3\right).$$

This expression is maximized (subject to our condition for this case) when $\alpha_3 = \frac{1}{8} + \frac{\alpha_2}{2}$ and $\alpha_2 = \frac{1}{10}$. The value achieved is $(\frac{1}{16})(\frac{49}{320})$, and hence the previous case achieves the maximum. This gives an upper bound on $\Gamma(w, A)$ of $\frac{1}{16}\left(\frac{5}{32}\right) = \frac{5}{4}\left(\frac{1}{2}\left(\frac{1}{4}\right)^3\right)$. So we get $\frac{1}{2}(\frac{1}{4})^3 \leq \Gamma(w) \leq \frac{5}{4}(\frac{1}{2}(\frac{1}{4})^3)$, as claimed.

4.2. Black, white and grey regions in Δ

For the remainder of this section, fix $w \in W_0$ with $\Gamma(w) > 0$. Let $\alpha = \Gamma(w)^{2/7}$, and let R_w be the Robinson approximation of w as described in Definition 3.1. To prove Theorem 3.2, we need to break down Δ into regions where R_w attains values between certain successive fractional multiples. We do so by having a closer look at the definition of R_w.

Remark 4.3. Fix an integer m; we will determine the optimal value for m later. For $1 \leq k \leq m$ and $(x, y) \in \Delta$, the inequality $\frac{k-1}{m} < R_w(x, y) \leq \frac{k}{m}$ holds precisely when the following two conditions are satisfied:

(i) There exists an $\alpha \times \alpha$ cell $S \times T$ contained in $\mathrm{UL}(x, y)$ on which $\overline{w}(S, T) > \frac{k-1}{m}$.

(ii) For every $\alpha \times \alpha$ cell $S \times T$ contained in $\mathrm{UL}(x, y)$, we have $\overline{w}(S, T) \leq \frac{k}{m}$.

Motivated by this observation, we use the graphon w to split Δ into smaller regions of three types, namely black, white and grey regions as defined below.

Definition 4.4. Let $\alpha = \Gamma(w)^{2/7}$, as was chosen in Definition 3.1. Let m be an integer. For $k = 1, \ldots, m$, define the k'th black region $\mathcal{B}_k$, the k'th white region $\mathcal{W}_k$ and the k'th grey region $\mathcal{G}_k$ as follows.

- $\mathcal{B}_k = \left\{(x, y) \in \Delta : \ x = y \quad \text{or} \quad \exists\ S \times T \subseteq \mathrm{UL}(x, y) \ \text{ with } \ |S| = |T| = \alpha \text{ and } \ \overline{w}(S, T) > \frac{k-1}{m}\right\}$.
- $\mathcal{W}_k = \left\{(x, y) \in \Delta \setminus \mathcal{B}_k : \ \exists\ S \times T \subseteq \mathrm{LR}(x, y) \ \text{ with } \ |S| = |T| = \alpha \ \text{ and } \ \overline{w}(S, T) \leq \frac{k-1}{m}\right\}$.
- $\mathcal{G}_k = \Delta \setminus (\mathcal{B}_k \cup \mathcal{W}_k)$.

Also, define $\mathcal{B}_0 = \mathcal{W}_{m+1} = \Delta$ and $\mathcal{W}_0 = \mathcal{B}_{m+1} = \emptyset$.

Finally, we introduce the regions of Δ on which the value of the Robinson approximation R_w is easy to predict.

Definition 4.5. For $0 \le k \le m$, define $\mathcal{R}_k := \mathcal{B}_k \cap \mathcal{W}_{k+1}$.

See Figure 4 for a demonstration of the black, white and $\mathcal{R}_k$ regions. Note that no set $\mathcal{B}_k$, $k > 0$, can contain any point within distance α of the border of $[0,1]^2$; in the figure this margin is assumed to be invisible. A similar statement holds for $\mathcal{W}_k$ and points within distance α of the diagonal, when $k \le m$. From Remark 4.3, it follows that, for $1 \le k \le m$, $\mathcal{B}_k \setminus \mathcal{B}_{k+1}$ is exactly the region on which $\frac{k-1}{m} < R_w \le \frac{k}{m}$. Note that $\mathcal{R}_k$ is a (possibly proper) subset of $\mathcal{B}_k \setminus \mathcal{B}_{k+1}$. This then leads to the following remark.

Remark 4.6. For $0 \le k \le m$,

$$\text{for all } (x,y) \in \mathcal{R}_k, \qquad \frac{k-1}{m} < R_w(x,y) \le \frac{k}{m}.$$

For $k > 0$ this follows immediately from the discussion above. For $k = 0$, note that for all $1 \le i \le m$, $\mathcal{R}_0 \cap (\mathcal{B}_i \setminus \mathcal{B}_{i+1}) = \emptyset$ and thus R_w cannot attain values in $\cup_{i=1}^m (\frac{i-1}{m}, \frac{i}{m}]$ on $\mathcal{R}_0$. Since R_w is bounded between 0 and 1, it follows that R_w equals zero on $\mathcal{R}_0$.

The intuition behind these definitions is the following. If $(x,y) \in \mathcal{R}_k = \mathcal{B}_k \cap \mathcal{W}_{k+1}$, then $\mathrm{UL}(x,y)$ contains an $\alpha \times \alpha$ cell with "high" w-average, and $\mathrm{LR}(x,y)$ contains an $\alpha \times \alpha$ cell with "low" w-average. These two conditions, together with Lemma 4.1, force us to assign a value to R_w on the region $\mathcal{R}_k$, that lies between these two averages. For $(\mathcal{B}_k \setminus \mathcal{B}_{k+1}) \setminus \mathcal{R}_k$, which lies in the grey region, we do not have such *a priori* knowledge, as Lemma 4.1 cannot be applied here. Luckily, this does not create problems for our error estimates, since the area of the grey regions turns out to be sufficiently small.

Lemma 4.7. *With assumptions as in Definition 4.4, we have the following.*

(i) For every $k = 0, \dots, m$, we have $\mathcal{B}_{k+1} \subseteq \mathcal{B}_k$ and $\mathcal{W}_k \subseteq \mathcal{W}_{k+1}$.

(ii) If $(x,y) \in \mathcal{B}_k$ then $\mathrm{LR}(x,y) \subseteq \mathcal{B}_k$. Similarly, if $(x,y) \in \mathcal{W}_k$ then $\mathrm{UL}(x,y) \subseteq \mathcal{W}_k$.

(iii) If $(x_1,y_1),(x_2,y_2) \in \mathcal{G}_k$ then $\mathrm{UL}(x_1,y_1) \cap \mathrm{LR}(x_2,y_2) \subseteq \mathcal{G}_k$.

(iv) For every $k = 0, \dots, m$, we have $\mathcal{B}_{k+1} \cap \mathcal{W}_k = \emptyset$.

Proof. Parts (i) and (ii) are easy to verify. For part (iv), note that $\mathcal{B}_{k+1} \cap \mathcal{W}_k \subseteq \mathcal{B}_k \cap \mathcal{W}_k$, and the latter set is empty by definition.

To prove (iii), suppose $(z,w) \in \mathrm{UL}(x_1,y_1) \cap \mathrm{LR}(x_2,y_2) \setminus \mathcal{G}_k$, which means that either $(z,w) \in \mathrm{UL}(x_1,y_1) \cap \mathrm{LR}(x_2,y_2) \cap \mathcal{B}_k$ or $(z,w) \in \mathrm{UL}(x_1,y_1) \cap \mathrm{LR}(x_2,y_2) \cap \mathcal{W}_k$. By (ii), the first case implies that $(x_1,y_1) \in \mathcal{B}_k$, and the second one implies that $(x_2,y_2) \in \mathcal{W}_k$. So either case leads to a contradiction. $\square$

Lemma 4.8. *For every $i \ne j$, we have $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$. Moreover,*

$$\Delta \setminus \Big(\bigcup_{k=1}^m \mathcal{G}_k\Big) = \bigcup_{k=0}^m \mathcal{R}_k. \tag{9}$$

Proof. By Definition 4.4, $\mathcal{B}_{k+1} \cap \mathcal{W}_{k+1} = \emptyset$, which implies that $\mathcal{R}_k \cap \mathcal{R}_{k+1} = \emptyset$. Also from Lemma 4.7 parts (i) and (iv), it follows that $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$ if $i \le j-2$, since $\mathcal{W}_{i+1} \cap \mathcal{B}_j \subseteq \mathcal{W}_{i+1} \cap \mathcal{B}_{i+2} = \emptyset$. This proves that the regions $\mathcal{R}_i$ are disjoint. To show Equation (9), observe that

$$\Delta \setminus \Big(\bigcup_{k=1}^m \mathcal{G}_k\Big) \;=\; \Delta \setminus \Big(\bigcup_{k=1}^m (\Delta \setminus (\mathcal{W}_k \cup \mathcal{B}_k))\Big) = \bigcap_{k=1}^m (\mathcal{W}_k \cup \mathcal{B}_k).$$

Now, consider the expansion of $(\mathcal{W}_1 \cup \mathcal{B}_1) \cap (\mathcal{W}_2 \cup \mathcal{B}_2) \cap \ldots \cap (\mathcal{W}_m \cup \mathcal{B}_m)$ into expressions $X_1 \cap \ldots \cap X_m$ with $X_i \in \{\mathcal{W}_i, \mathcal{B}_i\}$, and note that by Lemma 4.7, $X_1 \cap \ldots \cap X_m = \emptyset$ whenever $X_i = \mathcal{W}_i$ and $X_j = \mathcal{B}_j$ for some $i < j$. So, every nonempty term $X_1 \cap \ldots \cap X_m$ from the above expansion must be of one of the following forms:

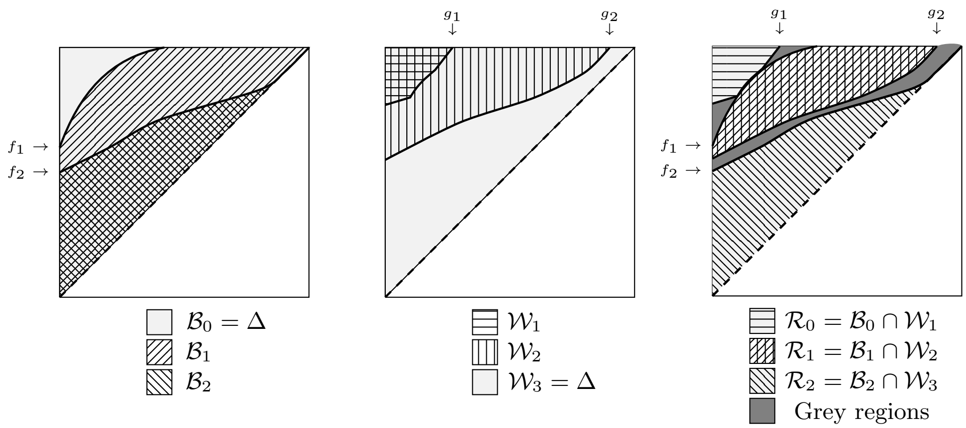

Figure 4: Black and white regions; example for $m = 2$.

(i) $X_1 \cap \ldots \cap X_m = \mathcal{B}_j \cap \mathcal{W}_{j+1}$ with $1 \leq j < m$, if there is at least one black and one white region amongst X_i's.

(ii) $X_1 \cap \ldots \cap X_m = \mathcal{W}_1 \cap \cdots \cap \mathcal{W}_m = \mathcal{W}_1$, if all X_i's are white.

(iii) $X_1 \cap \ldots \cap X_m = \mathcal{B}_1 \cap \cdots \cap \mathcal{B}_m = \mathcal{B}_m$, if all X_i's are black.

This finishes the proof, as $\mathcal{B}_m = \mathcal{B}_m \cap \mathcal{W}_{m+1}$ and $\mathcal{W}_1 = \mathcal{W}_1 \cap \mathcal{B}_0$. $\square$

Remark 4.9. For every $1 \leq k \leq m$, the region $\mathcal{G}_k$ is bounded between *lower* and *upper boundary* functions $f_k, g_k : [0,1] \to [0,1]$, as shown in Figure 4. Indeed, f_k is the upper boundary of $\mathcal{B}_k$, and g_k is the lower boundary of $\mathcal{W}_k$. To be precise, for every $x \in [0,1]$,

$$f_k(x) = \sup\{z \in [x,1] : (x,z) \in \mathcal{B}_k\}, \tag{10}$$

$$g_k(x) = \inf\{z \in [x,1] : (x,z) \in \mathcal{W}_k\}, \tag{11}$$

where we set $\inf \emptyset = 1$, if it appears in the definition of g_k. In addition, we define $f_0(x) = 1$ and $g_{m+1}(x) = x$ for all $x \in [0,1]$, to represent the corresponding boundaries for $\mathcal{B}_0 = \mathcal{W}_{m+1} = \Delta$. Finally, since f_k and g_{k+1} are the upper and lower boundaries of $\mathcal{B}_k$ and $\mathcal{W}_{k+1}$ respectively, if the region $\mathcal{R}_k$ is nonempty, then it is bounded from below by g_{k+1}, and from above by f_k.

From the definition of $\mathcal{B}_k$ and $\mathcal{W}_k$, it is clear that $f_k \leq g_k$. Moreover, both f_k and g_k are increasing functions. Indeed, towards a contradiction suppose f_k is not increasing, i.e. assume that there exist $x_1, x_2 \in [0,1]$, with $x_1 < x_2$ but $f_k(x_1) > f_k(x_2)$. Then, $(x_2, \frac{f_k(x_1)+f_k(x_2)}{2}) \notin \mathcal{B}_k$ but $(x_2, \frac{f_k(x_1)+f_k(x_2)}{2}) \in \mathrm{LR}(x_1, \frac{f_k(x_1)+f_k(x_2)}{2})$, which is a contradiction with $(x_1, \frac{f_k(x_1)+f_k(x_2)}{2}) \in \mathcal{B}_k$ by Lemma 4.7 (ii). The proof for g_k is similar.

Note that boundary functions f_k, g_k are not necessarily continuous. However, being increasing, the boundary functions can only admit jump discontinuities. So, we can naturally extend the graph of a boundary function to a curve by adding appropriate vertical line segments at its points of discontinuity. We call the resulting curve a *boundary curve*, and we denote it by f_k, g_k again.

Lemma 4.10. *Assume the notations and conditions of Definition 4.4, and let $k \in \{1, \ldots, m\}$. Then, $\overline{\mathcal{G}_k}$ does not contain any $\beta \times \beta$ square, where $\beta > \alpha$. Here, $\overline{\mathcal{G}_k}$ denotes the closure of $\mathcal{G}_k$ in $\mathbb{R}^2$.*

Proof. Let $k \in \{1, \ldots, m\}$ be fixed. Clearly, points of $\overline{\mathcal{G}_k}$ lie on or between the lower and upper boundary curves of $\mathcal{G}_k$. Towards a contradiction, suppose there are $\beta > \alpha$, and measurable subsets $S, T \subseteq [0,1]$ with $|S| = |T| = \beta$ for which $S \times T \subseteq \overline{\mathcal{G}_k}$. Let $a_1 = \inf S$, $a_2 = \sup S$, $b_1 = \inf T$, and

$b_2 = \sup T$; note that $a_2 - a_1 \geq \beta$ and $b_2 - b_1 \geq \beta$. Since $\overline{\mathcal{G}_k}$ is closed, we have $(a_i, b_j) \in \overline{\mathcal{G}_k}$ for $i, j = 1, 2$; thus using Lemma 4.7 (iii), one can easily see that $[a_1, a_2] \times [b_1, b_2] \subseteq \overline{\mathcal{G}_k}$. Note that every point in $\overline{\mathcal{G}_k}$, that is not on the lower or upper boundary curves, must be an inner point. Given that the lower and upper boundary functions of $\mathcal{G}_k$ are increasing, we conclude that $(a_1, a_2) \times (b_1, b_2) \subseteq \mathcal{G}_k$. Clearly, $(a_1, a_2) \times (b_1, b_2)$ contains a closed $\alpha \times \alpha$ rectangle, say $[a_1', a_2'] \times [b_1', b_2']$.

The two points $(a_1', b_2'), (a_2', b_1')$ belong to $\mathcal{G}_k$, so both of them fail to satisfy the conditions for $\mathcal{W}_k$ and $\mathcal{B}_k$. In particular, we have $\overline{w}([a_1', a_2'], [b_1', b_2']) \leq \frac{k-1}{m}$ as $(a_2', b_1') \notin \mathcal{B}_k$, and $\overline{w}([a_1', a_2'], [b_1', b_2']) > \frac{k-1}{m}$ as $(a_1', b_2') \notin \mathcal{W}_k$. This is a contradiction. □

4.3. Proof of Theorem 3.2.

We are now ready to prove Theorem 3.2. We use the contrapositive, by showing that, if $\|R_w - w\|_\square$ is large, then $\Gamma(w)$ cannot be small. First observe that if $\Gamma(w)^{1/7} > 1/14$, then $\|R_w - w\|_\square \leq 1 < 14\Gamma(w)^{1/7}$, and there is nothing to prove. So without loss of generality assume that $0 < \Gamma(w)^{1/7} \leq 1/14$.

Let m be an integer chosen from $[\frac{13}{14}\Gamma(w)^{-1/7}, \Gamma(w)^{-1/7}]$. Note that such an integer exists, as the interval has length at least 1. Moreover, from the upper bound on $\Gamma(w)$, we get $m \geq 13$. Let α be as in Definition 3.1, i.e. $\alpha = \Gamma(w)^{2/7}$. Note that we chose these parameters so that $\frac{\alpha^3}{m} \geq \Gamma(w)$. This choice of parameters allows us to use Lemma 4.1 to obtain a contradiction. In the rest of the proof, we will show that if $\|w - R_w\|_\square$ is not small, then we can find two rectangles satisfying conditions of Lemma 4.1, which will result in a contradiction.

Fix $\beta > \alpha$, and let $\delta := 14m\beta$. Towards a contradiction, assume that $\|w - R_w\|_\square > \delta$. So there exist measurable subsets $S, T \subseteq [0, 1]$ so that

$$\Big| \int\int_{S\times T} w - R_w \Big| > \delta.$$

Replacing $S \times T$ with $T \times S$ if necessary, we can assume *wlog* that $\left| \int\int_{(S\times T)\cap\Delta} w - R_w \right| > \frac{\delta}{2}$. We will now show that $(S \times T) \cap \Delta$ contains a region $S^* \times T^*$ so that w is substantially different from R_w on this region. The precise statement is given in the following claim.

Claim 4.11. *There exist sets S^*, T^* of size $|S^*| = |T^*| = \alpha$ and $0 \leq k \leq m$ so that $S^* \times T^* \subseteq \mathcal{R}_k$, and for which we have the following inequality:*

$$\Big| \int\int_{S^*\times T^*} w - R_w \Big| \geq m\alpha^3(3 - \frac{4}{m}). \tag{12}$$

Proof. Let S and T be as described above. Split S into $N_1 = \lceil |S|/\beta \rceil$ subsets $S_1 \leq S_2 \leq \cdots \leq S_{N_1}$ so that $|S_1| = |S_2| = \cdots = |S_{N_1-1}| = \beta$ and $|S_{N_1}| \leq \beta$. Likewise, we split T into $N_2 = \lceil |T|/\beta \rceil$ subsets $T_1 \leq T_2 \leq \cdots \leq T_{N_2}$ with $|T_1| = |T_2| = \cdots = |T_{N_2-1}| = \beta$ and $|T_{N_2}| \leq \beta$. The sets S_i, T_j form a "grid", in which each cell is of the form $S_i \times T_j$ with $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$. Since $|S_{N_1}| \leq \beta$, $|T_{N_2}| \leq \beta$, and $|w - R_w| \leq 1$, we can use triangle inequality to get

$$\begin{aligned}
\sum_{\substack{1\leq i<N_1,\ 1\leq j<N_2 \\ (S_i\times T_j)\cap\Delta\neq\emptyset}} \Big| \int\int_{S_i\times T_j} w - R_w \Big| &\geq \Bigg| \sum_{\substack{1\leq i<N_1,\ 1\leq j<N_2 \\ (S_i\times T_j)\cap\Delta\neq\emptyset}} \int\int_{S_i\times T_j} w - R_w \Bigg| \\
&\geq \Big| \int\int_{(S\times T)\cap\Delta} w - R_w \Big| - \Big| \int\int_{(S_{N_1}\times T)\cup(S\times T_{N_2})} w - R_w \Big| \\
&> \frac{\delta}{2} - 2\beta.
\end{aligned} \tag{13}$$

Recall that $\Delta = \bigcup_{k=0}^{m} \mathcal{R}_k \cup \bigcup_{k=1}^{m} \mathcal{G}_k$, and each of the regions $\mathcal{G}_k$ or $\mathcal{R}_k$ is bounded by boundary curves from the collection $\{f_k, g_l : 1 \le k \le m, 1 \le l \le m+1\}$ as defined in Remark 4.9. Thus, if a cell $S_i \times T_j$ does not cross the graph of any of these boundary curves, then it must be entirely contained inside one closed region $\overline{\mathcal{R}_k}$ or $\overline{\mathcal{G}_k}$. Here, by "a cell crossing a boundary", we mean that the top-left corner of the cell is strictly above the boundary curve, and its bottom-right corner is strictly below the curve. (Note that we need to use the concept of "a cell crossing a boundary" rather than "a cell intersecting with the boundary", as our cells are not necessarily connected subsets of $\mathbb{R}^2$. So a boundary curve can go through a cell, without having to intersect with it.) Next, by Lemma 4.10, none of the grey regions $\overline{\mathcal{G}_k}$ can contain any of the cells $S_i \times T_j$ with $1 \le i < N_1$ and $1 \le j < N_2$. Let $\mathcal{I}$ denote the collection of indices (i, j) with $i < N_1$ and $j < N_2$, for which the associated cells do not lie in a single region $\overline{\mathcal{R}_k}$. From the above discussion, we have

$$\mathcal{I} = \Big\{(i,j) : 1 \le i < N_1, 1 \le j < N_2, \;\text{ and } \exists\, 1 \le k \le m \text{ s.t. } (S_i \times T_j) \text{ crosses } f_k \text{ or } g_k \text{ or } g_{m+1}\Big\}.$$

Note that any cell with indices outside $\mathcal{I}$ is completely contained in the closure of one region $\mathcal{R}_k$.

Now, we bound the size of the set $\mathcal{I}$. By Remark 4.9, the lower and upper boundaries f_k, g_k are increasing functions. We claim that f_k (similarly g_k) crosses at most $2/\beta$ cells from the grid. Indeed, suppose that $S_{n_1} \times T_{n'_1}, S_{n_2} \times T_{n'_2}, \ldots, S_{n_p} \times T_{n'_p}$ is a sequence of distinct cells, all of which cross the graph of f_k. Since f_k is increasing, after relabeling if necessary, we have that $n_1 \le n_2 \le \ldots \le n_p$ and $n'_1 \le n'_2 \le \ldots \le n'_p$. However, Δ cannot contain any such sequence of length more than $N_1 - 1 + N_2 - 1 \le 2/\beta$. Thus, we have

$$|\mathcal{I}| \le \frac{2(2m+1)}{\beta}.$$

Since every cell indexed in $\mathcal{I}$ is of size β^2, we have $|\mathcal{F}| \le (4m+2)\beta$, where $\mathcal{F} = \bigcup_{(i,j)\in\mathcal{I}} S_i \times T_j$. By inequality (13), and the fact that $|w - R_w| \le 1$, we get

$$\sum_{\substack{1 \le i < N_1,\ 1 \le j < N_2 \\ (S_i \times T_j) \cap \Delta \ne \emptyset \\ (i,j) \notin \mathcal{I}}} \left| \int\int_{S_i \times T_j} w - R_w \right| > \frac{\delta}{2} - 2\beta - (4m+2)\beta = 3m\beta - 4\beta.$$

By the pigeonhole principle, since there are at most $(1/\beta)^2$ cells $S_i \times T_j$ of size $\beta \times \beta$, there must exist a cell $S_{i_0} \times T_{j_0} \subseteq \Delta \setminus \mathcal{F}$ so that $|S_{i_0}| = |T_{j_0}| = \beta$ and

$$\Big| \int\int_{S_{i_0} \times T_{j_0}} w - R_w \Big| \ge m\beta^3 (3 - \frac{4}{m}).$$

Since $(i_0, j_0) \notin \mathcal{I}$, we conclude that $S_{i_0} \times T_{j_0}$ lies entirely in the closure $\overline{\mathcal{R}_k} = \overline{\mathcal{B}_k \cap \mathcal{W}_{k+1}}$ for some $0 \le k \le m$. It is easy to see that, there are subsets $S^* \subseteq S_{i_0}$ and $T^* \subseteq T_{j_0}$ with $|S^*| = |T^*| = \alpha$, so that $S^* \times T^* \subseteq \mathcal{R}_k = \mathcal{B}_k \cap \mathcal{W}_{k+1}$ and

$$\Big| \int\int_{S^* \times T^*} w - R_w \Big| \ge m\alpha^2\beta(3 - \frac{4}{m}) \ge m\alpha^3(3 - \frac{4}{m}), \tag{14}$$

which proves the claim. (In the case where $\frac{\alpha}{\beta}$ is a rational number, the existence of such sets S^* and T^* is an easy application of the pigeonhole principle and the fact that $\frac{1}{\beta^2}\Big| \int\int_{S_{i_0} \times T_{j_0}} w - R_w \Big| \ge m\beta(3 - \frac{4}{m})$. The case where $\frac{\alpha}{\beta}$ is irrational follows from a standard density argument.) ☐

We now state and prove a second claim.

CLAIM 4.12. *If sets S^*, T^* and index k as in Claim 4.11 exist, then $\Gamma(w) > \alpha^3/m$.*

Proof. As explained in Remark 4.6, we have $\frac{k-1}{m} < R_w \leq \frac{k}{m}$ on $S^* \times T^*$. To prove the claim, we will consider two cases.

Case 1. Assume first that w has larger average than R_w on $S^* \times T^*$, so $\int\int_{S^* \times T^*} w - R_w \geq m\alpha^3(3 - \frac{4}{m})$. In this case, we have

$$\overline{w}(S^*, T^*) - \frac{k-1}{m} > \overline{w - R_w}(S^*, T^*) \geq \frac{m\alpha^3(3 - \frac{4}{m})}{|S^* \times T^*|} = m\alpha(3 - \frac{4}{m}), \tag{15}$$

where we use that $|S^* \times T^*| = \alpha^2$.

Next we argue that $k < m$. From our choice of parameters, we have that $m^2\alpha \geq (\frac{13}{14})^2 > \frac{1}{2}$, which implies that $2m\alpha - \frac{1}{m} > 0$. Assuming now that $S^* \times T^* \subseteq \mathcal{R}_m$, then by (15) and the facts that $0 \leq w \leq 1$ and $m \geq 13$, we have that

$$1 < 1 + (2m\alpha - \frac{1}{m}) < (1 - \frac{1}{m}) + m\alpha(3 - \frac{4}{m}) < \overline{w}(S^*, T^*) \leq 1,$$

which is a contradiction. This proves that $k < m$.

Now let (x, y) be the lower right corner of $S^* \times T^*$, so $x = \sup S^*$ and $y = \inf T^*$. Then $(x, y) \in \mathcal{W}_{k+1}$, and thus $\mathrm{LR}(x, y)$ contains a region $S_l \times T_l$ so that $|S_l| = |T_l| = \alpha$, and $\overline{w}(S_l, T_l) \leq \frac{k}{m}$. Applying Lemma 4.1 together with inequality (15), we now conclude that

$$\Gamma(w) \geq \alpha^3\big(m\alpha(3 - \frac{4}{m}) + \frac{k-1}{m} - \frac{k}{m}\big).$$

Since $m \geq \max\{\frac{13}{14}\Gamma(w)^{-1/7}, 13\}$, we conclude that

$$\Gamma(w)(\frac{m}{\alpha^3}) \geq m^2\alpha(3 - \frac{4}{m}) - 1 \geq (\frac{13}{14})^2\Gamma(w)^{-2/7}\Gamma(w)^{2/7}(3 - \frac{4}{13}) - 1 > 1, \tag{16}$$

Therefore,

$$\Gamma(w) > \frac{\alpha^3}{m},$$

which proves the claim for the first case.

Case 2. For the second case, assume $\int\int_{S^* \times T^*} R_w - w \geq m\alpha^3(3 - \frac{4}{m})$. This case can only happen if $k > 0$, as R_w attains 0 on $S^* \times T^*$ if $k = 0$. By a similar argument,

$$\frac{k}{m} - \overline{w}(S^*, T^*) \geq m\alpha(3 - \frac{4}{m}). \tag{17}$$

Now let (x, y) be the upper left corner of $S^* \times T^*$, so $x = \inf S^*$ and $y = \sup T^*$. Then $(x, y) \in \mathcal{B}_k$, and thus $\mathrm{UL}(x, y)$ contains a region $S_l \times T_l$ so that $|S_l| = |T_l| = \alpha$, and $\overline{w}(S_l, T_l) > \frac{k-1}{m}$. Applying Lemma 4.1 together with Equations (17) and (16), we now conclude that

$$\Gamma(w) \geq \alpha^3\big(m\alpha(3 - \frac{4}{m}) - \frac{1}{m}\big) = \frac{\alpha^3}{m}\big(m^2\alpha(3 - \frac{4}{m}) - 1\big) > \frac{\alpha^3}{m}.$$

□

Our second claim directly contradicts the assumption that $\Gamma(w) \leq \frac{\alpha^3}{m}$. Thus our assumption that $\|R_w - w\|_\square > \delta$ does not hold, which implies that $\|R_w - w\|_\square \leq \delta = 14m\beta$ for all $\beta > \alpha$. Thus, $\|R_w - w\|_\square \leq 14m\alpha \leq 14\Gamma(w)^{1/7}$. This completes the proof of Theorem 3.2.

5. Recognition of graph sequences sampled from Robinson graphons

Consider a graph sequence $\{G_n\}$ sampled from a graphon $w \in W_0$, i.e. a sequence $\{G_n\}$ that converges to the graphon w in the sense of Lovász-Szegedy, or equivalently $\delta_\square(w_{G_n}, w) \to 0$. By [CGH+15, Theorem 6.4], $\widetilde{\Gamma}$ is continuous, and in particular,

$$\text{If } \ G_n \to w \ \text{ and } \ \Gamma(w) = 0, \ \text{ then } \ \widetilde{\Gamma}(G_n) \to 0. \tag{$\star$}$$

Continuity guarantees a weaker version of the converse: if $G_n \to w$ and $\widetilde{\Gamma}(G_n) \to 0$, then $\widetilde{\Gamma}(w) = 0$. Since $\widetilde{\Gamma}(w) = \inf\{\Gamma(u) : \delta_\square(w, u)\}$, these earlier results do not imply that the $\delta_\square$-equivalence class of w contains a Robinson graphon. In Theorem 5.3 to follow, we will use Theorem 3.2 and the weak* topology of the space of graphons to show the strong version of the converse. The proof of Theorem 5.3 involves approximations of a graphon by step graphons, which we discuss in the following subsection.

5.1. Stepping operator

For an integer $N \in \mathbb{N}$ and a graphon $u \in W_0$, we define the step graphon $u^{(N)}$ as follows: split the interval $[0, 1]$ into N equal-sized subintervals $I_1, \ldots, I_N$, and define

$$u^{(N)}(x, y) = \overline{u}(I_i, I_j), \text{ if } (x, y) \in I_i \times I_j.$$

To avoid clutter of notation, we use $u^{(N)}_{i,j}$ to denote the value of $u^{(N)}$ at every point in $I_i \times I_j$. The operator that assigns to every $u \in W_0$, the step graphon $u^{(N)}$ is called a *stepping operator.* Step graphons approximate u in $\|\cdot\|_1$-norm, and hence in $\|\cdot\|_\square$-norm. In the following lemma, we obtain a universal upper bound for the rate of convergence of the step graphon approximation of a Robinson graphon. Note that the proof of Lemma 5.1 relies heavily on the Robinson structure of the graphon. For a general graphon, a uniform bound for the rate of convergence of step graphon approximations can be obtained by applying the (Weak) Regularity Lemma for graphons (for example, see [Lov12, Lemma 9.9]). However, the bound from the Regularity Lemma cannot be used for our methods of proving Theorem 5.3, since the bound is of the order of $\frac{1}{\sqrt{N}}$, and the partition sets are not necessarily intervals, but rather measurable sets. The following lemma gives us better control on both the partition sets and the error bound.

Lemma 5.1. *Let $u \in W_0$ be a Robinson graphon. Then for every $N \in \mathbb{N}$, $\|u - u^{(N)}\|_1 \leq \frac{7}{N}$*

Proof. Fix $N \in \mathbb{N}$, and note that $u^{(N)}$ is a Robinson graphon as well. Consider the symmetric graphons $u^{(N)}_-$ and $u^{(N)}_+$ that are obtained from $u^{(N)}$ by shifting every cell towards the diagonal (for $u^{(N)}_-$) or away from the diagonal (for $u^{(N)}_+$). That is,

$$u^{(N)}_-(x, y) := u^{(N)}_-(y, x) = \begin{cases} u^{(N)}_{i-1,j+1} & \text{if } (x, y) \in I_i \times I_j \ \text{ and } \ 1 < i \leq j < N \\ 0 & \text{if } (x, y) \in I_i \times I_N \ \text{ or } \ (x, y) \in I_1 \times I_j \end{cases}$$

and

$$u^{(N)}_+(x, y) := u^{(N)}_+(y, x) = \begin{cases} u^{(N)}_{i+1,j-1} & \text{if } (x, y) \in I_i \times I_j, \ 1 \leq i < j \leq N \ \text{ and } j - i \geq 2 \\ 1 & \text{if } (x, y) \in I_i \times I_j \ \text{ and } \ 0 \leq j - i \leq 1 \end{cases}$$

Since u is a Robinson graphon, it can be easily checked that $u^{(N)}_- \leq u \leq u^{(N)}_+$ and $u^{(N)}_- \leq u^{(N)} \leq u^{(N)}_+$.

On the other hand,

$$\begin{aligned}
\|u_+^{(N)}-u_-^{(N)}\|_1 &= \int_{[0,1]^2} u_+^{(N)}-u_-^{(N)} \\
&= \frac{1}{N^2}\big(2\sum_{i=1}^{N-2}\sum_{j=i+2}^{N} u^{(N)}_{i+1,j-1}+\sum_{i=1}^{N}\sum_{j=i-1}^{i+1}1-2\sum_{i=2}^{N-1}\sum_{j=i+1}^{N-1}u^{(N)}_{i-1,j+1}-\sum_{i=2}^{N-1}u^{(N)}_{i-1,i+1}\big) \\
&\leq \frac{7N}{N^2},
\end{aligned}$$

because every pair (i,j) with $1<i<j<N$ and $j-i\geq 3$, contributes exactly once in a positive sum and exactly once in a negative sum, and there are only $7N$ other terms left. Thus, we have $\|u-u^{(N)}\|_1\leq \|u_+^{(N)}-u_-^{(N)}\|_1\leq \frac{7}{N}$. □

5.2. Robinsonian graphons

We are now ready to prove Theorem 1.2, which will follow from Theorem 5.3. First we introduce the following definition, which matches a similar concept in matrix theory.

Definition 5.2. A graphon $w\in W_0$ is called *Robinsonian* if there exists a Robinson graphon $u\in W_0$ such that $\delta_\square(u,w)=0$. In other words, a graphon is Robinsonian if its $\delta_\square$-equivalence class contains a Robinson graphon.

Theorem 5.3. *Let $\{G_n\}_{n\in\mathbb{N}}$ be a growing sequence of graphs converging to a graphon $w\in W_0$. Then, w is Robinsonian if and only if $\widetilde{\Gamma}(G_n)\to 0$.*

Proof. The forward direction is a consequence of continuity of $\widetilde{\Gamma}$. To prove the backward direction, suppose $\widetilde{\Gamma}(G_n)\to 0$. Without loss of generality, assume that every G_n is labeled so that $\widetilde{\Gamma}(G_n)$ is achieved, that is $\widetilde{\Gamma}(G_n)=\Gamma(w_{G_n})$ where w_{G_n} denotes the graphon that represents G_n. From the assumption we have $\Gamma(w_{G_n})\to 0$, and from the definition of convergence of graph sequences, we have that $\delta_\square(w_{G_n},w)\to 0$.

Applying Theorem 3.2, for every $n\in\mathbb{N}$, there exists a Robinson graphon $u_n\in W_0$ such that $\|u_n-w_{G_n}\|_\square\leq 14\Gamma(w_{G_n})^{1/7}$. So $\{u_n\}_{n\in\mathbb{N}}$ is a sequence of Robinson graphons such that $\delta_\square(u_n,w)\to 0$ as $n\to\infty$, because

$$\delta_\square(u_n,w)\leq \delta_\square(u_n,w_{G_n})+\delta_\square(w_{G_n},w)\leq \|u_n-w_{G_n}\|_\square+\delta_\square(w_{G_n},w).$$

Now consider the graphon space W_0 as a subset of $B_1(L^\infty[0,1]^2):=\{f\in L^\infty[0,1]^2:\ \|f\|_\infty\leq 1\}$, namely the closed unit ball of $L^\infty[0,1]^2$. The Banach space $L^\infty[0,1]^2$ is isometrically isomorphic to the (Banach space) dual of $L^1[0,1]^2$, so one can equip $B_1(L^\infty[0,1]^2)$ with the weak* topology induced by this duality. By the Banach-Alaoglu theorem, $B_1(L^\infty[0,1]^2)$ is compact in the weak* topology ([Con90, Theorem 3.1 of Chapter V]). In addition, since $L^1[0,1]^2$ is separable, the unit ball $B_1(L^\infty[0,1]^2)$ is a metrizable space in the weak* topology, and thus it is sequentially compact as well ([Con90, Theorem 5.1 of V]). Thus the sequence $\{u_n\}_{n\in\mathbb{N}}$ in $B_1(L^\infty[0,1]^2)$ has a weak* convergent subsequence.

By going down to that subsequence if necessary, *wlog* we can assume that $\{u_n\}_{n\in\mathbb{N}}$ converges to some $z\in B_1(L^\infty[0,1]^2)$ in the weak* topology, i.e. for every $h\in L^1[0,1]^2$ we have $\int_{[0,1]^2}u_nh\to\int_{[0,1]^2}zh$ as $n\to\infty$. In particular, for every measurable subsets $S,T\subseteq[0,1]$, we have

$$\int_{S\times T}u_n\to\int_{S\times T}z,\ \text{as } n\to\infty. \tag{18}$$

By (18), we get that for every $N\in\mathbb{N}$, the sequence $\{u_n^{(N)}\}$ converges point-wise to $z^{(N)}$. Note that every graphon u_n is Robinson. So for every $N\in\mathbb{N}$, the corresponding step graphons $u_n^{(N)}$ and their point-wise limit $z^{(N)}$ are also Robinson. Finally, since $\|z - z^{(N)}\|_1 \to 0$, z is a Robinson graphon a.e. as well.

Next, we claim that $\delta_\Box(z, w) = 0$. Fix $\epsilon > 0$, and choose $N_0 \in \mathbb{N}$ such that $\frac{7}{N_0} \le \frac{\epsilon}{3}$. From the convergence of $\{u_n^{(N_0)}\}_{n\in\mathbb{N}}$ to $z^{(N_0)}$, pick $m \in \mathbb{N}$ so that $\|u_m^{(N_0)} - z^{(N_0)}\|_\infty \le \frac{\epsilon}{3}$. Applying Lemma 5.1 to the Robinson graphons $z^{(N_0)}$ and $u_m^{(N_0)}$, we have

$$\|u_m - z\|_1 \le \|u_m - u_m^{(N_0)}\|_1 + \|u_m^{(N_0)} - z^{(N_0)}\|_1 + \|z^{(N_0)} - z\|_1 \le \frac{7}{N_0} + \frac{\epsilon}{3} + \frac{7}{N_0} \le \epsilon. \qquad (19)$$

Since $\delta_\Box(w, z) \le \delta_\Box(w, u_m) + \delta_\Box(u_m, z) \le \delta_\Box(w, u_m) + \|u_m - z\|_1$ for all m, and $\delta_\Box(w, u_m) + \|u_m - z\|_1 \to 0$, we have that $\delta_\Box(w, z) = 0$. $\square$

Corollary 5.4. *$\widetilde{\Gamma}(w) = 0$ if and only if w is Robinsonian.*

Proof. Note that $\widetilde{\Gamma}(w)$ is defined as an infimum over all graphons in the $\delta_\Box$-equivalence class of w. The above theorem tells us that this infimum is in fact achieved at 0. $\square$

Remark 5.5. (i) A $\delta_\Box$-equivalence class of graphons may include more than one Robinson graphon, i.e. there is no concept of a "unique Robinson representation" of a $\delta_\Box$-equivalence class of graphons. The same holds when $\widetilde{\Gamma}(w)$ is small, i.e. there may be $u_1, u_2 \in W_0$ with $\delta_\Box(u_1, w) = \delta_\Box(u_2, w) = 0$, $\widetilde{\Gamma}(w) = \Gamma(u_1) = \Gamma(u_2)$, but $\|u_1 - u_2\|_\Box$ is large. This phenomenon is the root of complications in the proof of Theorem 5.3, and the reason that a purely combinatorial proof could not be derived easily.

(ii) A function $\Psi : W \to [0, 1]$ is called a Γ-type function if it is $\|\cdot\|_\Box$-norm continuous, satisfies the condition of Lemma 4.1, and attains 0 when applied to Robinson graphons. It is very easy to verify that Theorem 3.2 and Theorem 5.3 hold for any Γ-type function. In that sense, we think of Γ as a prototype of Γ-type functions, whose definition is very natural.

6. Rate of decay for $\widetilde{\Gamma}$ of samples from Robinson graphons

Let w be a Robinson graphon, and consider a graph sequence $\{G_n\}_{n\in\mathbb{N}}$ with G_n sampled from $\mathbb{G}(n, w)$. From Corollary 6.5 in [CGH+15], we have that $\lim_{n\to\infty} \widetilde{\Gamma}(G_n) = 0$ almost surely. In this section, we obtain upper and lower bounds for the speed of this convergence. In particular, we prove Theorem 1.3, which gives order bounds on $\widetilde{\Gamma}(G_n)$ for graph sequences sampled from graphons with a flat region (rectangular region in $[0, 1]^2$ on which w is constant) and steep graphons (graphons with partial derivatives bounded away from zero). Our results show that the decay is an order of magnitude faster for steep graphons than for graphons with a flat region. This confirms the intuitive notion that for graphons with stronger linearly embedded structures $\widetilde{\Gamma}(G_n)$ converges to zero faster.

6.1. Some definitions and notations for this section

Due to the graph-theoretic nature of our results, we will replace the graphon-based definition $\widetilde{\Gamma}(G)$ by the graph-based parameter $\Gamma^*(G)$, originally defined in [CGH+15] as follows: Let G be a graph with a linear order $\prec$ on its vertex set $V(G)$, and denote $|V(G)| = n$. For every $v \in V(G)$, the collection of all the neighbours of v is denoted by $N(v)$. Also, the down-set $D(v)$ and the up-set $U(v)$ of v are defined as follows:

$$D(v) = \{x \in V(G) : x \prec v\} \ \text{ and } \ U(v) = \{x \in V(G) : v \prec x\}.$$

DEFINITION 6.1. Let $A \subseteq V(G)$, and $\prec$ be a linear order on the vertex set of G. We define,

$$\begin{aligned}\Gamma^*(G,\prec,A) &= \frac{1}{n^3}\sum_{u\prec v}\big[|N(v)\cap A\cap D(u)| - |N(u)\cap A\cap D(u)|\big]_+ \\ &+ \frac{1}{n^3}\sum_{u\prec v}\big[|N(u)\cap A\cap U(v)| - |N(v)\cap A\cap U(v)|\big]_+.\end{aligned}$$

We also define

$$\Gamma^*(G,\prec) = \max_{A\subseteq V(G)} \Gamma^*(G,\prec,A) \quad \text{and} \quad \Gamma^*(G) = \min_{\prec} \Gamma^*(G,\prec),$$

where the minimum is taken over all the linear orderings of $V(G)$.

It follows from [CGH+15, Corollary 5.2] that the parameters Γ^* and $\widetilde{\Gamma}$ are asymptotically equal, i.e. for a sequence of graphs $\{G_n\}$ of increasing order n,

$$\Gamma^*(G_n) = \widetilde{\Gamma}(G_n) + O(\frac{1}{n}). \tag{20}$$

We will use the term *with exponential probability (w.e.p.)* to denote that an event holds asymptotically with probability at least $1-\exp(-c\log^2 n)$ for some positive constant c. Thus, a polynomial number of events that all hold *w.e.p.* will hold simultaneously *w.e.p.* as well. We will make extensive use of a well-known inequality, quoted below, which shows that the sum of bounded independent variables is close to its expected value.

THEOREM 6.2 (Hoeffding's inequality). *Let $\{X_i\}_{i=1}^N$ be a sequence of independent random variables bounded by the interval $[0,1]$, and let $X=\sum_{i=1}^N X_i$. Then*

$$\mathbb{P}(|X - \sum_{i=1}^N \mathbb{E}(X_i)| \geq \alpha N) \leq 2\exp(-2\alpha^2 N). \tag{21}$$

Further concentration bounds used in this section can be found in [Wai19]. For the convenience of the reader, we also include these results in Appendix A.

The proof of Theorem 1.3 will follow from two propositions presented in the following two subsections, each of which addresses one case of the theorem.

6.2. Graphons with a flat region

In this section, we prove Proposition 6.4 from which case (i) of Theorem 1.3 will follow. Namely, we give a lower bound for the convergence rate of Γ^* of graphs sampled from graphons with a constant rectangular region. In particular, this bound may be applied to the constant graphon, which is an extreme case of a graphon that is Robinson, but does not have a clear linear embedding.

Recall that Γ^* is a graph parameter defined as $\min_\prec \Gamma^*(G,\prec)$, where the minimum is taken over all the linear orderings of $V(G)$. For any particular ordering $\prec$, one can use standard probabilistic techniques to show that *w.e.p.* $\Gamma^*(G,\prec) \geq cn^{-1/2}$ for some constant c. This fact is due to the random fluctuation of the outcomes of the vertex pairs whose labels fall in the constant region of w. In order to obtain a lower bound on $\Gamma^*(G)$, we have to establish that the same lower bound remains true for *all* orderings. This cannot be achieved using a simple union bound argument, as the number of orderings of $V(G)$ is superexponential. To address this issue, we partition the orderings into classes according to a very coarse partial order of the vertex set. Next, we establish the desired lower bound on $\Gamma^*(G,\prec)$ for an entire class of orderings simultaneously, and then show that the number of classes is small enough so that a union bound argument guarantees that *w.e.p.* the lower bound holds for all classes simultaneously.

We will need the following lemma.

Lemma 6.3. *Let V_S, V_T be two subsets of a set V, and let $\prec$ be an ordering of V. Let $V_1 \subseteq V_S$ and $V_2 \subseteq V_T$ so that*

$$V_1 \prec V_S \setminus V_1 \text{ and } V_2 \prec V_T \setminus V_2. \tag{22}$$

Then $V_1 \prec V_T \setminus V_2$ or $V_2 \prec V_S \setminus V_1$.

Proof. Let $V = \{v_1, \ldots, v_n\}$ where the vertices are labeled according to $\prec$, i.e. we have $v_1 \prec v_2 \prec \cdots \prec v_n$. Our assumption on $\prec$ implies that if $v_i \in V_1$ and $v_j \in V_S \setminus V_1$, then $i < j$; an equivalent statement holds for V_T and V_2.

For $i = 1, 2$, let k_i be the largest index in V_i, that is, $k_i = \max\{k : v_k \in V_i\}$. Suppose $k_1 \le k_2$. Then $V_1 \subseteq \{v_1, \ldots, v_{k_1}\}$, and by Condition (22) we have $V_T \setminus V_2 \subseteq \{v_{k_2+1}, \ldots, v_n\}$. Therefore $V_1 \prec V_T \setminus V_2$. Similarly, if $k_2 < k_1$ then $V_2 \prec V_S \setminus V_1$. □

Proposition 6.4. *Let w be a graphon with a constant rectangular region, that is, there are measurable sets $S, T \subseteq [0, 1]$ with positive measure and a constant $p \in (0, 1)$ such that w assumes p a.e. on $S \times T$. Then there exist constants $c, \alpha > 0$ so that, for a w-random outcome $G \sim \mathbb{G}(n, w)$ we have*

$$\mathbb{P}(\Gamma^*(G) \ge cn^{-1/2}) \ge 1 - \exp(-\alpha n).$$

Moreover, the constants c, α depend only on p and $\min\{|S|, |T|\}$.

Proof. Let w be as stated, and denote $s := \min\{|S|, |T|\}$. Let $G \sim \mathbb{G}(n, w)$ be an outcome with vertex set $V = \{1, 2, \ldots, n\}$ and labels $x_1, \ldots, x_n$. Define

$$V_S = \{i \in V : x_i \in S\} \text{ and } V_T = \{i \in V : x_i \in T\}.$$

Clearly, $w(x, y) = p$ for all $x \in V_S$, $y \in V_T$. By Hoeffding's inequality, with probability at least $1 - 2\exp(-2(s/4)^2 n)$, we have

$$|V_S| \ge \frac{3}{4} sn \quad \text{and} \quad |V_T| \ge \frac{3}{4} sn. \tag{23}$$

We assume this is the case.

Let $0 < \epsilon < \frac{s}{4}$. (An appropriate choice for ϵ will become apparent at the end of this proof.) Consider partitions of V_S and V_T into sets V_1^S, V_2^S, V_3^S and V_1^T, V_2^T, V_3^T respectively, so that

$$|V_1^S| = |V_2^S| = |V_1^T| = |V_2^T| = \lfloor \epsilon n \rfloor \quad \text{and} \quad (V_1^T \cup V_2^T) \cap V_3^S = \emptyset. \tag{24}$$

We will show that *w.e.p.* for *any* such partition, there exists a sufficiently large subset of V_3^S containing vertices that have larger than expected number of neighbours in V_1^T and smaller than expected number of neighbours in V_2^T, and vice versa.

Let $m = \lfloor \epsilon n \rfloor$. For $i = 1, 2$, let

$$\begin{aligned} B_i^S &= \{y \in V_3^S : |N(y) \cap V_i^T| \ge pm + \sqrt{m}\} \\ C_i^S &= \{y \in V_3^S : |N(y) \cap V_i^T| \le pm - \sqrt{m}\} \end{aligned} \tag{25}$$

We are interested in the sets $B_1^S \cap C_2^S$ and $B_2^S \cap C_1^S$. First, we establish a lower bound on their sizes.

For each $y \in V_3^S$, the variable $|N(y) \cap V_1^T|$ is the result of m Bernoulli trials, with probability of success equal to p, so $\mathbb{E}(|N(y) \cap V_1^T|) = pm$. Moreover, since $V_3^S \cap V_1^T = \emptyset$, these variables are independent. Thus by the central limit theorem, for $y \in V_3^S$, the random variable $Z_m = \frac{|N(y) \cap V_1^T| - pm}{\sqrt{mp(1-p)}}$ converges in distribution to a standard normal random variable with *cdf* Φ as m, or equivalently n, converges to infinity. So if m is large enough, for every $y \in V_3^S$,

$$\mathbb{P}(y \in B_1^S) \ge \frac{3}{4}\Phi(\frac{1}{\sqrt{p(1-p)}}) \quad \text{and} \quad \mathbb{P}(y \in C_2^S) \ge \frac{3}{4}\Phi(\frac{1}{\sqrt{p(1-p)}}). \tag{26}$$

Next, we show that *w.e.p.* $|B_1^S \cap C_2^S|$ has a suitable lower bound. Let $a_p = \Phi(\frac{1}{\sqrt{p(1-p)}})$. Since $V_1^T \cap V_2^T = \emptyset$, the events $\{y \in B_1^S\}$ and $\{y \in C_2^S\}$ are independent, so (26) implies that $\mathbb{P}(y \in B_1^S \cap C_2^S) \geq (\frac{3}{4}a_p)^2$. Note that $|B_1^S \cap C_2^S|$ is a sum of Bernoulli random variables with probability of success at least $(\frac{3}{4}a_p)^2$, so $\mathbb{E}(|B_1^S \cap C_2^S|) \geq \frac{9}{16}a_p^2|V_3^S|$. From the condition $V_3^S \cap (V_1^T \cup V_2^T) = \emptyset$ in (24), these Bernoulli random variables are independent as well, so by Hoeffding's inequality, we have that, with probability at least $1 - 2\exp(-2(a_p^2/16)^2|V_3^S|)$, the size $|B_1^S \cap C_2^S|$ is bounded below by $(1/2)a_p^2|V_3^S| \geq (1/2)a_p^2(\frac{3}{4}s - 2\epsilon)n$. Employing a similar argument for $|B_2^S \cap C_1^S|$ and using a union bound, we obtain that if (23) is satisfied, then with probability at least $1 - 4\exp(-2(a_p^2/16)^2(\frac{3}{4}s - 2\epsilon)n)$, the set sizes $|B_1^S \cap C_2^S|$ and $|B_2^S \cap C_1^S|$ are bounded below by $(1/2)a_p^2(\frac{3}{4}s - 2\epsilon)n$.

To show that these lower bounds hold *w.e.p.* for *every* choice of V_i^S, V_i^T, $i = 1, 2, 3$, satisfying (24), we count the number of all such partitions. The number of ways to choose V_1^S and V_2^S of size $\lfloor \epsilon n \rfloor$ is $\binom{|V_S|}{\lfloor \epsilon n \rfloor}\binom{|V_S| - \lfloor \epsilon n \rfloor}{\lfloor \epsilon n \rfloor}$; recall that V_3^S is determined by V_1^S and V_2^S. It is known, as a consequence of Stirling's approximation of the factorial, that $\lim_{n\to\infty} \frac{1}{n}\log_2\binom{n}{m} = H(\epsilon)$, if $\lim_{n\to\infty}\frac{m}{n} = \epsilon$ (see [Spe14]). Here, H is the binary entropy function defined as $H(\epsilon) = -\epsilon\log_2\epsilon - (1-\epsilon)\log_2(1-\epsilon)$. Given that H is increasing on $[0, 1/2]$, we obtain the following upper bound on the number of ways to choose V_i^S, V_i^T:

$$\binom{|V_S|}{\lfloor \epsilon n \rfloor}\binom{|V_S| - \lfloor \epsilon n \rfloor}{\lfloor \epsilon n \rfloor}\binom{|V_T|}{\lfloor \epsilon n \rfloor}\binom{|V_T| - \lfloor \epsilon n \rfloor}{\lfloor \epsilon n \rfloor} \leq \binom{n}{\lfloor \epsilon n \rfloor}^4 \leq 2^{5H(\epsilon)n}$$

Define

$$f_p(x) = 2(a_p^2/16)^2(\frac{3}{4}s - 2x) - 5\ln(2)H(x).$$

Note that f_p is strictly decreasing on $[0, 1/4)$, and has positive value at $x = 0$. Let ϵ^* be the value of x in $[0, s/4)$ where f attains zero, or $\epsilon^* = s/4$ if no such value exists. We have now established the following claim:

Claim 6.5. *For each $0 < \epsilon < \epsilon^*$, with probability at least $1 - 2\exp(-2(s/4)^2n) - 4\exp(-f_p(\epsilon)n)$, V_S and V_T both have size at least $(3/4)sn$, and for* every *choice of $V_1^S, V_2^S, V_1^T, V_2^T$ satisfying Condition (24), $|B_1^S \cap C_2^S|$ and $|B_2^S \cap C_1^S|$ are bounded below by $(1/2)a_p^2(\frac{3}{4}s - 2\epsilon)n$.*

Now consider any ordering $\prec$ of V, and let V_1^S be the first $m = \lfloor \epsilon n \rfloor$ elements of V_S according to $\prec$, and V_2^S the next m elements of V_S. Similarly, let V_1^T and V_2^T be the first m elements and the next m elements of V_T. By Lemma 6.3, it follows that $(V_1^S \cup V_2^S) \prec V_3^T$ or $(V_1^T \cup V_2^T) \prec V_3^S$. By switching S and T if necessary, we may assume *wlog* that $(V_1^T \cup V_2^T) \prec V_3^S$. In particular, this implies that $(V_1^T \cup V_2^T) \cap V_3^S = \emptyset$, so Condition (24) is satisfied.

We now have that, for $i = 1, 2$,

$$\begin{aligned}
n^3\Gamma(G, \prec, V_i^T) &\geq \sum_{x,y \in V_3^S,\, x \prec y} \left[|N(y) \cap V_i^T| - |N(x) \cap V_i^T|\right]_+ \\
&\geq \sum_{x \in C_i^S,\, y \in B_i^S,\, x \prec y} \left[|N(y) \cap V_i^T| - |N(x) \cap V_i^T|\right]_+ \\
&\geq 2\sqrt{m}\,\left|\left\{(x, y) \in C_i^S \times B_i^S :\ x \prec y\right\}\right|. \qquad (27)
\end{aligned}$$

Note that any pair (x, y) with $x \in B_2^S \cap C_1^S$ and $y \in B_1^S \cap C_2^S$ belongs either to the set described in (27) with $i = 1$ (if $x \prec y$), or to the same set with $i = 2$ (if $y \prec x$). So

$$n^3\Gamma(G, \prec, V_1^T) + n^3\Gamma(G, \prec, V_2^T) \geq 2\sqrt{m}\,|B_2^S \cap C_1^S|\,|B_1^S \cap C_2^S|,$$

which implies that $2\Gamma(G, \prec) \geq \Gamma(G, \prec, V_1^T) + \Gamma(G, \prec, V_2^T) \geq \frac{2\sqrt{m}}{n^3}\,|B_2^S \cap C_1^S|\,|B_1^S \cap C_2^S|$.

Taking $\epsilon = \epsilon^*/2$, $\alpha = (\ln 6)\left(\min\{\frac{s^2}{8}, f(\epsilon^*/2)\}\right)$ and $c = (1/5)a_p^4(\frac{3}{4}s - 2\epsilon)^2\sqrt{\epsilon}$, and using Claim 6.5, we get that with probability at least $1 - \exp(-\alpha n)$, we have $\Gamma^*(G, \prec) \geq cn^{-1/2}$ for all orderings $\prec$. $\square$

6.3. Steep graphons

The next proposition gives an upper bound on Γ^* for graphs sampled from a steep graphon, and thus covers case (ii) of Theorem 1.3. Steep graphons are graphons for which the partial derivatives exist and are bounded away from zero.

Since we are establishing an upper bound on Γ^*, we only need to show the upper bound for $\Gamma^*(G, \prec)$ for some ordering $\prec$. Guided by the Robinson property, we use the natural ordering of vertices induced from the labels assigned by the random process $\mathbb{G}(n,w)$. We formalize this concept as follows: For $G \sim \mathbb{G}(n,w)$, we denote the vertex set of G by $V = \{1,\ldots,n\}$, and identify each vertex i with the *label* $x_i \in [0,1]$, which is its sampled value. We assume that the vertices are ordered so that $x_1 \leq x_2 \leq \cdots \leq x_n$. The graph G, labeled in this manner, is called a *labeled outcome* of $\mathbb{G}(n,w)$. Since the labels assigned to different vertices are almost surely distinct, we can assume that the vertex ordering mentioned above is unique.

The rather complicated nature of our proof is due to the fact that $\Gamma^*(G,\prec)$ is defined as the maximum of $\Gamma^*(G,\prec,A)$ over all possible choices of A. Thus, we must prove that the proposed upper bound dominates $\Gamma^*(G,\prec,A)$ for *every* subset A of $V(G)$. To do so, we proceed by partitioning the interval $[0,1]$ into a large number of small intervals. We then show that the following two facts hold with exponential probability: on the one hand, pairs of vertices chosen from intervals that are far apart do not have a positive contribution to Γ^*, as w is a steep Robinson graphon; on the other hand, the contribution to Γ^* coming from pairs of vertices chosen from intervals that are close together can be bounded from above efficiently.

PROPOSITION 6.6. *Let $w : [0,1]^2 \to (0,1)$ be a Robinson graphon, whose partial derivatives on $[0,1]^2$ exist and are bounded away from 0. That is, there exists a positive constant c, such that*

$$\Big|\frac{\partial w}{\partial x}\Big|, \Big|\frac{\partial w}{\partial y}\Big| \geq c. \tag{28}$$

Then, for an outcome $G \sim \mathbb{G}(n,w)$,

$$\mathbb{P}\left(\Gamma^*(G) \leq \left(\frac{26}{c}\right) n^{-2/3}\right) \geq 1 - \exp(-\Omega(\log^2 n)).$$

Proof. Let $m = \phi(n)$, where $\phi(n)$ is an integer-valued function so that $\lim_{n\to\infty} \frac{\phi(n)}{\log^3(n)} > 1$ and $\lim_{n\to\infty} \frac{\phi(n)\log n}{\sqrt{n}} = 0$. An appropriate choice for a function ϕ, satisfying both of these properties, becomes apparent at the end of the proof. Assume n is large enough so that

$$\log^3(n) < \phi(n) < 0.1\sqrt{n}\log^{-1} n. \tag{29}$$

Next, divide $(0,1]$ into m equal-sized intervals; namely $I_j = (\frac{j}{m}, \frac{j+1}{m}]$, for $0 \leq j < m$.

Let $G \sim \mathbb{G}(n,w)$ be a labeled outcome, i.e. $V(G)$ is labeled so that $0 < x_1 < x_2 < \ldots < x_n < 1$. Throughout the proof, we identify a vertex i in V with its label x_i in $(0,1)$, which allows us to think of V both as a set of vertices and as a subset of $(0,1)$.

The partition $\{I_j\}_{0\leq j<m}$ of $(0,1]$ results in an analogous partition of the vertex set V, which we denote by $\{V_j\}_{0\leq j<m}$. Namely, for every $0 \leq j < m$, $V_j = \{i \in V : x_i \in I_j\}$. We will see that *w.e.p.*, these sets will all be close to their expected size. Precisely, fix $0 \leq j < m$, and note that $|V_j|$ is the sum of n independent Bernoulli variables with success probability $|I_j|$. Applying Hoeffding's inequality with parameter $\alpha = \frac{\log n}{\sqrt{n}}$, we conclude that, with probability at least $1 - 2\exp(-2\log^2 n)$, we have

$$\big||V_j| - n|I_j|\big| \leq \sqrt{n}\log n. \tag{30}$$

Since the number of intervals I_j is sub-linear in n, we can assume that *w.e.p.*, (30) holds for every $0 \leq j < m$. Because of the fact that $\sqrt{n}\log n \leq 0.1\frac{n}{\phi(n)}$ (from (29)), together with $|I_j| = \frac{1}{\phi(n)}$, the inequality (30) implies that $(0.9)\frac{n}{\phi(n)} \leq |V_j| \leq (1.1)\frac{n}{\phi(n)}$. Thus, the event

$$\mathcal{B}_1 = \bigcap_{0\leq j<m} \left\{(0.9)\frac{n}{\phi(n)} \leq |V_j| \leq (1.1)\frac{n}{\phi(n)}\right\} \tag{31}$$

holds *w.e.p.*

Recall that the definition of Γ^* consists of two parts, which we call Γ^*_ℓ and Γ^*_u. We will only bound Γ^*_ℓ; an identical bound applies to Γ^*_u. Given a set $A \subseteq V$, let $A_k = A \cap V_k$, for $k = 0, \ldots, m-1$, and view each A_k as a subset of $[0,1]$. Note that $[0,x) \cap A_k = A_k$ when $x \in V_i$ and $k < i$. So we get,

$$\begin{aligned}
n^3\Gamma^*_\ell(G,A,\prec) &= \sum_{x,y\in V,\ x<y} \left[\sum_{k=0}^{m-1} |N(y)\cap[0,x)\cap A_k| - |N(x)\cap[0,x)\cap A_k|\right]_+ \\
&\leq \sum_{x,y\in V,\ x<y} \sum_{k=0}^{m-1} \Big[|N(y)\cap[0,x)\cap A_k| - |N(x)\cap[0,x)\cap A_k|\Big]_+ \\
&= \sum_{0\leq k<i\leq j<m} \sum_{\substack{x\in V_i\\ y\in V_j\\ x<y}} \Big[|N(y)\cap A_k| - |N(x)\cap A_k|\Big]_+ \qquad (32)\\
&+ \sum_{0\leq k=i\leq j<m} \sum_{\substack{x\in V_i\\ y\in V_j\\ x<y}} \Big[|N(y)\cap[0,x)\cap A_k| - |N(x)\cap[0,x)\cap A_k|\Big]_+. \qquad (33)
\end{aligned}$$

To obtain an upper bound for the above sum, we need to prove a few concentration results on random variables of the form $|N(x)\cap S|$. The graph G is determined by two sets of random variables: firstly the random variables x_i with values in $[0,1]$ which are assigned as labels to the vertices of G, and secondly the binary variables $Y_{x,z}$, where $Y_{x,z} = 1$ precisely when the pair of vertices labeled as x and z form an edge in G. According to the definition of $\mathbb{G}(n,w)$, the random variables $Y_{x,z}$ are independent Bernoulli variables with probability of success $w(x,z)$. Event $\mathcal{B}_1$ defined in (31) is only a function of the labels. From now on, we assume that this event occurs.

Recall that V is ordered so that the labeling $x_1,\ldots,x_n$ is increasing. For a vertex $x \in V$ and a subset $S \subseteq V\setminus\{x\}$, define the following random variable.

$$\delta_{x_1,\ldots,x_n}(x,S) = \sum_{s\in S} Y_{s,x} - \sum_{s\in S} \mathbb{E}\, Y_{s,x} = |N(x)\cap S| - \mathbb{E}(|N(x)\cap S|).$$

Note that $\mathbb{E}(|N(x)\cap S|) = \sum_{z\in S} w(x,z)$. We will now show how the definition of $\delta_{x_1,\ldots,x_n}$ allows us to bound $n^3\Gamma^*_\ell(G,A,\prec)$ further. To simplify notation, we denote $\delta_{x_1,\ldots,x_n}(x,S)$ by $\delta(x,S)$, when the assignment $\{x_i\}_{i=1}^n$ is understood.

Suppose $x\in V_i$, $y \in V_j$, $x<y$, and $k<i\leq j$. Since $\mathbb{E}|N(x)\cap A_k| - \mathbb{E}|N(y)\cap A_k| = \sum_{z\in A_k} w(x,z) - w(y,z)$, we have

$$\Big[|N(y)\cap A_k| - |N(x)\cap A_k|\Big]_+ = \Big[\delta(y,A_k) - \delta(x,A_k) - \sum_{z\in A_k}(w(x,z)-w(y,z))\Big]_+.$$

Given that $z<x<y$ and w is Robinson, every summand $w(x,z)-w(y,z)$ in the above sum is non-negative. Moreover, by the Mean Value Theorem and condition (28), we have $w(x,z)-w(y,z) \geq$

$c(y-x)$ for every $z \in A_k$. Thus,

$$\begin{aligned}
\Big[|N(y) \cap A_k| - |N(x) \cap A_k|\Big]_+ &= \Big[\delta(y, A_k) - \delta(x, A_k) - \sum_{z \in A_k} (w(x,z) - w(y,z))\Big]_+ \\
&\le \Big[\delta(y, A_k) - \delta(x, A_k) - c|A_k|(y-x)\Big]_+ \\
&\le \Big[|\delta(y, A_k)| - \frac{c|A_k|(y-x)}{2}\Big]_+ \\
&\qquad + \Big[|\delta(x, A_k)| - \frac{c|A_k|(y-x)}{2}\Big]_+ . \qquad (34)
\end{aligned}$$

Since $\mathbb{E}(\delta(x,S)) = 0$ for each set S, an upper bound will depend on concentration results for sums of random variables of type $\delta(x,S)$ or $\delta^2(x,S)$. The following three claims will establish the necessary bounds.

Claim 6.7. *For a labeled outcome $G \sim \mathbb{G}(n,w)$ with labeling $x_1, \dots, x_n$ assigned to $V(G)$, define the event*

$$\mathcal{B}_2 = \bigcap_{0 \le k < j < m} \bigcap_{S \subseteq V_k} \Big\{ \sum_{x \in V_j} \delta^2_{x_1,\dots,x_n}(x,S) < 7\frac{n}{\phi(n)}|S| \Big\}. \qquad (35)$$

For n large enough so that (29) holds, we have

$$\mathbb{P}(\mathcal{B}_2 \,|\, \mathcal{B}_1) \ge 1 - \exp\left(-\frac{0.0025\,n}{\phi(n)}\right).$$

That is, assuming $\mathcal{B}_1$, w.e.p. for every $0 \le k < j < m$ and subset $S \subseteq V_k$, $\sum_{x \in V_j} \delta^2(x,S) < 7\frac{n}{\phi(n)}|S|$.

Proof of claim. Fix an outcome $G \sim \mathbb{G}(n,w)$ with labeling $x_1, \dots, x_n$ for V. Fix $k < j$, a subset $S \subset V_k$ and a vertex $x \in V_j$. Since the random variables $Y_{s,x}$, $s \in S$, are independent, by Hoeffding's inequality (21), we have for all $t > 0$,

$$\mathbb{P}(\delta(x,S) \ge t) \le \exp(-2t^2/|S|) \quad \text{and} \quad \mathbb{P}(\delta(x,S) \le -t) \le \exp(-2t^2/|S|).$$

This means that $\delta(x,S)$ satisfies the conditions of the proposition in Appendix A.7 with $\sigma^2 = \frac{|S|}{4}$, and $\delta^2(x,S) - \mathbb{E}(\delta^2(x,S))$ is a sub-exponential random variable with parameters $\nu = 2\sqrt{2}|S|, \alpha = 2|S|$. Moreover, $\{\delta^2(x,S) - \mathbb{E}(\delta^2(x,S))\}_{x \in V_j}$ are independent random variables, as $V_j \cap S = \emptyset$. Thus, by the tail bound on sums of sub-exponential variables as in Theorem Appendix A.6, part (ii), we have

$$\mathbb{P}\Big(\sum_{x \in V_j} \delta^2(x,S) - \sum_{x \in V_j} \mathbb{E}(\delta^2(x,S)) < t|V_j| \Big) \ge 1 - \exp\big(- \frac{|V_j|}{2} \min\big\{ \frac{t^2}{8|S|^2}, \frac{t}{2|S|} \big\}\big).$$

Taking $t = \beta|S|$ with $\beta \ge 4$, the above inequality implies that, with probability at least $1 - \exp(-\frac{\beta}{4}|V_j|)$,

$$\sum_{x \in V_j} \delta^2(x,S) - \sum_{x \in V_j} \mathbb{E}(\delta^2(x,S)) < \beta|V_j|\,|S|. \qquad (36)$$

There are at most $m^2 2^{|V_k|}$ choices for j, k, and S. Assuming $\mathcal{B}_1$ holds, and using conditions (29) and (31), we have that $m^2 \le n \le \exp(0.1\frac{\sqrt{n}}{\phi(n)}) \le \exp(\frac{1}{9}|V_k|) \le \exp((1 - \ln 2)|V_k|)$. Using (31) again, we get:

$$m^2 2^{|V_k|} \exp(-\frac{\beta}{4}|V_j|) \le \exp(|V_k|)\exp(-\frac{\beta}{4}|V_j|) \le \exp\left(-(\frac{\beta}{4}(0.9) - 1.1)(\frac{n}{\phi(n)}) \right)$$

is exponentially small if $\beta > \frac{4(1.1)}{0.9}$. Taking $\beta = 4.9$ satisfies this condition.

Finally, note that $\mathbb{E}(\delta^2(x,S)) = \sum_{y \in S} \mathrm{Var}(Y_{x,y}) = \sum_{y \in S} w(x,y) - w(x,y)^2 \le |S|$. Therefore, $\sum_{x \in V_j} \mathbb{E}(\delta^2(x,S)) \le \frac{1.1n|S|}{\phi(n)}$, and this finishes the proof. ∎

To bound summand (33), we need concentration results for a new type of sets defined as follows. Given a set $A \subseteq V$, an index k and a vertex $x \in V_k$, let $A_k = A \cap V_k$ and $A_{k,x} = A \cap V_k \cap [0, x)$.

Claim 6.8. *For a labeled outcome $G \sim \mathbb{G}(n, w)$ with labeling $x_1, \dots, x_n$ assigned to the vertex set V, define the event*

$$\mathcal{B}_3 = \bigcap_{0 \le i < m} \bigcap_{T, A_i \subseteq V_i} \Big\{ \big| \sum_{x \in T} \delta_{x_1,\dots,x_n}(x, A_{i,x}) \big| < \left(\frac{n}{\phi(n)} \right)^{\frac{3}{2}} \Big\}.$$

For n large enough so that (29) holds, we have

$$\mathbb{P}(\mathcal{B}_3 \,|\, \mathcal{B}_1) \ge 1 - 2 \exp\left(-0.2 \frac{n}{\phi(n)} \right).$$

That is, assuming $\mathcal{B}_1$, w.e.p. $\mathcal{B}_3$ holds.

Proof of claim. Fix an index $0 \le i < m$, two sets $A_i, T \subseteq V_i$, and a labeling $x_1, \dots, x_n$ for V. As before, we denote $\delta_{x_1,\dots,x_n}(x, S)$ by $\delta(x, S)$. Recall that

$$\sum_{x \in T} \delta(x, A_{i,x}) = \sum_{x \in T, z \in A_{i,x}} (Y_{x,z} - w(x, z)),$$

where $Y_{x,z}$ denotes the indicator variable of $z \in N(x)$ with probability of success $w(x, z)$. A random variable $Y_{x,z}$ contributes to the above sum, only if $x \in T \subseteq V_i$, $z \in A_{i,x} \subseteq V_i$ and $z < x$. Therefore, $\sum_{x \in T} \delta(x, A_{i,x})$ is the sum of at most $|V_i|^2/2$ independent variables, each with expected value zero. Applying Hoeffding's inequality, we get

$$\mathbb{P}\left(\big| \sum_{x \in T} \delta(x, A_{i,x}) \big| \ge t \right) \le 2 \exp\left(\frac{-2t^2}{|V_i|^2/2} \right). \tag{37}$$

Let $t = \frac{3}{4}|V_i|^{3/2}$, so the above bound on the probability is exponentially small. Assume that $\mathcal{B}_1$ holds, and n satisfies (29). Note that there are at most $m2^{2|V_i|} \le \exp(2|V_i|)$ choices for i, A_i and T. So, applying a union bound to (37), we get

$$\mathbb{P}\Big(\bigcup_{0 \le i < m} \bigcup_{T, A_i \subseteq V_i} \Big\{ \big| \sum_{x \in T} \delta(x, A_{i,x}) \big| \ge \frac{3}{4} |V_i|^{3/2} \Big\} \,\Big|\, \mathcal{B}_1 \Big) \le 2 \exp(-\frac{9}{4}|V_i| + 2|V_i|) \le 2 \exp(-0.2 \frac{n}{\phi(n)}).$$

Since $\frac{3}{4}|V_i|^{3/2} \le \frac{3}{4}(1.1)^{3/2} \left(\frac{n}{\phi(n)} \right)^{3/2} < \left(\frac{n}{\phi(n)} \right)^{3/2}$, the result follows. $\square$

Finally, we show that with exponential probability, certain sums of variables $\delta(y, S)$ can be bounded from below.

Claim 6.9. *For a labeled outcome $G \sim \mathbb{G}(n, w)$ with labeling $x_1, \dots, x_n$ assigned to the vertex set V, define the event*

$$\mathcal{B}_4 = \bigcap_{0 \le i < m} \bigcap_{S \subseteq V_i} \bigcap_{T \subseteq V \setminus S} \Big\{ \sum_{y \in T} \delta_{x_1,\dots,x_n}(y, S) < \frac{1.1 n^{3/2}}{\sqrt{\phi(n)}} \Big\}.$$

For n large enough so that (29) holds, we have

$$\mathbb{P}(\mathcal{B}_4 \,|\, \mathcal{B}_1) \ge 1 - \exp(-0.1n).$$

That is, assuming $\mathcal{B}_1$, w.e.p. $\mathcal{B}_4$ holds.

Proof of claim. Fix $0 \leq i < m$, $S \subseteq V_i$, $T \subseteq V \setminus S$, and a labeling $x_1, \ldots, x_n$ for V. Let $\delta(y, S)$ denote $\delta_{x_1,\ldots,x_n}(y, S)$ for short. Similar to the proof of Claim 6.8, we can write $\sum_{y \in T} \delta(y, S)$ as a sum of $|T||S|$ independent random variables. Suppose $x_1, \ldots, x_n$ are so that $\mathcal{B}_1$ holds. Then, $|T||S| \leq \frac{1.1n^2}{\phi(n)}$, and applying Hoeffding's inequality we get

$$\mathbb{P}\left(\sum_{y \in T} \delta(y, S) < \frac{1.1n^{3/2}}{\sqrt{\phi(n)}}\right) \geq 1 - \exp\left(\frac{-2.42\, n^3}{\phi(n)|T||S|}\right) \geq 1 - \exp(-2.2n).$$

The result will then follow by applying a union bound, as there are at most $\phi(n)2^{\frac{1.1n}{\phi(n)}+n}$ choices for i, S and T, and $\phi(n)2^{\frac{1.1n}{\phi(n)}+n}\exp(-2.2n) \leq \exp(-0.1n)$. ∎

We have now shown that the probability that $\mathcal{B}_1$ does not hold is exponentially small, and, if $\mathcal{B}_1$ holds, then the probability that any of $\mathcal{B}_2$, $\mathcal{B}_3$ or $\mathcal{B}_4$ does not hold is also exponentially small. Using a union bound, we can then conclude that *w.e.p.* $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ and $\mathcal{B}_4$ all hold. We will assume in the rest of the proof that this is the case.

Bounding summand (32): The assumption that $\mathcal{B}_1$ $\mathcal{B}_2$, $\mathcal{B}_3$ and $\mathcal{B}_4$ all hold allows us to bound $\Gamma^*_\ell(G, A, \prec)$ as required. First, we bound the summand (32). Suppose $x \in V_i$, $y \in V_j$, $x < y$, and $k < i \leq j$. As we saw earlier in (34),

$$\begin{aligned}\Big[|N(y) \cap A_k| - |N(x) \cap A_k|\Big]_+ \quad &\leq \quad \left[|\delta(y, A_k)| - \frac{c|A_k|(y-x)}{2}\right]_+ &(38)\\ &+ \quad \left[|\delta(x, A_k)| - \frac{c|A_k|(y-x)}{2}\right]_+. &(39)\end{aligned}$$

Since $x \in V_i$ and $y \in V_j$, and $i \leq j$, we have

$$\frac{j-i-1}{\phi(n)} \leq y - x \leq \frac{j+1-i}{\phi(n)}. \tag{40}$$

Note that if $y - x \geq \frac{2|\delta(y,A_k)|}{c|A_k|}$ then summand (38) will attain zero. Using the lower bound on $y - x$, we can see that summand (38) is nonzero only if $j - 1 - \frac{2\phi(n)|\delta(y,A_k)|}{c|A_k|} < i$. Fix $y \in V_j$. If $i \in \left(j - 1 - \frac{2\phi(n)|\delta(y,A_k)|}{c|A_k|}, j - \frac{2\phi(n)|\delta(y,A_k)|}{c|A_k|}\right]$, the term $\left[|\delta(y, A_k)| - \frac{c|A_k|(y-x)}{2}\right]_+$ in summand (38) is at most $\frac{c|A_k|}{2\phi(n)}$ for each $x \in V_i$. There is exactly one such value of i and this value gives a total contribution of at most $\frac{c|V_i|\,|A_k|}{2\phi(n)} \leq \frac{c(1.1)^2n^2}{2\phi(n)^3}$ to (41). On the other hand, there are at most $\frac{2\phi(n)|\delta(y,A_k)|}{c|A_k|}$ values of i such that $j - \frac{2\phi(n)|\delta(y,A_k)|}{c|A_k|} < i \leq j$.

Similarly, if $y - x \geq \frac{2|\delta(x,A_k)|}{c|A_k|}$ then summand (39) will attain zero. Using the upper bound from (40) we see that, given $x \in V_i$, there are at most $\frac{2\phi(n)|\delta(x,A_k)|}{c|A_k|} + 1$ choices for j so that summand (39) can possibly have nonzero values for $y \in V_j$, and one of those values contributes at most $\frac{c(1.1)^2n^2}{2\phi(n)^3}$ to (41).

Finally, since $\left[|\delta(y, A_k)| - \frac{c|A_k|(y-x)}{2}\right]_+ \leq |\delta(y, A_k)|$ is valid for every $x \leq y$, and $|V_i| \leq \frac{1.1\,n}{\phi(n)}$ for every i, we have

$$\begin{aligned}
\text{Summand (32)} &\leq \sum_{\substack{0 \leq i \leq j < m \\ x \in V_i, y \in V_j \\ x < y \\ k < i}} \left[|\delta(y, A_k)| - \frac{c|A_k|(y-x)}{2}\right]_+ + \left[|\delta(x, A_k)| - \frac{c|A_k|(y-x)}{2}\right]_+ \\
&\leq 2 \sum_{\substack{0 \leq j < m \\ y \in V_j \\ k < j}} \left(\left(\frac{2\phi(n)|\delta(y, A_k)|}{c|A_k|}\right)\left(\frac{1.1\, n}{\phi(n)}\right)|\delta(y, A_k)| + \frac{c(1.1)^2}{2}\left(\frac{n^2}{\phi(n)^3}\right)\right) \quad (41) \\
&\leq \left(\frac{4.4\, n}{c} \sum_{\substack{0 \leq j < m \\ y \in V_j \\ k < j}} \frac{\delta(y, A_k)^2}{|A_k|}\right) + c(1.4)\left(\frac{n^3}{\phi(n)^2}\right). \quad (42)
\end{aligned}$$

Using our assumption that $\mathcal{B}_2$ holds, and the fact that there are only $m = \phi(n)$ many choices for j and k, we get

$$\text{Summand (42)} \leq \phi(n)^2 \left(\frac{4.4\, n}{c}\right)\left(\frac{7\, n}{\phi(n)}\right) + \left(\frac{c(1.4)n^3}{\phi(n)^2}\right) = \left(\frac{30.8}{c}\right)\phi(n)n^2 + c(1.4)\left(\frac{n^3}{\phi(n)^2}\right). \quad (43)$$

Bounding summand (33): Using our assumption that $\mathcal{B}_1$, $\mathcal{B}_3$ and $\mathcal{B}_4$ hold, we now bound summand (33). For $x \in V_i$, recall the notation $A_{i,x} = [0, x) \cap A_i$. Similar to the argument leading to (34), we have

$$\left[|N(y) \cap A_{i,x}| - |N(x) \cap A_{i,x}|\right]_+ \quad \leq \quad \left[\delta(y, A_{i,x}) - \delta(x, A_{i,x}) - c|A_{i,x}|(y-x)\right]_+.$$

Since $[\cdot]_+$ is sub-additive, we get

$$\begin{aligned}
\text{Summand (33)} &\leq \sum_{\substack{0 \leq i < m \\ x \in V_i}} \sum_{\substack{i \leq j < m \\ y \in V_j \\ x < y}} \left[-\delta(x, A_{i,x}) - \frac{c|A_{i,x}|(y-x)}{2}\right]_+ \quad (44) \\
&+ \sum_{\substack{0 \leq i < m \\ x \in V_i}} \sum_{\substack{y \in V \\ x < y}} \left[\delta(y, A_{i,x}) - \frac{c|A_{i,x}|(y-x)}{2}\right]_+. \quad (45)
\end{aligned}$$

For a fixed $x \in V_i$, a nonzero contribution of $y \in V_j$ in summand (44) is possible only if $\delta(x, A_{i,x}) < 0$ and $\frac{j-i-1}{\phi(n)} < \frac{-2\delta(x, A_{i,x})}{c|A_{i,x}|}$, as $\frac{j-i-1}{\phi(n)}$ is a lower bound for $y - x$. Therefore, there are at most $\frac{2\phi(n)|\delta(x, A_{i,x})|}{c|A_{i,x}|}$ choices for $j \geq i$ with nonzero contribution in (44). Let V_i^- denote the set $\{x \in V_i : \delta(x, A_{i,x}) < 0\}$.

Since for every j, $|V_j|$ is bounded above by $\frac{1.1\,n}{\phi(n)}$, we have

$$\begin{aligned}
\text{Summand (44)} &= \sum_{\substack{0 \le i < m \\ x \in V_i}} \sum_{\substack{i \le j < m \\ y \in V_j \\ x < y}} \left[-\delta(x, A_{i,x}) - \frac{c|A_{i,x}|(y-x)}{2}\right]_+ \\
&\le \sum_{\substack{0 \le i < m \\ x \in V_i^-}} \Big(\frac{2\phi(n)|\delta(x, A_{i,x})|}{c|A_{i,x}|}\Big)\Big(\frac{1.1\,n}{\phi(n)}\Big)|\delta(x, A_{i,x})| \\
&= \frac{2.2\,n}{c} \sum_{\substack{0 \le i < m \\ x \in V_i^-}} \frac{\delta(x, A_{i,x})^2}{|A_{i,x}|}.
\end{aligned}$$

By definition, it is clear that $|\delta(x, A_{i,x})| \le |A_{i,x}|$. So, we have

$$\begin{aligned}
\text{Summand (44)} &\le \frac{2.2\,n}{c} \sum_{0 \le i < m} \sum_{x \in V_i^-} -\delta(x, A_{i,x}) = \frac{2.2\,n}{c} \sum_{0 \le i < m} \Big| \sum_{x \in V_i^-} \delta(x, A_{i,x})\Big| \\
&\le \frac{2.2\,n}{c} \phi(n) \Big(\frac{n^{\frac{3}{2}}}{\phi(n)^{\frac{3}{2}}}\Big) = \frac{2.2}{c} \Big(\frac{n^{\frac{5}{2}}}{\sqrt{\phi(n)}}\Big),
\end{aligned} \tag{46}$$

where the last inequality comes from our assumption that $\mathcal{B}_3$ holds. This bounds summand (44). To bound summand (45), we need to take a different approach, as the distance between x and y in this case, cannot be bounded only in terms of y. So, we proceed as follows. For each x, define $T_x = \{y \in V : y > x \text{ and } \delta(y, A_{i,x}) > 0\}$. Note that

$$\begin{aligned}
\text{Summand (45)} &= \sum_{\substack{0 \le i < m \\ x \in V_i}} \sum_{\substack{y \in V \\ x < y}} \left[\delta(y, A_{i,x}) - \frac{c|A_{i,x}|(y-x)}{2}\right]_+ \\
&\le \sum_{\substack{0 \le i < m \\ x \in V_i}} \sum_{y \in T_x} \delta(y, A_{i,x}).
\end{aligned}$$

Given the assumption that $\mathcal{B}_4$ from Claim 6.9 holds, we have

$$\text{Summand (45)} \le \sum_{\substack{0 \le i \le m-1 \\ x \in V_i}} \sum_{y \in T_x} \delta(y, A_{i,x}) \le n \left(\frac{1.1 n^{\frac{3}{2}}}{\sqrt{\phi(n)}}\right) = \frac{1.1 n^{\frac{5}{2}}}{\sqrt{\phi(n)}}. \tag{47}$$

Putting inequalities (43), (46) and (47) together, we conclude that

$$n^3 \Gamma^*_\ell(G, A, \prec) \le \left(\frac{30.8}{c}\right) \phi(n) n^2 + c(1.4) \left(\frac{n^3}{\phi(n)^2}\right) + \frac{2.2}{c} \left(\frac{n^{\frac{5}{2}}}{\sqrt{\phi(n)}}\right) + \frac{1.1\, n^{\frac{5}{2}}}{\sqrt{\phi(n)}}.$$

Thus, the best upper bound, in order, will be obtained when $\phi(n) = \theta(n^{1/3})$, so that $\phi(n)n^2$, $n^3/\phi(n)^2$ and $\frac{n^{\frac{5}{2}}}{\sqrt{\phi(n)}}$ are of the same order. From the conditions on w and the fact that $w \ge 0$, we have that $c \le 1$. Using this, and taking $\phi(n) = \frac{1}{2} n^{\frac{1}{3}}$, we have

$$n^3 \Gamma^*_\ell(G, A, \prec) \le \left(\frac{30.8}{2} + 4(1.4) + \sqrt{2}(2.2 + 1.1)\right) \left(\frac{1}{c}\right) n^{7/3} \le \left(\frac{26}{c}\right) n^{7/3}.$$

□

7. Conclusions and further work

We have established that the graphon parameter Γ, first defined in [CGH+15], is indeed a suitable gauge for measuring the extent to which a graphon lacks the Robinson property. It was already known that the function Γ attains zero precisely when applied to a Robinson graphon. In this paper, we show that for any graphon w, there is a Robinson graphon R_w so that $\|R_w - w\|_\square \leq 14\Gamma(w)^{1/7}$. Thus, any graphon with small Γ-value is close, in cut norm, to a Robinson graphon. As a corollary of this approximation result, we prove that Γ detects graph sequences sampled from Robinson graphons. That is, for a convergent graph sequence $\{G_n\}_{n\in\mathbb{N}}$, we have $\Gamma(G_n) \to 0$ *iff* the limit object has a Robinson representative. In this paper, we have focused on proving the existence of the Robinson approximation R_w and the corresponding bound $\Gamma(w)^{1/7}$. We believe that a worthwhile direction of further study is to improve the exponent on the bound.

Computing, or even approximating, the Robinson approximation R_w is another important topic of investigation, which is beyond the scope of this work. The definition of R_w involves calculating the supremum of the average of w over sets of certain shape and size. Even for graphons corresponding to matrices, it is not clear to us how R_w can be efficiently computed. We believe there may be efficient algorithms to approximate R_w, and we find this topic worthy of further study.

Our results show that for a converging sequence of graphs, the equivalence class of the limiting graphon contains a Robinson graphon *iff* the sequence of Γ-values of the graphs converge to zero. However, to compute Γ of a sampled graph, one would have to come up with a labeling of the graph that approximates the Robinson ordering of the graphon. This relates to the problem of *graph seriation*, with important practical applications. We hope to use the theory developed in this paper as a stepping stone for making progress on the seriation problem for Robinson similarity data with errors.

Finally, we show that the convergence rates of Γ for sequences of random graphs sampled from a Robinson graphon depend on how strongly the graphon exhibits the Robinson property. The relevant result for steep graphons, only applies to graphons that are nowhere zero. Graphons that are zero outside a band around the main diagonal form a natural model for many applications. It would be worthwhile to extend this result to such graphons.

Acknowledgements

The first author acknowledges support from National Science Foundation Grant DMS-1902301 during the preparation of this article. The second author was funded by an NSERC Discovery Grant and a sabbatical grant from Dalhousie University. The authors completed this project while on a Research in Teams visit at the Banff International Research Center. We are grateful to BIRS for the financial support and hospitality. Finally, we sincerely thank the anonymous reviewers for critically reading the manuscript and suggesting numerous improvements.

Appendix A.

This appendix includes some of the definitions and concentration results that we need for Section 6. A beautiful exposition of this material can be found in [Wai19, Section 2.1].

Let X be a Gaussian random variable with mean μ and variance σ^2. It is easy to see that X has the moment generating function $\mathbb{E}(e^{\lambda X}) = e^{\mu\lambda + \frac{\sigma^2\lambda^2}{2}}$ valid for all $\lambda \in \mathbb{R}$. The following upper deviation inequality for X follows directly from Chernoff bound:

$$\mathbb{P}(X \geq \mu + t) \leq e^{-\frac{t^2}{2\sigma^2}}, \text{ for all } t \geq 0.$$

This deviation/concentration result can be generalized to a large class of non-Gaussian random variables, which mimic the behaviour of Gaussian random variables to some extend; such random variables are described in the following definition.

Definition Appendix A.1. A random variable X with mean μ is said to be *sub-Gaussian* if there exists $\sigma > 0$ such that for all $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}(e^{\lambda(X-\mu)}) \leq e^{\frac{\sigma^2\lambda^2}{2}}. \tag{A.1}$$

The constant σ satisfying the above equation is called the *sub-Gaussian parameter*, and σ^2 is called the *variance proxy*.

In this article, we use equivalent characterizations of sub-Gaussian random variables as listed below.

Theorem Appendix A.2 (Equivalent characterizations of sub-Gaussian random variables). *(Cf. [Wai19, Theorem 2.6]) Let X be a random variable with zero mean. Then the following are equivalent.*

(i) X is a sub-Gaussian random variable with parameter $\sigma \geq 0$, that is, for all $\lambda \in \mathbb{R}$ we have

$$\mathbb{E}(e^{\lambda X}) \leq e^{\frac{\sigma^2\lambda^2}{2}}.$$

(ii) There is a constant $c \geq 0$ and Gaussian random variable Z with zero mean and variance τ^2 such that

$$\mathbb{P}(|X| \geq s) \leq c\,\mathbb{P}(|Z| \geq s), \text{ for all } s \geq 0.$$

(iii) There is a constant $\theta \geq 0$ such that

$$\mathbb{E}(X^{2k}) \leq \frac{(2k)!}{2^k k!}\theta^{2k}, \text{ for all } k = 1, 2, \ldots.$$

Clearly, any Gaussian random variable with variance σ^2 is sub-Gaussian with parameter σ. More generally, any bounded random variable, supported in some interval $[a, b]$, is sub-Gaussian with parameter at most $\frac{b-a}{2}$ ([Wai19, Exercise 2.4]). An important concentration bound for sums of independent sub-Gaussian random variables is provided by Hoeffding's inequality. A special case of this inequality is stated in Theorem 6.2, where Hoeffding's bound is applied to the case of bounded random variables supported in $[0, 1]$.

Theorem Appendix A.3 (Hoeffding bound). *[Wai19, Proposition 2.5] Let $\{X_i\}_{i=1}^N$ be a collection of independent sub-Gaussian random variables with sub-Gaussian parameters σ_i. Then for all $t \geq 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}(X_i) \geq t\right) \leq \exp\big(-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2}\big).$$

An identical upper bound holds for the left-hand-side event $\big\{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}(X_i) \leq t\big\}$.

A relaxation of the notion of sub-Gaussian random variables leads to the following definition.

Definition Appendix A.4. A random variable X with mean μ is said to be *sub-exponential* if there exist non-negative parameters (ν, α) such that for all $\lambda \in (-\frac{1}{\alpha}, \frac{1}{\alpha})$, we have

$$\mathbb{E}(e^{\lambda(X-\mu)}) \leq e^{\frac{\nu^2\lambda^2}{2}}. \tag{A.2}$$

Clearly, any sub-Gaussian random variable is sub-exponential as well. However, many sub-exponential random variables are not sub-Gaussian. Some equivalent characterizations of sub-exponential random variables are listed in the following theorem.

THEOREM Appendix A.5 (Equivalent characterizations of sub-exponential random variables). *(Cf. [Wai19, Theorem 2.13]) Let X be a random variable with zero mean. Then the following are equivalent.*

(i) X is a sub-exponential random variable with parameters (ν, α), that is, for all $\lambda \in (-\frac{1}{\alpha}, \frac{1}{\alpha})$, we have

$$\mathbb{E}(e^{\lambda X}) \leq e^{\frac{\nu^2\lambda^2}{2}}.$$

(ii) There is a constant $c_0 > 0$ such that $\mathbb{E}(e^{\lambda X}) < \infty$ for all $|\lambda| \leq c_0$.

(iii) There are constants $c_1, c_2 > 0$ such that

$$\mathbb{P}(|X| \geq s) \leq c_1 e^{-c_2 s}, \text{ for all } s > 0.$$

(iv) The value $\gamma = \sup_{k\geq 2} \left(\frac{\mathbb{E}(X^k)}{k!}\right)^{\frac{1}{k}}$ is finite.

Similar to sub-Gaussian random variables, sub-exponential random variables also satisfy certain deviation/concentration inequalities, as stated below.

THEOREM Appendix A.6 (Tail bound for (sums of independent) sub-exponential random variables). *(Cf. [Wai19, Section 2.1.3])*

(i) Let X be a sub-exponential random variable with parameters (ν, α). Then

$$\mathbb{P}\big(X - \mathbb{E}(X) \geq t\big) \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}} & 0 \leq t \leq \frac{\nu^2}{\alpha} \\ e^{-\frac{t}{2\alpha}} & t > \frac{\nu^2}{\alpha} \end{cases}$$

Consequently, for every $t > 0$, we have $\mathbb{P}\big(X - \mathbb{E}(X) \geq t\big) \leq e^{-\frac{1}{2}\min(\frac{t}{\alpha}, \frac{t^2}{\nu^2})}$. An identical upper bound holds for the corresponding left-hand-side event.

(ii) Let $X_1, \ldots, X_n$ be independent random variables, such that each X_i is sub-exponential with parameters (ν_i, α_i). Then $\sum_{i=1}^n (X_i - \mathbb{E}(X_i))$ is sub-exponential with parameters $\nu_* := \sqrt{\sum_{i=1}^n \nu_i^2}$ and $\alpha_* := \max_{1\leq i\leq n} \alpha_i$. Moreover, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \geq t\right) \leq \begin{cases} e^{-\frac{nt^2}{2\nu_*^2/n}} & 0 \leq t \leq \frac{\nu_*^2}{n\alpha_*} \\ e^{-\frac{nt}{2\alpha_*}} & t > \frac{\nu_*^2}{n\alpha_*} \end{cases}$$

Consequently, for every $t > 0$, we have $\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \geq t\right) \leq e^{-\frac{n}{2}\min(\frac{t}{\alpha_*}, \frac{t^2}{\nu_*^2/n})}$. An identical upper bound holds for the corresponding left-hand-side event.

In this article, the notions of sub-Gaussian and sub-exponential random variables and the relation between them are used, in particular, in the following format.

PROPOSITION Appendix A.7. *Let X be a random variable with zero mean. Suppose there exists $\sigma > 0$ such that for all $s > 0$, we have*

$$\mathbb{P}(X \geq s) \leq \exp\big(-s^2/2\sigma^2\big) \ \text{ and } \ \mathbb{P}(X \leq -s) \leq \exp\big(-s^2/2\sigma^2\big).$$

Then X is a sub-Gaussian random variable with parameter $2\sqrt{2}\sigma$. Moreover, the random variable $X^2 - \mathbb{E}(X^2)$ is a sub-exponential random variable with parameters $(8\sqrt{2}\sigma^2, 8\sigma^2)$.

Proof. We observe that $\mathbb{P}(|X| \geq s) \leq 2\, \exp\big(-s^2/2\sigma^2\big)$, thus, we have

$$\begin{aligned}\mathbb{E}(X^{2k}) &= \int_0^\infty \mathbb{P}(X^{2k} > s)\, ds = \int_0^\infty \mathbb{P}(|X| > s^{\frac{1}{2k}})\, ds \\ &\leq 2\int_0^\infty \exp\big(-s^{\frac{1}{k}}/2\sigma^2\big)\, ds = 2^{k+1}k\sigma^{2k}\int_0^\infty t^{k-1}\exp(-t)\, dt = 2^{k+1}\sigma^{2k}k!,\end{aligned}$$

where we have used the change of variable $s^{\frac{1}{k}}/2\sigma^2 = t$ in the penultimate equality. So,

$$\mathbb{E}(X^{2k}) \leq \frac{(2\sigma)^{2k}k!}{2^{k-1}}, \tag{A.3}$$

which implies that (iii) of Theorem Appendix A.2 holds, when $\theta = 2\sigma$. Now from the proof of implication (iii) $\Rightarrow$ (i) of Theorem Appendix A.2 (see e.g. Page 46 of [Wai19]), we conclude that X is a sub-Gaussian random variable with parameter at most $2\sqrt{2}\sigma$.

To prove the second statement, let $Z = X^2 - \mathbb{E}(X^2)$. Note that for every $k \in \mathbb{N}$, convexity of $f(x) = x^k$, together with the triangle inequality, implies that $|X^2 - \mathbb{E}(X^2)|^k \leq 2^{k-1}(X^{2k} + (\mathbb{E}(X^2))^k)$. Moreover, Jensen's inequality guarantee that $(\mathbb{E}(X^2))^k \leq \mathbb{E}(X^{2k})$. So, we have

$$\begin{aligned}\mathbb{E}(e^{\lambda Z}) &\leq \mathbb{E}(e^{|\lambda Z|}) = 1 + \sum_{k=2}^\infty \frac{|\lambda|^k}{k!}\mathbb{E}\big|X^2 - \mathbb{E}(X^2)\big|^k \leq 1 + \sum_{k=2}^\infty \frac{|\lambda|^k 2^k}{k!}\mathbb{E}(X^{2k}) \\ &\leq 1 + \sum_{k=2}^\infty \frac{|\lambda|^k 2^k}{k!}\big(2^{k+1}\sigma^{2k}k!\big) = 1 + 2(4|\lambda|\sigma^2)^2\sum_{k=0}^\infty (4|\lambda|\sigma^2)^k,\end{aligned}$$

where we have used the bound in (A.3) for $\mathbb{E}(X^{2k})$. Since the sum $\sum_{k=0}^\infty (4|\lambda|\sigma^2)^k$ is bounded by 2, when $4|\lambda|\sigma^2 \leq \frac{1}{2}$, we get

$$\mathbb{E}(e^{\lambda Z}) \leq 1 + 4(4|\lambda|\sigma^2)^2 \leq e^{64\sigma^4\lambda^2}, \quad \text{whenever } |\lambda| \leq \frac{1}{8\sigma^2},$$

that is, Z is sub-exponential with parameters $(8\sqrt{2}\sigma^2, 8\sigma^2)$. ∎

References

[Abb17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *J. Mach. Learn. Res.*, 18:1–86, 2017.

[BCL+11] Christian Borgs, Jennifer Chayes, László Lovász, Vera Sós, and Katalin Vesztergombi. Limits of randomly grown graph sequences. *European J. Combin.*, 32(7):985–999, 2011.

[BJR12] Béla Bollobás, Svante Janson, and Oliver Riordan. Monotone graph limits and quasimonotone graphs. *Internet Mathematics*, 8(3):187–231, 2012.

[CGH+15] Huda Chuangpishit, Mahya Ghandehari, Matthew Hurshman, Jeannette Janssen, and Nauzer Kalyaniwalla. Linear embeddings of graphs and graph limits. *J. Combin. Theory Ser. B*, 113:162–184, 2015.

[CHW19] Xiangyu Chang, Danyang Huang, and Hansheng Wang. A popularity-scaled latent space model for large-scale directed social network. *Statist. Sinica*, 29(3):1277–1299, 2019.

[Con90] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.

[CS11] Victor Chepoi and Morgan Seston. Seriation in the presence of errors: a factor 16 approximation algorithm for l_∞-fitting Robinson structures to distances. *Algorithmica*, 59(4):521–568, 2011.

[DHJ08] Persi Diaconis, Susan Holmes, and Svante Janson. Threshold graph limits and random threshold graphs. *Internet Math.*, 5(3):267–320 (2009), 2008.

[FdV16] Fajwel Fogel, Alexandre d'Aspremont, and Milan Vojnovic. Spectral ranking using seriation. *J. Mach. Learn. Res.*, 17:Paper No. 88, 45, 2016.

[FK99] Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[FMR19] Nicolas Flammarion, Cheng Mao, and Philippe Rigollet. Optimal rates of statistical seriation. *Bernoulli*, 25(1):623–653, 2019.

[GJ19] Mahya Ghandehari and Jeannette Janssen. An Optimization Parameter for Seriation of Noisy Data. *SIAM J. Discrete Math.*, 33(2):712–730, 2019.

[HRH02] Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.*, 97(460):1090–1098, 2002.

[Lii10] Innar Liiv. Seriation and matrix reordering methods: An historical overview. *Stat. Anal. Data Min.*, 3(2):70–91, April 2010.

[Lov12] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[LST17] Monique Laurent, Matteo Seminaroti, and Shin-ichi Tanigawa. A structural characterization for certifying Robinsonian matrices. *Electron. J. Combin.*, 24(2):Paper 2.21, 22, 2017.

[PF14] Pascal Préa and Dominique Fortin. An optimal algorithm to recognize Robinsonian dissimilarities. *J. Classification*, 31(3):351–385, 2014.

[SAC19] Anna L. Smith, Dena M. Asta, and Catherine A. Calder. The geometry of continuous latent space models for network data. *Statist. Sci.*, 34(3):428–453, 2019.

[Spe14] Joel Spencer. *Asymptopia*, volume 71 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2014. With Laura Florescu.

[Wai19] Martin J. Wainwright. *High-dimensional statistics*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019.