# A LECTURE ON GRAPH LIMITS

JAEHOON KIM

Abstract. In this course, we learn basics of graph limit theory. The materials on this note are based on [1, 2, 3]

## 1. Short motivations

Let $G_1, G_2, \ldots\ldots$ be a sequence of graphs. Our aim is to define when this sequence is convergent, and what the limit of this sequence should be. Obviously if a sequence is convergent, the graphs on the sequence eventually must share some characteristics. Consider some examples.

**Example 1.1.** *Let $K_1, K_2, K_3, \ldots$ be the sequence of $n$-vertex cliques.*

How do we feel about the above sequence? Do we feel that the graphs in the sequence shares some characteristic? Do we feel that the sequence must converges to? What should be the limit of this sequence? Let's consider one more example.

**Example 1.2.** *Let $P_1, P_2, P_3, \ldots$ be the sequence of $n$-vertex paths.*

In both of the examples, we feel that both sequences should be convergent. If we ask ourselves why we feel this way, one reason is that their local shapes are alike. For $n_1, n_2 \gg r$, if we take random $s$ vertices from $K_{n_1}$ and $K_{n_2}$, they induces the same graph $K_r$ with high probability (in fact, they are always the same). For $n_1, n_2 \gg r$, if we take a random vertex in $P_{n_1}$ and $P_{n_2}$, the $r$-neighborhood $B_r(v)$ induces the same graph $P_{2r+1}$ in both graphs unless the vertex $v$ is very close to the end of the paths. In this two cases, 'local shapes' are a bit different. In one case, we take arbitrary vertices and in the other case, we take a neighborhood of a vertex. The former case leads to graph limits of dense graphs and the latter case lead to graph limits of sparse graphs. In sum, we will study about the probability distributions of certain random sampling on a given graph.

To further motivate the study on the subgraph statistics (which is the same to the probability distributions of certain random sampling), consider the following problem.

**Question 1.1.** *For an $n$-vertex labeled graph $G$ with labeled edge density at least $1/2$, how small the labeled $C_4$-density can be?*

Here, labeled edge density is $2e(G)/n(n-1)$ and labeled $C_4$-density is the number of labeled $C_4$s divided by $n(n-1)(n-2)(n-3)$. Erdős showed that the labeled $C_4$-density of such a graph is at least $1/16$. However, this number $1/16$ is not achieved by any graph. On the other hand, there exists a sequence $(G_n)_{n\in\mathbb{N}}$ of graphs with edge density at least $1/2$ whose labeled 4-cycles density approaches to $1/16$. This example motivates the study of limits of subgraph densities as it can be used to study certain inequalities from the extremal graph theory.

*Date*: July 26, 2020.

## 2. Graphons

We want to first define homomorphism density, which is a random sampling we will use. Then we define convergence and objects which will be the limit of convergent sequence of dense graphs.

Motivated from Example 1.1, we want to define 'local shape' or 'local behavior' of graphs. As in the previous section, we want to analyze what graphs a set of $s$ chosen vertices induces. For that, we can consider the number of sets $S \in \binom{V(G)}{s}$ which induces a certain graph $F$. However, for several reasons it's more convenient to work with the following set up.

(1) Instead of choosing $s$ vertices at once, we choose vertices one by one. In other words, we distinguish choosing $v_1, v_2$ in order with choosing $v_2, v_1$ in order.
(2) We allow to choose the same vertex multiple times.
(3) We normalize the number, i.e. divide the number by all possible choices.

First two choices allows us to simplify many calculations, and such choices do not make any essential difference. The third choice enables us to deal with graphs with different number of vertices. As we are allowing to choose same vertex multiple times, the following notions of homomorphism is more appropriate than just considering subgraphs. Let's write $|G|, e(G)$ for the number of vertices and the number of edges of $G$.

**Definition 2.1.** *A function $f : V(F) \to V(G)$ is a homomorphism from $F$ to $G$ if $f(u)f(v) \in E(G)$ for all $uv \in E(F)$. Let the homomorphism number $hom(F, G)$ be the number of homomorphisms from $F$ to $G$. Let the homomorphism density $t(F, G)$ be*

$$\frac{hom(F,G)}{|G|^{|F|}}.$$

The homomorphism density $t(F, G)$ captures the local behavior that we wanted to define. This is a probability that a randomly chosen map $f : V(F) \to V(G)$ is a homomorphism. Then, $\{t(F, G)\}_{|F|=s}$ provides the probability distribution of random sampling of $s$ vertices. As this reflects 'local behavior' of our interest, we can propose the following definition.

**Definition 2.2.** *A sequence $(G_n)_{n\in\mathbb{N}}$ is convergent if the sequence $(t(F, G_n))_{n\in\mathbb{N}}$ converges for all graphs $F$.*

Another way of considering this is the following. We enumerate all the graphs into $F_1, F_2, \dots$ and we identify $G$ with an infinite tuple $(t(F_1, G), t(F_2, G), \dots)$. Note that this is a sequence which has only finitely many nonzero terms. With the above definition, the sequence $G_1, G_2, \dots$ converges if and only if the sequence of the corresponding tuples component-wise converges.

With this definition, it is easy to define the limit. The limit of a sequence $G_1, G_2, \dots$ should be the probability distribution

$$(\lim_{n\to\infty} t(F_i, G_n) : i \in \mathbb{N})$$

This is our object of interest. We want to know what kind of probability distribution can be obtained from a sequence of graphs, and what property do they have. Note that every graph corresponds to an infinite tuple with finitely many nonzero components, but the limit can be an infinite tuple with infinitely many nonzeros. Hence, the limits of a convergent sequence of graphs may not be graphs.

Although this definition is very natural, this is not descriptive of what happens and it does not provide much intuition to us. Hence we want to find another more descriptive (equivalent) definition of graph limits.

As we mention before, $t(F, G)$ captures the local behavior of our interest. Assume $V(F) = [s]$. This can be interpreted as the following. We choose random vertices

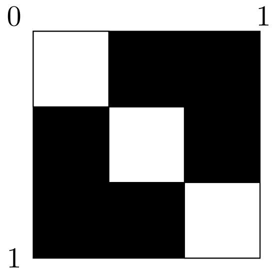


FIGURE 1. An associated graphon $W_G$ for $G = K_{n,n,n}$.

$x_1, x_2, \ldots, x_s$ in $G$ uniformly at random and compute the probability that the map $f(i) = x_i$ is a homomorphism. Here, the graph $G$ provides a distribution of the probability that $x_i x_j$ forms an edge. In above discussion, we identified the graph with an infinite tuple $(t(F_i, G))_{i \in \mathbb{N}}$ which is same as the probability distributions based on this sampling procedures for all possible $s \in \mathbb{N}$. Moreover, what kind of properties this distribution have? For given graph $G$, we can check the following.

(1) For $\{i, j\} \cap \{i', j'\} = \emptyset$, the event of $x_i x_j$ forming an edge and the event of $x_{i'} x_{j'}$ forming an edge are independent.
(2) If we fix $x_i \in V(G)$, then the distribution of $x_i x_j$ forming an edge is determined.

Once $x_i = v$ is anchored, the distribution of $v x_j$ forming an edges are determined. This motivates us to write the distribution for each $v$ in a row to make a matrix. This provides the adjacency matrix. For this adjacency matrix $A$, what is $hom(F, G)$? This can be written as follows where $V(F) = [s]$.

$$t(F, G) = \sum_{f:[s] \to V(G)} \prod_{ij \in E(F)} A(f(i), f(j)) \times \frac{1}{|G|^s}.$$

Now we forget the graph, and try to mimic this for some given (more general) distribution $W(x, y)$ on two variables. We have a measure space $V$, and we choose elements $x_1, \ldots, x_s$ according to some measure, and compute $W(x_i, x_j)$ for all $i, j$ and use this to compute something similar to the homomorphism density. From this motivation, we define as follows by taking $[0, 1]$ as our measure space.

**Definition 2.3.** *Let $\mathcal{W}_0$ denote the set of all symmetric measurable functions $[0,1]^2 \to [0,1]$. The elements of $\mathcal{W}_0$ are called graphons. We also define $\mathcal{W}_1$ be the set of all symmetric measurable functions $[0,1]^2 \to [-1,1]$ and $\mathcal{W}$ be the set of all symmetric measurable functions $[0,1]^2 \to \mathbb{R}$ which are called kernels.*

Let's consider some examples of graphon.

**Example 2.1.** *Given an $n$-vertex (edge)-weighted graph $H$ with edge weights $\beta(ij) \in [0,1]$, we define its associated graphon $W_H : [0,1]^2 \to [0,1]$ as follows. Let $I_1 \cup \cdots \cup I_n$ be a partition of $[0,1]$ into $n$ sets of same Lebesgue measure and let $W_H(x,y) = \beta(ij)$ if $ij \in E(H)$ and $x \in I_i, y \in I_j$. Given an $n$-vertex graph $G$, we identify this with weighted graphs with $\beta(ij) = 1$ for all $ij \in E(G)$ this naturally defines its associated graphon $W_G$. In a similar way, we can defined associated graphon for a symmetric matrix with entries in $[0,1]$.*

Similarly as for the graphs, for given graphon, we can also define homomorphism density as follows. For given graph $F$ on the vertex sets $[s]$,

$$t(F, G) = \int_{[0,1]^s} \prod_{ij \in E(F)} W(x_i, x_j) \prod_{i \in V(F)} dx_i$$

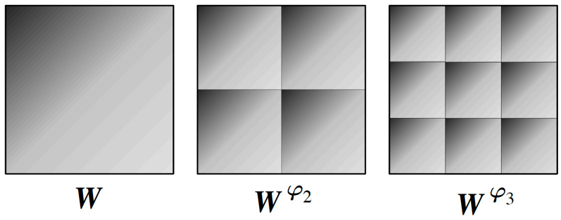


FIGURE 2. $W^{\varphi_2}, W^{\varphi_3}$ are weakly isomorphic but there are no measure-preserving transformation up to a null set from one to another.

If $W(x,y) \in \{0,1\}$ and if we interpret $W(x,y)$ being 1 as $xy$ being an edge, then this is exactly the probability that a map $V(F) \to [0,1]$ chosen at random in uniform measure induces a homomorphism. Hence, $(t(F_i, G))_{i\in\mathbb{N}}$ is a weighted version of probability distribution of random sampling. Again, we define a convergent sequence of graphons using this probability distribution.

**Definition 2.4.** *A sequence $(W_n)_{n\in\mathbb{N}}$ of graphons is convergent if the sequence $(t(F, W_n))_{n\in\mathbb{N}}$ converges for all graphs $F$.*

Recall our purpose again. Our purpose is to define objects which will be limits of convergent graph sequences. Are graphons the objects we wanted? Not quite. Consider two graphons $W$ and $W'$ which are same everywhere except a measure zero set in $[0,1]^2$. Then we have $t(F,W) = t(F,W')$ for all graphs $F$. We wanted to say that a convergent sequence $G_1, G_2, \ldots$ has limit $W$ if $\lim_{i\to\infty} t(F, G_i) = t(F, W)$ for all graphs $F$. However, if the limit of a sequence $G_1, G_2, \ldots$ is $W$, then $W'$ also can be its limit. So, we want to 'identify' two graphons $W$ and $W'$ if $t(F,W) = t(F,W')$ for all graphs $F$.

**Definition 2.5.** *Two graphons $W$ and $W'$ are weakly isomorphic if $t(F,W) = t(F,W')$ for all graphs $F$.*

However, when does two graphons are weakly isomorphic? Consider Example 2.1. There the partition $I_1 \cup \cdots \cup I_n$ of $[0,1]$ into sets of equal measures are arbitrary. Let $W, W'$ be two graphons we get from two different partitions. Then we again have $t(F,W) = t(F,W')$ for any graph $W$. Among two partitions, one partition can be obtained from the other by taking an invertible measure-preserving map of $[0,1]$ and changes on a measure-zero set.

**Definition 2.6.** *We say that a map $\varphi : [0,1] \to [0,1]$ is measure-preserving with respect to a measure $\lambda$ if $\lambda(X) = \lambda(\varphi^{-1}(X))$ for all measurable $X \subseteq [0,1]$.*

**Definition 2.7.** *Two graphons $W, W'$ are isomorphic up to a null set if there exists an invertible measure-preserving map $\varphi : [0,1] \to [0,1]$ such that $W'(\varphi(x), \varphi(y)) = W(x,y)$ almost everywhere. Here, almost everywhere meaning that the equality holds for all $(x,y)$ up to a set of measure zero.*

So, if two graphons $W, W'$ are isomorphic up to a null set, then they are weakly isomorphic. However, does the converse hold? No. One easy way to see this is to observe that the term 'invertible' is not really necessary. Meaning that if we define $W^{\varphi}(x,y) = W(\varphi(x), \varphi(y))$ for a given measure-preserving map $\varphi : [0,1] \to [0,1]$, we have

$$t(F,W) = t(F, W^{\varphi})$$

for all graphs $F$. With this example, it is not difficult to check $t(F,W) = t(F,W^{\varphi})$. However, again there are two weakly isomorphic graphons $W, W'$ which are not obtained by using this relation. For example, consider measure preserving maps $\varphi_k : x \mapsto kx \pmod 1$. Then the graphons $W, W^{\varphi_2}, W^{\varphi_3}$ are weakly isomorphic. But there are no measure-preserving transformation up to a null set from $W^{\varphi_2}$ to $W^{\varphi_3}$. On the other hand, it is known that if two graphons $W_1, W_2$ are weakly isomorphic, then there exists a graphon $W_0$ and measure preserving maps $\varphi_i$ such that $W_i = W_0^{\varphi_i}$ up to a null set for each $i \in [2]$.

## 3. Cut distance

In order to better understand graphons, we want to measure distances between two given graphons and turn the space of graphons into a metric space. Obviously, we want two graphons $W, W'$ to be close if and only if $|t(F,W) - t(F,W')|$ is small for all graphs $F$. Moreover, we want two graphons $W, W'$ are at distance zero if and only if they are weakly isomorphic. How can we define distances between graphons?

Let's consider graphs first. For two graphs $G$ and $G'$ with the same vertex set $[n]$, the most natural way to measure the distance between them is to consider the following.

$$d_1(G,G') = \frac{|E(G)\triangle E(G')|}{n^2}.$$

Of course, we want to identify two graphs if one can be obtained from the other by permuting the vertices.

$$\delta_1(G,G') = \min_{G'' \text{ is a permutation of } G'} \frac{|E(G)\triangle E(G'')|}{n^2}.$$

However, even with this definition, $\delta_1$ does not capture the homomorphism density as we want. Consider two independent random graphs $G, G'$ with edge probability $1/2$, then with high probability, $\delta_1(G,G') = 1/4$. However, for any small graph $F$, we have $t(F,G) = t(F,G') \pm o(1)$. Hence, $\delta_1$ is not the right distance we want.

Also consider the following example.

**Example 3.1.** *Let $G_1, G_2$ be two graphs having the same partition $V_1 \cup \cdots \cup V_r$ such that $G_k[V_i, V_j]$ is $\varepsilon$-regular with density $d_{i,j}$ for all $k \in [2]$ and $ij \in \binom{[r]}{2}$. Then for any small graph $F$, we have $t(F,G_1) = t(F,G_2) \pm O(|e(F)|\varepsilon + 1/r)$.*

Recall that a pair $(S,T)$ of vertex sets in a graph $G$ is $\varepsilon$-regular if

$$\left|\frac{e_G(S',T')}{|S'||T'|} - \frac{e_G(S,T)}{|S||T|}\right| < \varepsilon$$

for all $S' \subseteq S, T' \subseteq T$ with $|S'| \geq \varepsilon|S|, |T'| \geq \varepsilon|T|$.

The above example roughly tells us the following. For two graphs $G_1$ and $G_2$, for not-too-small sets $S, T$, if the number of edges between $S$ and $T$ is similar for the two graphs, then they must be close in terms of our desired distance. Hence this tells us that whether each edge is at the right positions (which is measured by edit distance) is not important but we have to care whether the number of edges between two large sets are correct. This motivates the following definition of cut distance between two graphs with the same vertex set.

$$d_\square(G,G') = \max_{S,T\subseteq V(G)} \frac{|e_G(S,T) - e_{G'}(S,T)|}{n^2} \text{ and } \delta_\square(G,G') = \min_{G'' \text{ is a permutation of } G'} d_\square(G,G'')$$

Of course this can be only defined for two graphs with the same vertex set, and it can be easily generalized into two graphs with the same number of vertices. There are ways to generalize this to two graphs with different number of vertices. However, we rather directly consider graphons. The following norms are natural.

**Definition 3.1.** *For $W : [0,1]^2 \to [0,1]$, we define $L^p$ norm as $\|W\|_p = (\int_{[0,1]^2} |W|^p)^{1/p}$ and $L^\infty$ norm $\|W\|_\infty$ to be the infimum of all the real numbers $\alpha$ such that $\{(x,y) \in [0,1]^2 : W(x,y) > \alpha\}$ has measure zero.*

Similarly to the graph case, let $d_1(U, W) = \|U - W\|_1$. The following cut norm for graphon is also similarly defined as the graph case.

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} |\int_{S\times T} W(x,y)dxdy| \text{ and } d_\square(U,W) = \|U - W\|_\square$$

where the supremum is taken over all measurable subsets $S$ and $T$.

Then we have

$$\|W\|_\square \le \|W\|_1 \le \|W\|_2 \le \|W\|_\infty \le 1.$$

From this we define cut distance as

$$\delta_\square(U,W) = \inf_\varphi d_\square(U, W^\varphi)$$

where $\varphi$ is taken over all invertible measure-perserving

Note that this cut-distance is not a metric as two different graphons can have distance zero. We identify two graphons with cut-distance zero to obtain the set $\widetilde{\mathcal{W}_0}$ of *unlabeled graphons*. Later we will see this cut-distance is the distance we wanted, meaning that two graphons are weakly isomorphic if and only if cut-distance between them is zero.

For $R \subseteq \mathcal{W}_0$, define

$$B_1(R,\varepsilon) = \{W \in \mathcal{W}_0 : \exists U \in R \text{ such that } d_1(W,U) < \varepsilon\}$$
$$B_\square(R,\varepsilon) = \{W \in \mathcal{W}_0 : \exists U \in R \text{ such that } d_\square(W,U) < \varepsilon\}$$

Using these as bases, $L_1$-norm and cut norm defines topologies on $\mathcal{W}$. These two topologies are different. Note that an open set in the topology defined by cut norm is also an open set in $L_1$-norm, but the converse is not true.

## 4. Szemerédi partitions

A function $W \in \mathcal{W}$ is called a *stepfunction* if there exists a partition $S_1 \cup \cdots \cup S_k$ of $[0,1]$ into measurable sets such that $W$ is constant on $S_i \times S_j$ for all $i, j \in [k]$. It is an easy fact that any integrable function can be approximated by a stepfunction in terms of $L^1$-distance.

It would be very useful if we can also approximate graphons by stepfunctions with respect to cut distance. Indeed, this is possible by the following regularity lemma.

**Lemma 4.1** (Weak regularity lemma). *For every $W \in \mathcal{W}$ and $k \ge 1$ there is stepfunction $W$ with $k$ steps such that*

$$\|W - U\|_\square < \frac{4\|W\|_2}{\sqrt{\log k}}.$$

Furthermore, the following counting lemma helps us to approximate $t(F,W)$.

**Lemma 4.2** (Counting lemma). *For $U, W \in \mathcal{W}_0$ and a graph $F$,*

$$|t(F,U) - t(F,W)| \le e(F)\delta_\square(U,W).$$

Hence, to approximate $t(F,W)$, we first find a stepfunction $U$ with small $\|W-U\|_\square$ with $k$ steps. As $U$ is a stepfunction, computing $t(F,U)$ can be done with a finite computation. Then, using the above counting lemma, $t(F,U)$ is an approximation of $t(F,W)$. Moreover, we can ensure such stepfunctions are steppings of $W$, which we define as follows.

**Definition 4.3.** *For a partition $\mathcal{P} = \{S_1,\dots,S_k\}$ of $[0,1]$ into measurable subsets and $W \in \mathcal{W}$. The stepping $W_\mathcal{P}$ is defined by*

$$W_\mathcal{P}(x,y) = \frac{1}{\lambda(S_i)\lambda(S_j)}\int_{S_i\times S_j} W \text{ if } (x,y) \in S_i \times S_j.$$

Here $\lambda(S_i)$ is the Lebesgue measure of $S_i$. Instead of proving weak regularity lemma we will prove the following version.

**Lemma 4.4.** *Let $W \in \mathcal{W}$ and $1 \le m \le k$. For every $m$-partition $\mathcal{Q}$ of $[0,1]$ there is a $k$-partition $\mathcal{P}$ refining $\mathcal{Q}$ such that*

$$\|W - W_{\mathcal{P}}\|_\square \le \frac{4\|W\|_2}{\sqrt{\log(k/m)}}.$$

To prove lemma 4.4, we prove the following lemmas first.

**Lemma 4.5.** *For any $U \in \mathcal{W}$ there are two sets $S, T \subseteq [0,1]$ and a real number $0 \le a \le 1$ such that*

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 \le \|U\|_2^2 - \|U\|_\square^2.$$

*Proof.* Let $S$ and $T$ be measurable subsets of $[0,1]$ such that the cut-norm is achieved with $S$ and $T$, meaning

$$\|U\|_\square = \left| \int_{[0,1]^2} U \cdot \mathbb{1}_{S\times T} \right|.$$

Let $a = \frac{1}{\lambda(S)\lambda(T)}\|U\|_\square$. Then

$$\|U - a\mathbb{1}_{S\times T}\|_2^2 = \int_{[0,1]^2} (U^2 - 2aU \cdot \mathbb{1}_{S\times T} + a^2) = \|U\|_2^2 - \|U\|_\square^2.$$

□

Note that for any $W \in \mathcal{W}$ and a partition $\mathcal{P}$ of $[0,1]$ into measurable sets, we have

$$\|W_{\mathcal{P}}\|_\square \le \|W\|_\square.$$

To see this, consider sets $S, T \subseteq [0,1]$. For some $S_i \in \mathcal{P}$, if $S_i \subsetneq S$, we consider $S' = S \cup S_i$ and $S'' = S \setminus S_i$. Depending on the sign of $\int_{S_i\times T} W_{\mathcal{P}}$, one of $|\int_{S'\times T} W_{\mathcal{P}}|$ and $|\int_{S''\times T} W_{\mathcal{P}}|$ is as large as $|\int_{S\times T} W_{\mathcal{P}}|$. By repeating this, one can obtain $\hat{S}, \hat{T}$ such that

$$|\int_{S\times T} W_{\mathcal{P}}| \le |\int_{\hat{S}\times\hat{T}} W_{\mathcal{P}}|$$

and $\hat{S} = \bigcup_{i\in I} S_i, \hat{T} = \bigcup_{j\in J} T_j$ for some $I, J \subseteq [k]$. For such sets, we have $|\int_{\hat{S}\times\hat{T}} W_{\mathcal{P}}| = |\int_{\hat{S}\times\hat{T}} W| \le \|W\|_\square$. Using this, we can prove the following lemma.

**Lemma 4.6.** *Let $W \in \mathcal{W}_0$ and $U$ be a stepfunction with steps $\mathcal{P}$. Then*

$$\|W - W_{\mathcal{P}}\|_\square \le 2\|W - U\|_\square.$$

*Proof.* As we have $\|(U-W)_{\mathcal{P}}\|_\square \le \|U - W\|_\square$,

$$\|W - W_{\mathcal{P}}\|_\square \le \|W - U\|_\square + \|U - W_{\mathcal{P}}\|_\square = \|W - U\|_\square + \|(U-W)_{\mathcal{P}}\|_\square \le 2\|W - U\|_\square.$$

□

*Proof of Lemma 4.4.* Let $s = \lceil \frac{1}{2}\log(\frac{k}{2m}) \rceil$. By repeatedly applying Lemma 4.5, we obtain pairs of sets $S_i, T_i$ and real numbers $a_i$ such that $W_j = W - \sum_{i=1}^{j} a_i \mathbb{1}_{S_i\times T_i}$ satisfying

$$0 \le \|W_j\|_2^2 \le \|W\|_2^2 - \sum_{i=0}^{j-1} \|W_i\|_\square^2.$$

Thus, there exists $t \le s$ such that $\|W_t\|_\square \le \frac{\|W\|_2}{\sqrt{s}}$. Consider $U = \sum_{i=1}^{t} \frac{a_i}{2}(\mathbb{1}_{S_i\times T_i} + \mathbb{1}_{T_i\times S_i})$. As $W$ is symmetric, we have

$$\|W - U\|_\square \le \|\frac{1}{2}W - \frac{1}{2}\sum_{i=1}^{j} a_i \mathbb{1}_{S_i\times T_i}\|_\square + \|\frac{1}{2}W - \frac{1}{2}\sum_{i=1}^{j} a_i \mathbb{1}_{T_i\times S_i}\|_\square \le \|W_t\|_\square \le \frac{\|W\|_2}{\sqrt{s}}.$$

Note that $U$ is a stepfunction with at most $2^{2s}$ steps. Let $\mathcal{Q}'$ be the set of these steps. Let $\mathcal{P}$ be the common refinement of $\mathcal{Q}$ and $\mathcal{Q}'$, then $\mathcal{P}$ has at most $m2^{2s} \leq k$. Applying Lemma 4.6, we finish the proof. □

We have defined the metric space $(\widetilde{\mathcal{W}}, \delta_\square)$. We can prove that this space is compact. In particular, this implies that the limit of a convergent sequence of graphons is again a graphon. (Every compact metric space is complete, so the limit of convergent graphon sequence is a graphon.) To prove that, we need some concepts.

**Definition 4.7.** *A sequence $X_0, X_1, \dots$ of random variables is a martingale if $\mathbb{E}[X_n \mid X_{n-1}, \dots, X_0] = X_{n-1}$ for all $n$.*

Imagine that each $X_{i+1}$ is determined after we have $X_i$. As $\mathbb{E}[X_n] = X_{n-1}$, it is not likely that $|X_n - X_{n-1}|$ is too big and the sequence is not likely to deviate too much from $X_0$. This intuition is captured in the following fact. We omit the proof of this theorem.

**Theorem 4.8.** *Let $X_0, X_1, \dots$ be a martingale where each taking real value and we have $\sup_n \mathbb{E}[|X_n|] < \infty$. Then $X_0, X_1, \dots$ is convergent with probability 1.*

We now prove the following theorem.

**Theorem 4.9.** *The metric space $(\widetilde{\mathcal{W}_0}, \delta_\square)$ is compact.*

*Proof.* Based on the definition of compactness, our goal is to prove the following: every sequence $W_1, W_2, \dots$ of graphons has a convergent subsequence.

For each $n \in \mathbb{N}$, we apply Lemma 4.4 to obtain the partitions $\mathcal{P}'_{n,k}$ of $[0,1]$ satisfying the followings, where $\mathcal{P}'_{n,1} = \{[0,1]\}$ is a trivial partition. Here, we allow the partitions $\mathcal{P}'_{n,k}$ to have several empty parts to ensure (3).

(1) $\|W_n - W'_{n,k}\| \leq 1/k$ where $W'_{n,k} = (W_n)_{\mathcal{P}'_{n,k}}$.
(2) $\mathcal{P}'_{n,k+1}$ refines $\mathcal{P}'_{n,k}$
(3) $|\mathcal{P}'_{n,k}| = m_k$ where $m_k$ depends only on $k$.

Note that $\mathcal{P}_{n,1} = \mathcal{P}'_{n,1}$ is a trivial partition, thus it is a partition of $[0,1]$ into intervals. We also want to ensure this for all partitions with larger second indices. Let $W_{n,1} = W'_{n,1}$.

For a fixed $n$ and $k \geq 2$, assume $\mathcal{P}_{n,k-1}$ is a partition of $[0,1]$ into intervals. We apply a measure preserving bijection to $W'_{n,k}$ and $\mathcal{P}'_{n,k}$ to obtain $W_{n,k}$ and $\mathcal{P}_{n,k}$ such that $\mathcal{P}_{n,k}$ is a partition of $[0,1]$ into intervals and it refines $\mathcal{P}_{n,k-1}$. From this we obtain the following for each $k \geq 1$.

(W1) $\delta_\square(W_n, W_{n,k}) \leq 1/k$.
(W2) $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$ and all $\mathcal{P}_{n,k}$ partitions $[0,1]$ into intervals.
(W3) $|\mathcal{P}_{n,k}| = m_k$ where $m_k$ depends only on $k$.

Let
$$\ell(n,k) = (\ell_{n,1}, \dots, \ell_{n,m_k}) \in [0,1]^{m_k}$$
where $\ell_{n,i}$ is the length of the $i$-th interval in $\mathcal{P}_{n,k}$. We first find an increasing sequence $n'_1, n'_2, \dots$ of natural numbers satisfying the following.

(N$'$1) For each $k \in \mathbb{N}$, the sequence $\ell(n'_1, k), \ell(n'_2, k), \dots$ converges.

To show that this is possible, assume that we have an infinite sequence $n_1^{k-1}, n_2^{k-1}, \dots$ such that $\ell(n_1^{k-1}, j), \ell(n_2^{k-1}, j) \dots$ converges in $[0,1]^{m_j}$ for all $j \leq k-1$. Now, we consider the sequence $\ell(n_1^{k-1}, k), \ell(n_2^{k-1}, k), \dots$ in the compact space $[0,1]^{m_k}$. This has a convergent subsequence $\ell(n_1^k, k), \ell(n_1^k, k), \dots$ where $n_j^k = n_j^{k-1}$ for $j \leq k$. By repeating this, we obtain the desired sequence $n'_1, n'_2, \dots$ where $n'_k = n_k^k$ is well-defined by our construction.

Now, we take a subsequence $n_1, n_2, \dots$ of $n'_1, n'_2, \dots$ and graphons $U_k$ as follows.

(N2) For each $k \in \mathbb{N}$, the sequence $W_{n_1,k}, W_{n_2,k} \dots$ converges to $U_k$ on almost all points on $[0,1]^2$.

To show that this is possible, we only focus on the values of the graphon on $j$-th interval $\times$ $j'$-th interval. Fix $k$ and consider $f_{i,k} : [m_k]^2 \to [0,1]$ be a function where $f_{i,k}(j,j')$ is the value of $W_{n'_i,k}$ on the $j$-th interval $\times$ $j'$-th interval. Note that $f_{i,k} \in [0,1]^{m_k^2}$ for each $k$. As before, we can repeatedly taking subsequences for each $k$ in an increasing order, we can find subsequence $n_1, n_2, \dots$ where $f_{n_1,k}, f_{n_2,k}, \dots$ converges in $[0,1]^{m_k^2}$ for all $k \geq 1$. For fixed $k$, together with (N$'$1), the limit of this subsequence and the limit of $\ell(n_1,k), \ell(n_2,k), \dots$ yields $U_k$ as desired. Let $\mathcal{P}_k$ be the partition corresponding to $U_k$. Note that this is the partition corresponding to the limit of $\ell(n'_1,k), \ell(n'_2,k), \dots$.

By deleting all $W_i$ with $i \notin \{n_1, n_2, \dots\}$, and renaming the remaining graphons to $W_1, W_2, \dots$, we assume that $W_{1,k}, W_{2,k}, \dots$ converges to $U_k$ and $U_k$ is a step graphon on $\mathcal{P}_k$. By (W1), each $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$, we have $W_{n,k} = (W_{n,k+1})_{\mathcal{P}_{n,k}}$. From this, we obtain $U_k = (U_{k+1})_{\mathcal{P}_{n,k}}$.

Let $(x,y)$ be a point in $[0,1]^2$ chosen uniformly at random. Then the above relation shows that $U_1(x,y), U_2(x,y), \dots$ forms a bounded martingale. Hence Theorem 4.8 implies that this sequence is convergent with probability 1. In other words, $U_1, U_2, \dots$ converges on almost all points $(x,y)$ on $[0,1]^2$. Let $U$ be the limit where we assign 0 on the null set of points of non-convergence.

Now we claim that $\delta_\square(W_n, U)$ converges to zero as $n$ tends to infinity. Let $\varepsilon > 0$. As $U_k$ converges to $U$, we have $\|U_k - U\|_1 \leq \varepsilon/3$ for some $k > 3/\varepsilon$. Fix this $k$ and choose $n_0$ such that $\|U_k - W_{n,k}\|_1 \leq \varepsilon/3$ for all $n \geq n_0$. Then we have

$$\begin{aligned}\delta_\square(W_n, U) &\leq \delta_\square(W_n, W_{n,k}) + \delta_\square(W_{n,k}, U_k) + \delta_\square(U_k, U) \\ &\leq 1/k + \|U_k - W_{n,k}\|_1 + \|U_k - U\|_1 \leq \varepsilon.\end{aligned}$$

This shows that the subsequence $W_1, W_2, \dots$ converges in $\delta_\square$ metric. □

Recall that we have defined the convergence of graphs in terms of its homomorphism densities. To quantify this, we introduced the notion of the cut distance. To show that the cut distance serves our purpose well, we need to prove the counting lemma.

*Proof of Lemma 4.2.* It suffices to prove $|t(F,W) - t(F,U)| \leq e(F)\|W-U\|_\square$ as one can taking infimum over all right hand side over all $U^\varphi$ where $\varphi$ is is measure-preserving bijections.

We first prove the following claim.

**Claim 1.**

$$\|W\|_\square = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} W \right| = \sup_{u,v:[0,1]\to[0,1]} \left| \int_{[0,1]^2} W(x,y)u(x)v(y)dxdy \right|,$$

*where the supremum is taken over all measurable functions $u$ and $v$.*

*Proof.* The first equality is the definition of cut norm. We only consider the second equality. Right hand side being as big as the left hand side is clear by taking $u = \mathbb{1}_S, v = \mathbb{1}_T$.

Consider $u, v : [0,1] \to [0,1]$. If there exist a set $A$ of measure $\varepsilon > 0$ where $\varepsilon < u(x) < 1-\varepsilon$ for all $x \in A$, then consider $c = \int_{[0,1]^2} W(x,y)\mathbb{1}_A(x)v(y)$. Depending on the sign of $c$, either replacing $u$ with $u + \varepsilon\mathbb{1}_A$ or $u - \varepsilon\mathbb{1}_A$ yields a value of $|\int_{[0,1]^2} W(x,y)u(x)v(y)dxdy|$ as large as before. This shows that the supremum is obtained by the supremum over indicator functions up to null sets. Implying the desired equality. □

We now prove the counting lemma. We enumerate the edges of $F$, and let $i_s j_s$ be the $s$-th edge of $F$.

$$
\begin{aligned}
|t(F,W)-t(F,U)| &= \left|(\int_s \prod_s W(x_{i_s},x_{j_s}) - \prod_s U(x_{i_s},x_{j_s}))\prod_i dx_i\right| \\
&\le \sum_{s=1}^{e(F)} \left|\int \left(\prod_{s=1}^{s'-1} U(x_{i_s},x_{j_s})(W(x_{i_{s'}},x_{j_{s'}})-U(x_{i_{s'}},x_{j_{s'}}))\prod_{s=s'+1}^{e(F)} W(x_{i_s},x_{j_s})\right)\prod dx_i\right| \\
&\le e(F)\|W-U\|_\square.
\end{aligned}
$$

The last inequality comes from the Claim as all the $\prod_{s=1}^{s'-1} U(x_{i_s},x_{j_s})\prod_{s=s'+1}^{e(F)} W(x_{i_s},x_{j_s})$ can be expressed as

$$
\prod_{s<s':i_s=i_{s'}} U(x_{i_s},x_{j_s}) \prod_{s>s':i_s=i_{s'}} W(x_{i_s},x_{j_s}) \times \prod_{s<s':j_s=j_{s'}} U(x_{i_s},x_{j_s}) \prod_{s>s':j_s=j_{s'}} W(x_{i_s},x_{j_s})
$$
$$
\times \text{ remaining terms.}
$$

As the first two terms can play the roles of $u$ and $v$ and the last term is just constant when we consider all variables other than $x_{i_{s'}}, x_{j_{s'}}$ as constants. ☐

**Definition 4.10.** *A sequence $W_1, W_2, \dots$ converges if $t(F, W_i)$ converges for all graphs $F$.*

This definition extends the convergence of graphs. For graphs, its limit is not necessarily a graph. But for graphons, the limit of convergent graphon sequence is again a graphon.

**Theorem 4.11.** *Let $W_1, W_2, \dots$ be a convergent sequence of graphons. Then there exists a garaphon $W$ such that $t(F, W_n)$ converges to $t(F, W)$ for every graph $F$.*

*Proof.* As $\widetilde{W}_0$ is compact, there exists a subsequence $n_1, n_2, \dots$ and a graphon $W$ such that $\delta_\square(W_{n_i}, W) \to 0$ as $i$ tends to infinity. Using counting lemma, $t(F, W_{n_i}) \to t(F, W)$. However, our assumption tells that $\{t(F, W_n)\}_n$ converges, hence $t(F, W_n) \to t(F, W)$. ☐

It is known that $t(F, \cdot) : \widetilde{\mathcal{W}_0} \to [0, 1]$ is a continuous function. Hence the solutions on optimization problems on homomorphism densities can be attained. For example, there exists a graphon $W$ where $t(K_3, W)$ is the minimum over all graphons with $t(K_2, W) \ge 1/2$. Such a graphon provides an approximate solution to the graph case problem. Many problems in extremal graph theory can be reformulated in this way.

## 5. Sampling and inverse counting lemma

Now we want to prove the following theorem.

**Theorem 5.1.** *Let $W_1, W_2, \dots$ be a sequence of graphons. Then the sequence is convergent if and only if it is Cauchy with respect to $\delta_\square$.*

'If' direction of this theorem is proved by the counting lemma above. We want to prove the 'only if' direction. In other words, we want to prove the following.

**Lemma 5.2** (Inverse counting lemma)**.** *Let $k$ be a positive integer, let $U, W \in \mathcal{W}_0$, and assume that for every simple graph $F$ on $k$ nodes, we have $|t(F, U) - t(F, W)| \le 2^{-k^2}$. Then $\delta_\square(U, W) \le \frac{100}{\sqrt{\log k}}$.*

We also want to prove that every graphon is the limit of a sequence of graphs. To prove this and the above inverse counting lemma, we consider some random sampling on a graphon. We will choose some random points on $[0, 1]$ and construct some (weighted) graph based on the chosen points and the graphon.

Let's collect the following Azuma's inequality.

**Theorem 5.3.** *Let $X_1, X_2, \dots$ be a martingale such that $|X_{m+1} - X_m| \leq 1$ for every $m$. Then*
$$\mathbb{P}[X_m > X_0 + \lambda] < e^{-\lambda^2/(2m)}.$$

We also consider the following definitions for a given graphon $W$.

**Definition 5.4.** *Let $W$-random weighted graph $\mathbb{H}(n, W)$ be the edge-weighted graph obtained as follows: we pick $x_1, \dots, x_n$ from $[0, 1]$ uniformly at random and the resulting graph has a vertex set $[n]$ and $ij$ has the weight $W(x_i, x_j)$.*

**Definition 5.5.** *For a given weighted graph $H$ with edge weight $\beta$, let $\mathbb{G}(H)$ be random graph model which yields a graph $G$ with vertex set $V(H)$ and each pair $uv$ is an edge in $G$ independently at random with probability $\beta(uv)$.*

**Definition 5.6.** *A $W$-random graph $\mathbb{G}(n, W)$ is a probability distribution of graphs where the resulting simple graph is obtained as follows: we pick a random edge-weighted graph $H \in \mathbb{H}(n, W)$ and for each edge $ij$ with weight $\beta(ij)$, $G$ has the edge $ij$ with probability $\beta(ij)$ independently at random. In other words, $\mathbb{G}(n, W) = \mathbb{G}(\mathbb{H}(n, W))$.*

We identify a simple graph with an edge-weighted graph with weights 0 and 1.

**Theorem 5.7.** *Let $f$ be a graph parameter where $|f(H) - f(H')| \leq 1$ for two edge-weighted graphs whose weights differ on only edges incident to one vertex. Let $W \in \mathcal{W}_0$ and $k \geq 1$. Then for all $\lambda \geq 0$,*
$$\mathbb{P}\left(f(\mathbb{G}(k, W)) \geq \mathbb{E}[f(\mathbb{G}(k, W))] + \sqrt{2\lambda k}\right) \leq e^{-\lambda}, \text{ and}$$
$$\mathbb{P}\left(f(\mathbb{H}(k, W)) \geq \mathbb{E}[f(\mathbb{H}(k, W))] + \sqrt{2\lambda k}\right) \leq e^{-\lambda}.$$

*Proof.* Let's only consider $\mathbb{G}(k, W)$ as $\mathbb{H}(k, W)$ is simpler. We choose $x_1, \dots, x_k$ in $[0, 1]$ uniformly at random, and we obtain $G$ as $ij$ is an edge of $G$ with probability $W(x_i, x_j)$. This random process yields the distribution of $\mathbb{G}(k, W)$. For $i \geq 0$, let $X_i = \mathbb{E}[f(G) : x_1, \dots, x_i]$. Then one can check that $|X_{i+1} - X_i| \leq 1$. Hence, we can use the Azuma's inequality to obtain the desired result. □

For given $U \in \mathcal{W}$ and points $x_1, \dots, x_k$ in $[0, 1]$ write $X = (x_1, \dots, x_k)$, let $U[X]$ be the $k \times k$ matrix with $U[X]_{ij} = U(x_i, x_j)$. This symmetric matrix corresponds to a weighted graph, whose cut norm can be also defined as follows
$$\|A\|_\square = \max_{S,T \subseteq [k]} \frac{\sum_{i \in S, j \in T} A_{i,j}}{k^2}.$$
Note that $\mathbb{H}(k, U)$ and $U[X]$ are essentially the same object. Using this, we can prove the following sampling lemma.

**Lemma 5.8.** *Let $U \in \mathcal{W}_1$ and let $X$ be an ordered tuple of $k$ points in $[0, 1]$ chosen uniformly at random. Then with the probability at least $1 - 4e^{-\sqrt{k}/10}$,*
$$\frac{-3}{k} \leq \|U[X]\|_\square - \|U\|_\square \leq \frac{9}{k^{1/4}}.$$

To prove this lemma, we define the following for given kernel $W$.
$$\|W\|_\square^+ = \sup_{S,T \subseteq [n]} \int_{S \times T} W(x, y) dx dy.$$
Note that $\|W\|_\square = \max\{\|W\|_\square^+, \|-W\|_\square^+\}$.

Let $A = U[X]$ be a matrix. For any set $R$ of rows and $C$ of columns let $A(R, C) = \sum_{i \in R, j \in C} A_{ij}$. We write $R^+ = \{j : B(R, \{j\}) > 0\}, C^- = \{B(\{j\}, C) \leq 0\}$ and define $R^-, C^+$ analogously.

**Lemma 5.9.** *Let $A$ be a matrix with $\|A\|_\infty \le 1$ and $S_1, S_2 \subseteq [k]$ and let $Q$ be a random $q$-subset of $[k]$. Then*

$$A(S_1, S_2) \le \mathbb{E}_Q(A((Q \cap S_2)^+, S_2) + \frac{k^2}{\sqrt{q}}.$$

*Proof.* Note that the right hand side is independent of $S_1$. Hence, we can replace $S_1$ if necessary to assume that $A(S_1, S_2)$ is maximum among all choices of $S_1$. In other words, $S_1 = S_2^+$ and $[k] \setminus S_1 = S_2^-$.

Consider row $i$ of $A$, and and $a_i = \sum_{j \in S_2} A_{ij}$ and $c_i = \sum_{j \in S_2} A_{ij}^2$. For each $i \in S_1 = S_2^+$ and $j \in [k] \setminus S_1 = S_2^-$, let

$$f_i := a_i \mathbb{1}_{j \in (Q \cap S_2)^-} \text{ and } g_i := a_i \mathbb{1}_{j \in (Q \cap S_2)^+}.$$

Observe that

$$A((Q \cap S_2)^+, S_2) = A(S_1, S_2) - (\sum_{i \in S_1} f_i - \sum_{j \notin S_1} g_i).$$

Now we want bound the expectation of $\sum_{i \in S_1} f_i - \sum_{j \notin S_1} g_i$ from above.

For each $i \in [k]$, let $T_i = \sum_{j \in Q \cap S_2} A_{ij}$. Then we have $\mathbb{E}[T_i] = qa_i/k$. Let $E_j$ be $A_{ij}$ if $j \in Q$ and 0 otherwise. Then we can check that $Var(T_i) \le \frac{qc_i}{k}$ holds.[1] Hence,

$$\mathbb{P}(a_i T_i \le 0) \le \mathbb{P}\left(|T_i - \frac{qa_i}{k}| \ge \frac{qa_i}{k}\right) \le \frac{k^2 Var(T_i)}{q^2 a_i} < \frac{kc_i}{qa_i^2}.$$

As $\mathbb{P}(a_i T_i \le 0)$ is between 0 and 1, this implies that $\mathbb{P}(a_i T_i \le 0) \le \frac{\sqrt{kc_i}}{\sqrt{q}|a_i|}$. So, we have

$$\mathbb{E}_Q(\sum_{i \in S_1} f_i - \sum_{j \notin S_1} g_i) \le \sum_{i \in S_1} a_i \frac{\sqrt{kc_i}}{\sqrt{q}|a_i|} - \sum_{i \notin S_1} a_i \frac{\sqrt{kc_i}}{\sqrt{q}|a_i|} \le \frac{k^2}{\sqrt{q}}.$$

This proves the lemma. $\square$

The following lemma reduces the number of rectangles to consider in estimating cut norms.

**Lemma 5.10.** *Let $Q_1, Q_2$ be random $q$-subsets of $[k]$. Then*

$$\|A\|_\square^+ \le \frac{1}{k^2} \mathbb{E}_{Q_1, Q_2}\left[\max_{T_i \subseteq Q_i} A(T_2^+, T_1^+)\right] + \frac{2}{\sqrt{q}}.$$

*Proof.* By the previous lemma, for some $S_1, S_2 \subseteq [k]$, we have

$$k^2 \|A\|_\square^+ \le \mathbb{E}_{Q_2}(A((Q_2 \cap S_2)^+, S_2)) + \frac{k^2}{\sqrt{q}}.$$

We apply the same lemma again with rows and columns interchanged. Then we have

$$\begin{aligned} A((Q_2 \cap S_2)^+, S_2) &\le \mathbb{E}_{Q_1}((Q_2 \cap S_2)^+, (Q_1 \cap S_1 \cap (Q_2 \cap S_2)^+)^+) + \frac{k^2}{\sqrt{q}} \\ &\le \mathbb{E}_{Q_1}(\max_{T_i \subseteq Q_i} A(T_2^+, T_1^+)) + \frac{k^2}{\sqrt{q}}. \end{aligned}$$

---

1

$$\begin{aligned} Var(T_i) &= \sum_{j \in C} Var(E_j) + 2\sum_{j<j'} Cov(E_j, E_{j'}) = \sum_{j \in C} (\mathbb{E}[E_j^2] - \mathbb{E}[E_j]^2) + 2\sum_{j<j'} (\mathbb{E}[E_j E_{j'}] - \mathbb{E}[E_j]\mathbb{E}[E_{j'}]) \\ &= \frac{qc_i}{k} - \sum_{j \in C} (\frac{qa_i}{k})^2 + 2\sum_{j<j'} (\frac{A_{ij}A_{ij'}}{k(k-1)} - \frac{A_{ij}A_{ij'}}{k^2}) = \frac{qc_i}{k} - \frac{q^2 c_i}{k^2} + \frac{a_i^2 - c_i}{k^2(k-1)} \le \frac{qc_i}{k}. \end{aligned}$$

Here we use the Cauchy-Schwartz to show that $a_i^2 \le kc_i$.

Thus, we have

$$\|A\|_{\square}^{+} \leq \frac{1}{k^2}\mathbb{E}_{Q_1,Q_2}\left[\max_{T_i\subseteq Q_i} A(T_2^+,T_1^+)\right] + \frac{2}{\sqrt{q}}.$$

□

*Proof of Lemma 5.8.* We prove that the following holds with probability at least $1 - 2e^{-\sqrt{k}/10}$.

$$-\frac{3}{k} \leq \|U[X]\|_{\square}^{+} - \|U\|_{\square}^{+} \leq \frac{9}{k^{1/4}}.$$

Let $A = U[X]$. Recall that $X = (x_1, x_2, \ldots, x_k)$ is an ordered tuple of randomly chosen points. We will consider the expectation of $\|U[X]\|_{\square} - \|U\|_{\square}$. For two measurable sets $S_1, S_2 \subseteq [0,1]$, we have

$$\|A\|_{\square}^{+} \geq \frac{1}{k^2}U(S_1\cap X, S_2\cap X) = \sum_{x\in S_1\cap X, y\in S_2\cap X} U(x,y).$$

As we choose $X$ at random, we have

$$\begin{aligned}\mathbb{E}_X(\|A\|_{\square}^{+}) &\geq \frac{1}{k^2}\mathbb{E}_X(U(S_1\cap X, S_2\cap X)) = \frac{k-1}{k}\int_{S_1\times S_2} U(x,y)dxdy + \frac{1}{k}\int_{S_1\cap S_2} U(x,x)dx\\ &\geq \int_{S_1\times S_2} U(x,y)dxdy - \frac{2}{k}.\end{aligned}$$

By taking supremum of the right side over all measurable sets $S_1, S_2$ we obtain

$$\mathbb{E}_X(\|A\|_{\square}^{+}) \geq \|U\|_{\square}^{+} - \frac{2}{k}.$$

We apply Theorem 5.7 to conclude that the lower bound holds with probability at least $1 - e^{-\sqrt{k}/10}$.

We now prove the upper bound on the $\|U[X]\|_{\square} - \|U\|_{\square}$, we let $Q_1, Q_2$ be random $q$-subsets of $[k]$, where $q = \lfloor\sqrt{k}/4\rfloor$. The previous lemma implies that for every $X$ and $A = U[X]$, we have

$$\|A\|_{\square}^{+} \leq \frac{1}{k^2}\mathbb{E}_{Q_1,Q_2}\left[\max_{T_i\subseteq Q_i} A(T_2^+,T_1^+)\right] + \frac{2}{\sqrt{q}}.$$

Now we take the expectation of this over the choices of $X$ while fixing some points corresponding to $Q = Q_1 \cup Q_2$. We fix the set $T_j \subseteq Q_j \subseteq [k]$ with $j \in [2]$, and points $x_i \in [0,1]$ for which $i \in Q = Q_1 \cup Q_2$. Let

$$Y_1 = \{y \in [0,1] : \sum_{i\in T_1} U(x_i, y) > 0\} \text{ and } Y_2 = \{y \in [0,1] : \sum_{i\in T_2} U(y, x_i) > 0\}.$$

Let $X' = (x_i : i \in [k] \setminus Q)$, then

$$\begin{aligned}\mathbb{E}_{X'}[A(T_2^+,T_1^+)] &\leq \sum_{i\in T_2^+\setminus Q, j\in T_1^+\setminus Q} \int_{Y_1\times Y_2} U + \sum_{\{i,j\}\cap Q\neq\emptyset} 1\\ &\leq k^2\|U\|_{\square}^{+} + 4kq.\end{aligned}$$

We now want to prove concentration of $A(T_2^+, T_1^+)$. Without loss of generality, assume that $[k] \setminus Q = [k-q]$. As we have fixed $(x_i : i \in Q)$, the following is a random variable depending on $x_i : i \in [k] \setminus Q$ :

$$X_i = \mathbb{E}[A(T_2^+,T_1^+) : x_1, \ldots, x_i].$$

Moreover, changing one point $x_i$ changes the value of this by at most $4k$ as at most $2k$ entries changes by at most 2.

We use Azuma's inequality to conclude that with the probability at least $1-e^{-2q}$, we have
$$A(T_2^+,T_1^+) \le \mathbb{E}_{X'}[A(T_2^+,T_1^+)] + 8k\sqrt{kq} \le k^2\|U\|_\Box^+ + 4kq + 8k\sqrt{kq}.$$
As there are $4^q$ pairs of $(T_1,T_2)$, union bound yields that with probability at least $1-4^q e^{-2q} \ge 1-e^{q/2}$, this holds for all $T_1 \subseteq Q_1, T_2 \subseteq Q_2$, and so it holds also for the maximum over all choices. We take the expectation over all $Q_1,Q_2$, then with probability at least $1-e^{-q/2}$, we have
$$\|A\|_\Box^+ \le \|U\|_\Box^+ + \frac{2}{\sqrt{q}} + \frac{4q}{k} + \frac{8\sqrt{q}}{\sqrt{k}}.$$
If $k$ is large enough, this implies the upper bound.

Applying the above proof to both $U$ and $-U$, then with probability $1-4e^{-\sqrt{k}/10}$, we have
$$-\frac{3}{k} \le \|U[X]\|_\Box^+ - \|U\|_\Box^+ \le \frac{9}{k^{1/4}} \text{ and } -\frac{3}{k} \le \|U[X]\|_\Box^- - \|U\|_\Box^- \le \frac{9}{k^{1/4}},$$
implying the desired inequality. □

**Lemma 5.11.** *For every $q$-vertex edge-weighted graph $H$ with edge weights in* $[0,1]$ *and* $\lambda \ge 10/\sqrt{q}$*, we have* $\mathbb{P}(d_\Box(\mathbb{G}(H),H) > \lambda) \le e^{-\lambda^2 q^2/100}$.

*Proof.* For $i,j \in [q]$, let $X_{ij} = \mathbb{1}_{ij\in E(\mathbb{G}(H))}$ be the indicator random variable. For two disjoint sets $S,T \subseteq [q]$, $\{X_{ij} : i \in S, j \in T\}$ is a collection of independent random variables with $\mathbb{E}[X_{ij}]$ being the edge weight $\beta(ij)$. So,
$$e_{\mathbb{G}(H)}(S,T) - e_H(S,T) = \sum_{i\in S, j\in T} (X_{ij} - \mathbb{E}(X_{ij})).$$
By using Chernoff's inequality, we have
$$\mathbb{P}\left[ \Big| \sum_{i\in S,j\in T} (X_{ij} - \mathbb{E}(X_{ij})\Big| > \frac{1}{4}\lambda q^2 \right] \le 2\exp(\frac{-\lambda^2 q^4}{32|S||T|}) \le 2\exp(\frac{-\lambda^2 q^2}{32}).$$
As there are $3^q$ possible pairs $(S,T)$, so the probability that the above events hold for all $(S,T)$ is at most $2\exp(\frac{-\lambda^2 q^2}{32})3^q < e^{-\lambda^2 q^2/100}$. In this case, we have $d_\Box(\mathbb{G}(H),H) \le \lambda$. □

Using these, we can prove the following.

**Lemma 5.12.** *Let $k \ge 1$, and $W \in \mathcal{W}_0$. Then with probability at least* $1-\exp(-k/(2\log k))$,
$$d_\Box(\mathbb{G}(k,W),W) \le \frac{30}{\sqrt{\log k}}.$$

*Proof.* We wish to bound the expectation of $d_\Box(\mathbb{G}(k,W),W)$ and apply Theorem 5.7. We apply regularity lemma to find an equipartition $\mathcal{P} = \{V_1,\dots,V_m\}$ of $[0,1]$ into $m = \lfloor k^{1/4} \rfloor$ classes such that $d_\Box(W,W_\mathcal{P}) \le \frac{20}{\sqrt{\log k}}$.

Let $S$ be an ordered tuple of $k$ random points from $[0,1]$, then Lemma 5.8 implies $|d_\Box(W[S],W_\mathcal{P}[S]) - d_\Box(W,W_\mathcal{P})| \le \frac{9}{k^{1/4}}$ with probability at least $1-4e^{-\sqrt{k}/10}$. Hence
$$\mathbb{E}\left[\big|d_\Box(W[S],W_\mathcal{P}[S]) - d_\Box(W,W_\mathcal{P})\big|\right] \le \frac{9}{k^{1/4}} + 4e^{-\sqrt{k}/10}\cdot 1 \le \frac{10}{k^{1/4}}.$$
Then, we have
$$\mathbb{E}[d_\Box(W[S],W_\mathcal{P}[S])] \le d_\Box(W,W_\mathcal{P}) + \frac{10}{k^{1/4}} \le \frac{23}{\sqrt{\log k}}.$$

On the other hand, let $H = W_{\mathcal{P}}[S]$. Then the graphons $W_{\mathcal{P}}$ and $W_H$. Let $|V_i \cap S|/k = 1/m + r_i$, then two corresponding intervals of two step graphons have length difference $r_i$. Hence, $\delta_\square(W_{\mathcal{P}}, W_H) \leq \delta_1(W_{\mathcal{P}}, W_H) \leq 2\sum_i r_i$. Then we have

$$\mathbb{E}[\delta_\square(W_{\mathcal{P}}, W_H)] \leq 2\sum_i \mathbb{E}[|r_i|] = 2m\mathbb{E}(|r_1|) \leq 2m\sqrt{\mathbb{E}(r_1^2)} = 2\sqrt{\frac{m-1}{k}} < \frac{1}{k^{3/8}}.$$

[2] Hence,

$$\mathbb{E}[\delta_\square(W, W[S])] \leq \delta_\square(W, W_{\mathcal{P}}) + \mathbb{E}[\delta_\square(W_{\mathcal{P}}, W_{\mathcal{P}}[S])] + \mathbb{E}[\delta_\square(W_{\mathcal{P}}[S], W[S])] \leq \frac{25}{\sqrt{\log k}}.$$

Using Lemma 5.11 with $\lambda = \frac{10}{\sqrt{k}}$,

$$\mathbb{E}[\delta_\square(\mathbb{H}(k, W), \mathbb{G}(k, W))] \leq \frac{10}{\sqrt{k}} \cdot (1 - e^{-\sqrt{k}}) + e^{-\sqrt{k}} \cdot 1 \leq \frac{11}{\sqrt{k}}.$$

Hence,

$$\mathbb{E}[\delta_\square(W, \mathbb{G}(k, W)) \leq \mathbb{E}(\delta_\square(W, \mathbb{H}(k, W))) + \mathbb{E}[\delta_\square(\mathbb{H}(k, W), \mathbb{G}(k, W))] \leq \frac{27}{\sqrt{\log k}}.$$

Let $f(G) = |G|\delta_\square(G, W)$, and apply Theorem 5.7 with this function $f$ yields the conclusion. $\square$

Using the above lemma, we can see the following by taking a sequence $(G_n)_{n\in\mathbb{N}}$ of graphs with $G_n = \mathbb{G}(n, W)$: For given graphon $W$, there exists a sequence of finite graphs having limit $W$. Now we prove the following inverse counting lemma.

**Lemma 5.13** (Inverse counting lemma)**.** *Let $k$ be a positive integer, let $U, W \in \mathcal{W}_0$, and assume that for every simple graph $F$ on $k$ nodes, we have $|t(F, U) - t(F, W)| \leq 2^{-k^2}$. Then $\delta_\square(U, W) \leq \frac{100}{\sqrt{\log k}}$.*

*Proof.* Inclusion-exclusion implies that $\mathbb{P}[\mathbb{G}(k, U) = F] = \sum_{F' \supseteq F}(-1)^{e(F')-e(F)}t(F', W)$. This implies that for each $k$-vertex graph $F$,

$$|\mathbb{P}[\mathbb{G}(k, U) = F] - \mathbb{P}[\mathbb{G}(k, W) = F]| \leq 2^{\binom{k}{2}}2^{-k^2} = 2^{-\binom{k+1}{2}}. \tag{5.1}$$

Define

$$\mathcal{F}_U = \{F : |F| = k, \delta_\square(U, F) \leq \frac{30}{\sqrt{\log k}}\} \text{ and } \mathcal{F}_W = \{F : |F| = k, \delta_\square(W, F) \leq \frac{30}{\sqrt{\log k}}\}.$$

By Lemma 5.12, we have

$$\sum_{F \in \mathcal{F}_U} \mathbb{P}[\mathbb{G}(k, U) = F] \geq 1 - 2\exp(-k/(2\log k)) \text{ and}$$
$$\sum_{F \notin \mathcal{F}_W} \mathbb{P}[\mathbb{G}(k, W) = F] \leq 2\exp(-k/(2\log k)).$$

If $\mathcal{F}_U \cap \mathcal{F}_W$ is not empty and contains a graph $F$, then we have

$$\delta_\square(U, W) \leq \delta_\square(U, F) + \delta_\square(W, F) \leq \frac{30}{\sqrt{\log k}} + 0 + \frac{30}{\sqrt{\log k}} \leq \frac{100}{\sqrt{\log k}}.$$

Hence, it suffice to prove that $\mathcal{F}_U \cap \mathcal{F}_W$ is not empty. Assume $\mathcal{F}_U \cap \mathcal{F}_W = \emptyset$. Then

$$\sum_{F \in \mathcal{F}_U} (\mathbb{P}[\mathbb{G}(k, U) = F] - \mathbb{P}[\mathbb{G}(k, W) = F]) \geq 1 - 4\exp(-\frac{k}{2\log k}) \geq \frac{1}{2}.$$

[2] Note that we have $\mathbb{E}[r_1] = 0$, so $\mathbb{E}[(\frac{k}{m} + kr_1)^2] = k^2/m^2 + 2\mathbb{E}[r_1]k^2/m + k^2\mathbb{E}[r_1^2] = k^2/m^2 + k^2\mathbb{E}[r_1^2]$. Also, we have $\mathbb{E}[(\frac{k}{m} + kr_1)^2] = \mathbb{E}[|V_i \cap S|^2] = Var[|V_i \cap S|] + \mathbb{E}[|V_i \cap S|]^2 = k \cdot (1/m) \cdot (1 - 1/m) + k^2/m^2$. This implies that $\mathbb{E}[r_1^2] = \frac{m-1}{km^2}$.

As $|\mathcal{F}_U| \le 2^{\binom{k}{2}}$, this implies that there exists $F$ such that $\mathbb{P}[\mathbb{G}(k,U) = F] - \mathbb{P}[\mathbb{G}(k,W) = F] \ge 2^{-\binom{k}{2}-1} > 2^{-\binom{k+1}{2}}$, a contradiction to (5.1). Therefore, $\mathcal{F}_U \cap \mathcal{F}_W$ is not empty and this proves the theorem. □

This implies the following.

**Corollary 5.14.** *Two graphons $U, W \in \mathcal{W}_0$ are weakly isomorphic if and only if $\delta_\square(U,W) = 0$.*

## 6. Bounded degree graphs

We again want to define probability distribution on a sparse graph $G$ using random sampling. However, as $G$ is a sparse graph, the homomorphism density $t(F, G_n)$ always tends to zero as $|G_n|$ tends to infinity. Hence, we define another random sampling procedure which makes more sense for graphs with maximum degree bounded by constant. Let $\Delta$ be a constant. From now on, we will only consider graphs with maximum degree at most $\Delta$.

**Definition 6.1.** *Let $\mathfrak{B}_r$ be the set of rooted graph $F = (F', u)$ where every vertex in $F'$ has distance at most $r$ from $u \in V(F)$. We call each rooted graph in $\mathfrak{B}_r$ an $r$-ball. We say that two rooted graphs $(F', u')$ and $(F'', u'')$ are isomorphic if there exists a graph isomorphism $f : F' \to F''$ where $f(u') = u''$. Let $\deg(F) = \deg_{F'}(u)$ be the degree of the root.*

As $F'$ above has maximum degree bounded, $\mathfrak{B}_r$ is a finite set.

**Definition 6.2.** *For given graph $G$ and $F = (F', u) \in \mathfrak{B}_r$, let $\rho_G(F) = \rho_{G,r}(F)$ be the probability that the $r$-neighborhood $B_r(v)$ in $G$ with the root $v$ is isomorphic to $(F', u)$ for a vertex $v \in V(G)$ chosen uniformly at random.*

**Definition 6.3.** *A sequence of graphs $(G_n)_{n\in\mathbb{N}}$ of bounded degree is locally convergent if the $r$-neighborhood density $\rho_{G_n,r}(F)$ converges for every $r$ and every $r$-ball $F$.*

As $\rho_{G_n,r}$ is a probability distribution on $\mathfrak{B}_r$, and $\rho_{G_1,r}, \rho_{G_2,r}, \dots$ converges to a probability distribution $\sigma_r$, we can consider $\sigma = (\sigma_1, \sigma_2, \dots)$ as the limit of a convergent sequence $G_1, G_2, \dots$. However, as we have constructed graphons as a more descriptive limit object, we want to construct a more descriptive limit object of this convergent sequence. In order for this, we first need to know more properties of the $\sigma = (\sigma_1, \sigma_2, \dots)$.

First, this distribution is consistent in the following sense. We select a random $r$-ball using the distribution $\sigma_r$ and delete all vertices of distance $r$ from the root, then we get an $(r-1)$-ball in $\mathfrak{B}_{r-1}$. This provides the distribution $\sigma_{r-1}$.

Second, this is 'involution invariant' in the following sense. An $r$-ball $(F, u)$ chosen from $\sigma_r$ and consider a neighbor $v$ of the root $u$ and consider $(r-1)$-ball around $v$. This must also yields the distribution similar to $\sigma_{r-1}$ except a vertex with higher degree is more likely to be chosen. This motivates to define the following for each $F \in \mathfrak{B}_r$:

$$\sigma^*_r(F) = \frac{\deg(F)\sigma_r(F)}{\sum_{H\in\mathfrak{B}_r} \deg(H)\sigma_r(H)}.$$

Here, this can be considered as the probability of getting $F = (F', u)$ when we choose an edge $e = uv$ of $G$ uniformly at random, and choose an endpoint $u$ and taking an $r$-neighborhood of $u$. We can encode this information by the following definition.

**Definition 6.4.** *Select a random $r$-ball $F$ from $\sigma^*_r$ and an edge $uv$ from the root $u$ of $F$ uniformly at random. We delete all vertices of distance $r$ from $u$ to obtain a rooted graph $F_1$ with root $u$ and root edge $uv$, and we delete all vertices at distance more than $r-1$ from $v$ to obtain a rooted graph $F_2$ with the root $v$ and the root edge $uv$. If each of $F_1$*

*and $F_2$ has the same distribution with $\sigma^*_{r-1}$, then we say that the sequence $(\sigma_1, \sigma_2, \dots)$ is involution invariant.*

As we consider pointwise convergence of the distributions $(\rho_{G_1,r}, \rho_{G_2,r}, \dots)$, every convergent graph sequence gives rise to an involution invariant and consistent probability measure $\sigma$ on $\mathfrak{B}_r$.

In order to define a limit of convergent sequence of graphs, we want to consider a graph with infinitely many vertices. Recall that when we consider the sampling distribution for finite graphs $G$ above, we choose one vertex uniformly at random with probability $1/|G|$ and consider its $r$-neighborhood. However, for infinite graphs, the probability of choosing one vertex is not so important as it is zero if the choice is uniform. Instead, for a certain subset $A$ of vertices, we need to consider the probability of choosing the vertex from the set $A$. Hence, we need to instead consider an infinite graph equipped with a measure $\lambda$ on the vertex set so that we can consider a random sampling procedure where $\lambda(A)$ is the probability of choosing a random vertex from the set $A$. Of course, we can't expect to define a nontrivial measure $\lambda(A)$ for all sets $A \in 2^V$ for infinite set $V$ as we have learned in analysis (we know that the axiom of choice provides a non-measurable set). So, we will restrict our attention to certain 'good' sets. The following is what we will do.

(1) Define Polish space which will be the vertex set of our infinite graph $G$.
(2) Define Borel sets, which will be the 'good' subset of $V(G)$ whose measures can be computed.
(3) Define Borel graphs, which will allow us to use graph operations on Borel subsets of $V(G)$, in particular, the neighborhood $N(A)$ of a Borel set is again Borel.
(4) Define Graphings, which will allow us to use graph operations on Borel subsets while predicting how the measure of the sets changes during the operations.

(1) Possibly, we will work on a graph with vertex set $V$ where $V$ is a topological space like a unit interval $[0, 1]$. This motivates to work on Polish space, which will be the vertex set of our infinite graphs.

(2) Of course, we want to be able to compute the measure on all intervals inside $[0, 1]$, and their unions and intersections and etc. This motivates to use the definition of Borel sets, which is a smallest good collection containing all those intervals and their countable unions and countable intersections and etc.

(3) and (4) We want to be able to consider neighborhood of a given Borel set, and compute the measure of the neighborhood. One of the reason for this is to be able to make sure that the distribution from the infinite graph becomes involution invariant. Recall that, in finite graphs, we randomly choose a vertex $u$ proportional to its degree and an incident edge $uv$, and compare this with the probability of choosing the root $v$ with an edge $uv$.

For infinite graphs, we need to compute the probability of choosing $u \in A$ and an incident edge $uv$ with $v \in B$ for certain measurable sets $A$ and $B$ and we need to compare this to the probability of choosing $v \in B$ and an incident edge $uv$ with $u \in A$. In order for this, we need to make sure that the neighborhood $N_G(A)$ of a Borel set $A$ is also Borel in our graph $G$, so we can compute the measure of $N_G(A) \cap B$. This motivates our definition of Borel graphs. Moreover, the involution invariance can be stated as follows.

$$\int_A \deg_B(x) d\lambda(X) = \int_B \deg_A(x) d\lambda(x).$$

This motivates the definition of graphings.

## 7. Borel graphs

A subset $D \subseteq X$ of a topological space $X$ is *dense* if it meets every nonempty open set. A space $X$ admitting a countable dense set is called *separable.* If $X$ is metrizable, then $X$ is separable if and only if $X$ has a countable basis.

On a metric space $(X, d)$, a *Cauchy sequence* is a sequence $(x_n)$ of elements of $X$ such that $\lim_{m,n} d(x_m, x_n) = 0$. We call $(X, d)$ *complete* if every Cauchy sequence has a limit in $X$. We say that a topological space $X$ is *completely metrizable* if it admits a compatible metric $d$ such that $(X, d)$ is complete.

**Definition 7.1.** *A topological space $X$ is* **Polish** *if it is completely metrizable and separable.*

We call $(X, \mathcal{B}(X))$ *Standard Borel Space (SBS)* if $X$ is a Polish space. It is known that any SBS are either isomorphic to $(X, 2^X)$ with countable $X$ or isomorphic to $(I, \mathcal{B}(I))$ where $I \in \{[0, 1], (0, 1), [0, 1)\}$ is an interval. It is not difficult to see that any Polish space has a countable basis. Also we have the following.

**Proposition 7.2.** *The following spaces are also Polish: All closed/open subsets, all open subsets of Polish spaces, a disjoint union of Polish spaces and a product of Polish spaces.*

We say that a collection $\mathcal{A}$ an *$\sigma$-algebra on a set $X$* if it is a family of subsets containing $\emptyset$ and closed under complements and countable unions. For a family $\mathcal{F}$ of subsets of $X$, let $\sigma(\mathcal{F})$ be the smallest $\sigma$-algebra containing $\mathcal{F}$. This is well-defined as $2^X$ is a $\sigma$-algebra, and arbitrary intersection of $\sigma$-algebras are also $\sigma$-algebra.

**Definition 7.3.** *A measure space is a pair $(X, \mathcal{F}_X)$ when $X$ is a set and $\mathcal{F}_X$ is a $\sigma$-algebra on $X$. A map $f : X \to Y$ is called measurable if $f^{-1}(A) \in \mathcal{F}_X$ for all $A \in \mathcal{F}_Y$.*

**Definition 7.4.** *For a topological space $X$ with topology $\mathcal{T}$, we write $\mathcal{B}(X) = \sigma(\mathcal{T})$ and call it Borel $\sigma$-algebra.*

The following graphs give some idea on the graphs we will deal with.

**Example 7.1.** *Fix $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and consider a set $V = [0, 1)$ and a transformation*

$$T : x \mapsto (x + \alpha) \pmod 1$$

*from $V$ to $V$. Let $R_\alpha := (V, E)$ where we have $E = \{xT(x) \in \binom{V}{2}\}$*

Consider the above $V$ as a unit circle, by identifying 0 and 1. This consists of uncountably many components where each component is a bi-infinite path. This is different from a countable union of infinite paths, as we have a nontrivial measure equipped with the vertex set. It is known that such a graph cannot be partitioned into two Borel independent sets.

**Example 7.2.** *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, $V = [0, 1) \times \{0, 1\}$ and let $T(x, i) = (x + \alpha \mod 1,\ 1 - i)$ be a transformation from $V$ to $V$. Let $D_\alpha := (V, E)$ where $E = \{vT(v) : v \in V\}$.*

Note that $D_\alpha$ is again a 2-regular acyclic graph. To visualise this, consider two disjoint unit cycles corresponding to $[0, 1) \times \{i\}$ with $i \in \{0, 1\}$. As there's no edges inside each cycle, $D_\alpha$ can be colored with 2 colors in such a way that each cycle is monochromatics, hence a coloring with Borel color classes. So $R_\alpha$ and $D_\alpha$ are not equivalent under Borel coloring.

In this section, we will prove that the definition of Borel graphs below provides the following properties:

(a) $N(A)$ is a Borel set when $A$ is a Borel set
(b) Borel graph $G$ can be decomposed into finitely many Borel matchings.

(a) is as we intended to prove and (b) will be useful to deal with the properties we mentioned in (4) in the previous section.

**Definition 7.5.** *A graph $G = (V, \mathcal{B}, E)$ is a Borel graph if the followings hold.*

(1) *$(V, \mathcal{B})$ is standard Borel space, i.e. there exists a Polish topology $\mathcal{T}$ such that $\mathcal{B} = \sigma(\mathcal{T})$.*
(2) *$E \subseteq V \times V$ is Borel and $(x, y) \in E$ if and only if $(y, x) \in E$.*

Again, we always assume that every graph we deal with has maximum degree at most $\Delta$. In the second condition, we mean that say that an edge set $E$ is Borel in the space $\mathcal{B}(V) \times \mathcal{B}(V) = \mathcal{B}(V^2)$. The followings are some examples of Borel graphs.

**Example 7.3.**

(1) *Every finite graphs $G$ are Borel graphs.*
(2) *$R_\alpha$ and $D_\alpha$ are Borel graphs.*

For the second example, note that $T$ is continuous. Hence, $E = f^{-1}(\{0\})$ where $f(x, y) = y - T(x)$ is a closed set in $V^2$ thus Borel. So $R_\alpha$ and $D_\alpha$ are Borel graphs. We will use the following theorem to prove the next lemma.

**Theorem 7.6** (Lusin)**.** *For a continuous countable to 1 function $f$ and a Borel set $A$, $f(A)$ is again Borel.*

**Lemma 7.7.** *Let $V$ be a Polish space and $G = (V, E)$ be a graph with $\Delta(G) \le \Delta < \infty$. Then $G$ is a Borel graph if and only if $N(A) = \{y : \exists x \in A, xy \in E\}$ is Borel for all Borel subset $A \subseteq V$.*

*Proof.* ($\Leftarrow$) Suppose that for all Borel sets $A$, $N(A)$ is Borel. Consider a countable base $\{U_n\}$ of the Polish space $V$. Then we claim

$$V^2 \setminus E = \bigcup_n (U_n \times (N(U_n)^c)). \tag{7.1}$$

As $\supseteq$ is clear, we show $\subseteq$. Consider $(x, y) \in V^2 \setminus E$ and let $z_1, \ldots, z_s$ be the neighbors of $y$. As $x \notin \{z_1, \ldots, z_s\}$, there exists $\varepsilon > 0$ such that

$$B_\varepsilon(x) \cap \{z_1, \ldots, z_s\} = \emptyset.$$

Thus there exists $i$ such that $x \in U_i$ but $z_j \notin U_i$ for all $j \in [s]$ and $y \in N(U_i)^c$. This proves (7.1).

As $U_n$ is an open set, the assumption implies that $N(U_n)^c$ is a Borel set. So (7.1) implies that $V^2 \setminus E$ is a countable union of Borel sets, thus it is Borel, hence its complement, $E$ is also Borel.

($\Rightarrow$) Let $A$ be a Borel set. Then we have

$$N(A) = pr_2((A \times V) \cap E),$$

where $pr_2$ is a projection map to the second coordinate, which is a continuous map. Note that every $x \in V$ has at most $\Delta$, finitely many preimage. Hence, $N(A)$ is a image of $(A \times V) \cap E \in \mathcal{B}(V^2)$ via a continuous countable to 1 map and Theorem 7.6 implies that $N(A)$ is Borel. □

Let's consider one more example.

**Example 7.4.** *Let $\Gamma = \langle S \rangle$ be a group generated by a finite set $S$ with $S = S^{-1}$. Let $\Gamma$ act on a Polish space $X$ in a Borel way. In other words,*

- *For all $\gamma \in \Gamma$, the map $x \to \gamma.x$ is Borel, meaning that preimage of every Borel set under this map is Borel. (Equivalently, $(\gamma, x) \to \gamma.x$ is a Borel map when we give $\Gamma$ is countable sets with discrete topology.)*

*Then $G = (X, E)$ is a Borel graph where $E = \{\{x, \gamma.x\} : x \in X, \gamma \in S\} \setminus \{(x, x) : x \in X\}$. Note that $\Delta(G) \leq |S|$.*

Let $A$ be a Borel subset of $X$ and let $T_\gamma$ be the map $x \to \gamma.x$. Then

$$N(X) = \bigcup_{\gamma \in S} \{\gamma x : x \in X\} = \bigcup_{\gamma \in S} \{\gamma^{-1} x : x \in X\} = \bigcup_{\gamma \in S} T_\gamma^{-1}(X).$$

As $\Gamma$ act on $X$ in a Borel way, the set $T_\gamma^{-1}(X)$, the preimage of $T_\gamma$, is Borel. Hence $N(X)$ is a finite union of Borel sets, thus a Borel set. By Lemma 7.7, such a graph is actually Borel graph. Note that we know that the function is Borel if and only if its corresponding graph is Borel.

**Theorem 7.8** (Kechris-Solecki-Todorcevic)**.** *For a Borel graph $G$, we have $\chi_{\mathcal{B}}(G) \leq \Delta(G) + 1$.*

*Proof.* Consider a countable base $\{U_n\}$ of $(V, \mathcal{T})$. We first define the following assignment of a sequence to each point. For each $x \in V$, let

$$f(x) := (\mathbb{1}_{U_1}(x), \mathbb{1}_{U_2}(x), \dots) \in 2^{\mathbb{N}}.$$

As for any $x \neq y \in V$, there exists a base $U_i$ containing $x$ but not $y$. This shows that $f$ is an injective function. For $p = (p_1, \dots, p_k) \in 2^{<\infty} = \bigcup_{k \geq 0} \{0, 1\}^k$, let $W_p \subseteq 2^{\mathbb{N}}$ be the collection of sequences whose first $k$ terms are same with $p$.

We consider $2^{\mathbb{N}}$ as a countable product of discrete topologies, then $\{W_p : p \in 2^{<\infty}\}$ is a base for $2^{\mathbb{N}}$. For all $p = (p_1, \dots, p_k) \in 2^{<\infty}$, the set

$$X_p := f^{-1}(W_p) = \bigcap_{i : p_i = 1} U_i \cap \bigcap_{i : p_i = 0} U_i^c$$

is a Borel set. Hence $f$ is a Borel map while it is not necessarily continuous.

Now we first define a coloring with countably many colors and then use this coloring to obtain the desired coloring. Let $\ell : V \to 2^{<\infty}$ by $\ell(x) := p$ where $p$ is a shortest sequence with $f(x) \in W_p$ but $f(N(x)) \cap W_p = \emptyset\}$. For example, $f(x) = (0, 1, 1, 0, \dots)$ and $N(x) = \{y, z\}$ with $f(y) = (0, 1, 0, 0, \dots)$ and $f(z) = (1, 0, 1, 0, \dots)$, then we have $\ell(x) = (0, 1, 1)$.

Note that the injectivity of $f$ implies that $\ell$ is well-defined. Also we have the following claim stating that $\ell$ is a Borel function.

**Claim 2.** *For all $p \in 2^{<\infty}$, we have $\ell^{-1}(p) \in \mathcal{B}$.*

*Proof.* We use induction of the length of $p$. If $p = \emptyset$, it is easy as $\ell^{-1}(p) = \emptyset$. Assume that the claim holds for all $p$ with length at most $i$. For a sequence $p$ of length $i + 1$, we have

$$\ell^{-1}(p) = X_p \setminus (N(X_p) \cup \bigcup_{q < p} \ell^{-1}(q)).$$

where $q < p$ implies that $q$ is a proper prefix of $p$. Note that $N(X_p)$ is a Borel set by Lemma 7.7 and $\ell^{-1}(q)$ are Borel sets by Induction hypothesis. Hence $\ell^{-1}(p)$ is also a Borel set. This proves the claim. $\square$

By the claim and renumbering, we have a Borel partition $V_1 \cup V_2 \cup \dots$ of $V$ into countably many independent sets. Now we greedily construct the desired coloring $c : V \to [\Delta + 1]$. Assume that we have coloring $c : V_1 \cup \dots \cup V_i \to [\Delta + 1]$. For every $x \in V_{i+1}$, let

$$c(x) := \min\{k \in [\Delta + 1] : k \notin c(N(x))\}.$$

As $\Delta(G) \leq \Delta$, the function $c$ is well-defined on every point in $V_{i+1}$. This defines $c$ on every $V_i$, hence on every $V$. For each $i \in \mathbb{N}$ and $k \in [\Delta + 1]$, let $V_{i,k} := c^{-1}(k) \cap V_i$.

We claim that each $V_{i,k}$ is a Borel set. We prove this by using induction on $(i,k)$, where $(i',k') < (i,k)$ if either $i' < i$ or $i' = i$ and $k' < k$. Then $V_{1,1} = V_1$ is a Borel set. Also

$$V_{i,k} = V_i \setminus \bigcup_{k'<k} \bigcup_{i'<i} N(V_{i',k'})$$

is again Borel by Lemma 7.7. This shows that every $V_{i,k}$ is Borel sets and $c^{-1}(k) = \bigcup_{i=1}^{\infty} V_{i,k}$ is a Borel set. Hence we obtain a Borel coloring, hence $\chi_{\mathcal{B}}(G) \leq \Delta + 1$. □

Similarly, we can consider edge-colorings of Borel graphs.

**Definition 7.9.** *Borel chromatic index $\chi'_{\mathcal{B}}(G)$ is the minimum $k \in \mathbb{N}$ such that there exists a Borel edge coloring $c : E \to [k]$. In other words, we have $c(x,y) = c(y,x)$ for all $xy \in E$ and $c$ gives a Borel partition $E = E_1 \cup E_2 \cup \cdots \cup E_k$ to matchings.*

**Theorem 7.10.** *For a Borel graph $G$, we have $\chi'_{\mathcal{B}}(G) \leq 2\Delta(G) - 1$.*

*Proof.* One possible way to prove this is to apply Theorem 7.8 to the line graph. One can check that $(E, \mathcal{B}(V^2) \cap E)$ is a Standard Borel space. Also one can check that $\{ee' \in E \times E : e \cap e' \neq \emptyset\}$ is Borel.

Another way to prove this is to consider a square $G^2$ of $G$ where two vertices are adjacent in $G^2$ if and only if the distance between them is at most two in $G$. Then we obtain a Borel partition $V = V_1 \cup \cdots \cup V_k$ where $E = \bigcup_{ij \in \binom{k}{2}} E_{ij}$ where $E_{ij} = E \cap ((V_i \times V_j) \cup (V_j \times V_i))$ is a Borel matching. Hence, we can apply parallel greedy algorithm over $ij$ as we did in the proof of Theorem 7.8. □

**Definition 7.11.** *For given set $S$ of bijection on $V$, let* graph$(S)$ *be the Borel graph with vertex set $V$ and edge set*

$$\{x\phi(x) : x \in V, \phi \in S, \phi(x) \neq x\}.$$

**Theorem 7.12.** *Let $G = (V,E)$ be a Borel graph. Then there exists a finite set $S$ with $|S| \leq 2\Delta - 1$ satisfying the following.*

*(a) Each $\phi \in S$ is a Borel involutions. In other words, each $\phi \in S$ satisfies $\phi \circ \phi = Id_V$.*
*(b) $G = graph(S)$.*

*Proof.* By Theorem 7.10, there exists a Borel partition $E = \bigcup_{i=1}^{k} E_i$ with $k = 2\Delta - 1$. For each $i \in [k]$, let $\phi_i : V \to V$ by

$$\phi_i(x) = \begin{cases} y & \text{if } xy \in E_i \\ x & \text{otherwise.} \end{cases}$$

Clearly each $\phi_i$ is an involution. We still need to prove that each $E_i$ is a Borel set in $V \times V$.

**Claim 3.** *Let $X, Y$ be Standard Borel spaces, and let $f : X \to Y$. Then $f$ is Borel if and only if $F = \{(x, f(x)) : x \in X\} \in \mathcal{B}(X \times Y)$.*

*Proof.* Proof is similar for the proof of Lemma 7.7. Take a countable base $\{U_n\}$ of $Y$ and show that $F^c = \bigcup_n (f^{-1}(U_n) \times U_n^c)$ to conclude $(\Rightarrow)$. For the other direction, use Theorem 7.6. □

It is easy to use this claim to conclude that each $E_i$ is a Borel set. So $S = \{\phi_1, \ldots, \phi_k)$ suffice. □

## 8. Graphings

Now we will consider the objects which can be considered as a limit of bounded degree graphs.

**Definition 8.1.** *We say that $\mathcal{G} = (V, \mathcal{B}, E, \lambda)$ is a graphing if $(V, \mathcal{B}, E)$ is a Borel graph and $\lambda$ is a probability measure on $(V, \mathcal{B})$ such that*

*there exists Borel bijections $\psi_1, \dots \psi_k : V \to V$ such that $E = \mathrm{graph}(\psi_1, \dots, \psi_k)$ and each $\psi_i$ is measure preserving. I.e. for all $A \in \mathcal{B}$, $\psi_i^{-1}(A) \in \mathcal{B}$ has the same $\lambda$-measure as $A$.* (8.1)

For a graphing $\mathcal{G}$, there could be more than one choice of $\{\psi_1, \dots, \psi_k\}$ generating $E$. However, the following lemma shows that any choice of $\{\psi_1, \dots, \psi_k\}$ consists of measure preserving bijections as long as there is one such choice of measure preserving bijections. Here $[x]_E$ denote the component containing $x$.

**Lemma 8.2.** *Let $\mathcal{G}$ be a graphing and $A, B \in \mathcal{B}(V)$ and $f : A \to B$ be Borel bijection such that for all $x \in A$, $f(x) \in [x]_E$. Then $f$ is measure preserving.*

*Proof.* Let $\Gamma =< \psi_1, \dots, \psi_k >$ be the formal free group generated by symbols $\psi_1, \dots, \psi_k$ and their inverses. We enumerate $\Gamma = \{\gamma_0, \gamma_1, \dots\}$ with $|\gamma_0| \leq |\gamma_1| \leq \dots$. In particular, we have $\gamma_0 = \emptyset$. For each $i \geq 0$, define

$$D_i = \{x \in A : f(x) = \gamma_i x\} \setminus (\bigcup_{j<i} D_j).$$

Each $D_i$ is a Borel set. If $\gamma_i = \psi_{k_1}^{\varepsilon_1} \circ \dots \circ \psi_{k_s}^{\varepsilon_s}$, then $f\mid_{D_i}$ is $\psi_{k_1}^{\varepsilon_1} \circ \dots \circ \psi_{k_s}^{\varepsilon_s}\mid_{D_i}$, a measure preserving function. As $D_i$ are disjoint and forms a partition of $A$, $f$ is also measure preserving. This finishes the proof. $\square$

**Example 8.1.** *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and consider $R_\alpha = \mathrm{graph}(x \mapsto x + \alpha \pmod 1)$ with $V = [0, 1)$. Note that the Lebesgue measure $\lambda$ is preserved by this map.*

As a graphing has probability measure, we can sample a random vertex according to this probability distribution. Picking a random vertex $v$ and considering an $r$-neighborhood $B_r(v)$, we can get an $r$-ball in $\mathfrak{B}_r$.

We say that a sequence $(G_n)$ of graphs converge (in a sense of Benjamini-Schramm) to a graphing $\mathcal{G}$ if for all $r \in \mathbb{N}$ and any rooted graph $F$,

$$\mathop{\mathbb{P}}_{x \text{ uniform in } V(G_n)}[B_r(x) \simeq F] \to \mathop{\mathbb{P}}_{x \sim (V,\lambda)}[B_r(x) \simeq F].$$

Note that such a graphing $\mathcal{G}$ may not be unique.

**Example 8.2.** *Let $G_n = C_n$ be the $n$-cycle. Then $G_n \xrightarrow{BS} R_\alpha$ for any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.*

**Example 8.3.** *Let $G_n$ be $n \times n$ grid. A possible graphing is $\mathcal{G} = ([0, 1]^2, \mathcal{B}, \mathrm{graph}(\psi_1, \psi_2), \lambda)$ where $\psi_1 : (x, y) \mapsto (x + \alpha, y)$ and $\psi_2 : (x, y) \mapsto (x, y + \beta)$ where $\alpha, \beta$ are generic irrational number, algebraically independent each other.*

We say that $S = \{\phi_1, \dots, \phi_k\}$ is an involutive measure preserving family for $\mathbf{G}$ if $\mathbf{G} = \mathrm{graph}(S)$ and each $\phi_i$ is measure preserving involutions.

**Theorem 8.3.** *For each graphing $\mathbf{G}$, there is an involutive measure preserving family with at most $2\Delta - 1$ bijections.*

*Proof.* Use Theorem 7.10 to find $2\Delta - 1$ Borel matchings. Each matching yields a involutive Borel bijections. Lemma 8.2 implies that each such involution is measure preserving. $\square$

Using this, it is easy to check that the following holds for graphings.

$$\int_A \deg_B(x) d\lambda(X) = \int_B \deg_A(x) d\lambda(x).$$

In fact, any Borel graph with this condition is a graphing. Hence this condition can be taken as the definition of graphing.

## 9. Limit of sequences of bounded degree graphs

Our aim on this section is to show that any locally convergent sequence of graphs with bounded degree has a graphing as a limit. For such a sequence, we have a limiting sampling distribution $\sigma = (\sigma_1, \sigma_2, \dots)$. We will do the following.

(1) Define a Borel graph $\mathbf{H}$.
(2) Define sampling distributions on given graphings.
(3) Find a way to give measure to $\mathbf{H}$ to make it a graphing which has sampling distribution $\sigma$, hence it can be considered as a limit of the given sequence.

We wish to define a graphing. As seen before, a graphing in general consists of infinitely many components where each component is a countable graph. So it would be convenient to define one graph having all possible countable graphs. Our aim is to define a Borel graph $\mathbf{H}$ and give measures according to the given distribution $\sigma = (\sigma_1, \sigma_2, \dots)$. One issue is that when we have two components, $H_1, H_2$ of infinite graphs where $H_1$ has no automorphism while $H_2$ is an infinite path. Then some 'measure' of $\sigma_r(P_{2r-1})$ must be distributed over all vertices of $H_2$, while some 'measure' of $\sigma_r(F)$ must be distributed over only finitely many vertices of $H_1$. This can possibly make 'measure distribution' complicated. To avoid this issue, instead of choosing a vertex at random, we build a rooted graph at random using $\sigma = (\sigma_1, \sigma_2, \dots)$ in such a way that the $r$-ball around the root of the built graph is the graph from $\sigma_r$. In other words, each rooted graph will be the vertex we can choose. For this, we define the following set of all connected countable rooted graphs.

**Definition 9.1.** *Let $\mathfrak{B}^\bullet$ denote the set of connected countable graphs with maximum degree at most $\Delta$ that have a specified vertex called their root. For each $H \in \mathfrak{B}^\bullet$ we write $\mathrm{root}(H)$ to denote the root of $H$ or we write $H = (H', v)$. Let $\deg(H)$ be the degree of its root.*

We consider two graphs in $\mathfrak{B}^\bullet$ the same if there is an isomorphism between them preserving the root. For $H \in \mathfrak{B}^\bullet$, let $B_{H,r} \in \mathfrak{B}_r$ be the rooted graph induced by the neighborhood of the root with radius $r$. For every $r$-ball $F$, let $\mathfrak{B}^\bullet_F$ denote the set of those graphs $H \in \mathfrak{B}^\bullet$ for which $B_{H,r} \simeq F$ as rooted graphs.

Let $\sigma = (\sigma_1, \sigma_2, \dots)$ be any probability distribution on $(\mathfrak{B}^\bullet, \mathfrak{A})$. We choose a graph $F \in \mathfrak{B}^\bullet$ in the following way: we choose $F_1 \in \mathfrak{B}_1$ with respect to the measure $\sigma_1$, and $F_2$ with respect to $\sigma_2$ conditioning on $B_{F_2,1} = F_1$, and similarly, choose $F_{r+1}$ with respect to $\sigma_{r+1}$ conditioning on $B_{F_{r+1},r} = F_r$ for all $r \geq 1$. This sequence $(F_1, F_2, \dots)$ yields an infinite graph $F \in \mathfrak{B}^\bullet$. This procedure provides a probability distribution on $\mathfrak{B}^\bullet$, which again call this distribution $\sigma$.

**Definition 9.2.** *We define a graph $\mathbf{H}$ on the vertex set $\mathfrak{B}^\bullet$ as follows: For each $(H, v) \in \mathfrak{B}^\bullet$ and $vv' \in E(H)$, we connected $(H, v)$ and $(H, v')$ in $\mathbf{H}$. We call $\mathbf{H}$ "graph of graphs".*

If $H$ has no automorphism as unrooted graphs, then the $r$-neighborhood of a rooted graph $H$ in $\mathbf{H}$ is isomorphic to the $r$-neighborhood of the root in $H$. In order to make this a Borel graph, we define the following metric: for two graphs $H_1, H_2 \in \mathfrak{B}^\bullet$, we define the ball distance by

$$d^\bullet(H_1, H_2) = \inf\{2^{-r} : B_{H_1,r} \simeq B_{H_2,r}\}.$$

This turns $\mathfrak{B}^\bullet$ into a metric space where $\mathfrak{B}^\bullet_F$ forms an open basis. Moreover the space $(\mathfrak{B}^\bullet, d^\bullet)$ is compact. Moreover the graph $\mathbf{H}$ is a Borel graph with respect to the above topology and sigma-algebra $\mathfrak{A}$ obtained from the ball distance.

**Proposition 9.3.** *The graph of graphs* $\mathbf{H}$ *is Borel.*

*Proof.* We prove that $E(\mathbf{H})$ is closed with the topology given from the metric.

Let $H_1H_2 \notin E(\mathbf{H})$ and let $F_1, \dots, F_d$ be the neighbors of $H_1$ distinct from $H_1$, where all these rooted graphs is isomorphic to some graph $H$ when we forget the root. As $H_2$ is not adjacent to $H_1$, there exists $r \geq 1$ such that $B_{H_2,r}$ is not isomorphic to any of $B_{F_1,r}, B_{F_2,r}, \dots, B_{F_d,r}$.

We claim that if $H_1', H_2' \in \mathfrak{B}^\bullet$ such that $d^\bullet(H_1, H_1') < 2^{-r}$ and $d^\bullet(H_2, H_2') < 2^{-r}$ then $H_1'H_2' \notin \mathfrak{B}^\bullet$.

Indeed, assume $H_1'H_2' \in E(\mathbf{H})$. We know $B_{H_1,r+1} \simeq B_{H_1',r+1}$. Let $x = \mathrm{root}(H_1), y = \mathrm{root}(H_2), u = \mathrm{root}(H_1')$ and $v = \mathrm{root}(H_2')$. Assume that $y_1 = \mathrm{root}(F_1)$ corresponds to $v$ in the isomorphism $B_{H_1,r+1} \simeq B_{H_1',r+1}$. Then we have $B_{H_1,r}(y_1) \simeq B_{H_1',r}(v)$. But $d^\bullet(H_2, H_2') < 2^{-r}$ implies that $B_{H_2,r}(u) \simeq B_{H_2',r}(v) \simeq B_{H_1,r}(y_1) = B_{F_1,r}(y_1)$, a contradiction to our choice of $r$.

Therefore, for any $H_1H_2 \notin E(\mathbf{H})$, there exists an open set $B_{2^{-r}}(H_1) \times B_{2^{-r}}(H_2)$ containing $H_1H_2$ lying outside $E(\mathbf{H})$. Hence, $E(\mathbf{H})$ is a closed set, hence a Borel set. □

We have a Borel graph $\mathbf{H}$. To make this a graphing, we need to define measure on $\mathfrak{B}^\bullet$. There will be many ways to give measures to this Borel graph. Let $\sigma$ be a probability distribution on $\mathfrak{B}^\bullet$.

For each $d \in \mathbb{N}$, the set $\deg^{-1}(d) = \{F \in \mathfrak{B}^\bullet : \deg(F) = d\}$ is the union of finitely many $\mathfrak{B}^\bullet_F$, it is an open set. Hence, deg is a measurable function. As we did with finite graph cases, we will consider the following new measure.

$$\sigma^*(A) = \frac{\int_A \deg(F) d\sigma}{\int_{\mathfrak{B}^\bullet} \deg(F) d\sigma}.$$

Note that this is at most $\Delta$ and the denominator is nonzero unless the measure is an indicator measure on $K_1$. In this case, we set $\sigma^* = \sigma$.

Now we will consider 'involution invariance' as we did for finite graphs. We select a rooted graph $H$ according to the distribution $\sigma^*$ and then select an edge $e$ incident from the root uniformly at random. We consider $e$ as oriented away from the root. This provides a probability distribution $\sigma^\to$ on the set $\mathfrak{B}^\to$ of graphs in $\mathfrak{B}^\bullet$ with an oriented edge from the root are also specified.

**Definition 9.4.** *We say that $\sigma$ is involution invariant if the map $\mathfrak{B}^\to \to \mathfrak{B}^\to$ obtained by reversing the orientation of the root edge and changing the root is measure preserving with respect to $\sigma^\to$. We call the random rooted connected graph drawn from an involution invariant probability measure on $\mathfrak{B}^\bullet$ as involution invariant random graph.*

We first introduce how to define sampling distribution of graphing. Let $\mathbf{G} = (\Omega, \mathcal{B}, \lambda, E)$ be a graphing and we choose a random point $x \in V(G)$ with respect to the given measure of the graphing. The connected component $\mathbf{G}_x$ of $\mathbf{G}$ containing $x$ with the root $x$ is a graph in $\mathfrak{B}^\bullet$, which we call a random rooted component of $\mathbf{G}$. The map $x \to \mathbf{G}_x$ is called the component map, and this is measurable function and thus every graphing $\mathbf{G}$ defines a probability distribution $\sigma_{\mathbf{G}}$ on $(\mathfrak{B}^\bullet, \mathfrak{A})$.

Let $d_0 = \int_\Omega \deg(x) d\lambda(x)$ let $\lambda^*(A) = \int_A \deg(x) d\lambda(x)/d_0$ be the stationary distribution of $\mathbf{G}$. Let $\eta = \eta_{\mathbf{G}}$ on $(\Omega \times \Omega, \mathcal{B} \times \mathcal{B})$ by

$$\eta(A \times B) = \int_A \deg_B(x) d\lambda(x).$$

This extends to the sigma-algebra $\mathcal{B} \times \mathcal{B}$ (this is by Caratheodory's extension theorem). The measure $\eta/d_0$ can be considered as a uniform probability measure on $E(\mathbf{G})$. Moreover, the above expression gives

$$\eta(A \times B) = \int_A \deg_B(x) d\lambda(x) = \int_B \deg_A(x) d\lambda(x) = \eta(B \times A)$$

hence, $\eta$ is symmetric.

We select $x$ from the distribution $\lambda^*$, the graph $\mathbf{G}_x$ will be a random rooted connected graph from the distribution $\sigma^*$ on $\mathfrak{B}^\bullet$. Selecting an edge of $\mathbf{G}_x$ incident with $x$, we get an edge of $\mathbf{G}$ from the probability distribution $\eta_{\mathbf{G}}/d_0$ together with an orientation. Since $\eta_G$ is symmetric, shifting the root to the other endpoint does not change the distribution, hence $\sigma$ is involution invariant. Thus, every graphing yields an involution invariant random graph, we also say that the graphing represents this distribution. The converse also holds by the following theorem. Note that such a graphing representing an involution invariant probability distribution is not unique.

**Theorem 9.5.** *Every involution invariant probability distribution on $\mathfrak{B}^\bullet$ can be represented by a graphing.*

*Proof.* We generalize the construction of graph of graphs. Let $\mathfrak{B}^+$ denote the set of triples $(H, v, \alpha)$ where $(H, v) \in \mathfrak{B}^\bullet$ and $\alpha : V(H) \to [0, 1]$ is a weighting of the vertices of $H$ and $v$ is the root of $H$. Two rooted weighted graphs are considered the same, if there is an isomorphism between the graphs preserving the root and the weights. Let $\mathcal{A}^+$ be the sigma-algebra on $\mathfrak{B}^+$ generated by the following cylinder sets: for an $r \geq 0$, we fix an $r$-ball $B \in \mathfrak{B}_r$ and for every vertex $u$ in $B$, we specify a Borel set $A_u$ in $[0, 1]$ and consider a cylinder set $\{(H, v, \alpha) : H \in \mathfrak{B}_B, \alpha(w) \in A_u\}$ where $w$ is the image of $u$ in the isomorphism from $B$ to an $r$-neighborhood of $H$ around the root. Then it is not difficult to check that $(\mathfrak{B}^+, \mathcal{A}^+)$ becomes a Borel sigma-algebra.

Let $\mathbf{H}^+$ be the graph of weighted graphs on the vertex set $\mathfrak{B}^+$ as follows: we connect $(G, \alpha)$ and $(G', \alpha')$ be an edge if $G'$ is obtained from $G$ by shfting the root to one of its neighbors while keeping all the vertex weights.

Given a probability distribution $\sigma$ on $(\mathfrak{B}^\bullet, \mathcal{A})$, we can define a probability distribution $\sigma^+$ on $(\mathfrak{B}^+, \mathcal{A}^+)$ as follows: we select a random graph $H \in \mathfrak{B}^\bullet$ from the distribution $\sigma$ and assign weights to each vertex from $[0, 1]$ independently uniformly at random.

Now, we prove that $(\mathbf{H}^+, \sigma^+)$ is a graphing. Choose a rooted weighted graph $(H, v, \alpha) \in \mathfrak{B}^+$ from the distribution $(\sigma^+)^*$ and a random neighbor $u$ of $v$ uniformly from the neighbors. Then almost surely, $H$ has no nontrivial automorphisms as our choices of weights are uniformly at random from $[0, 1]$, the graph $(H, u, \alpha)$ is different from $(H, v, \alpha)$. The pair $(H, u, \alpha)(H, v, \alpha)$ is an edge of $\mathbf{H}^+$ and selecting another neighbor of $v$ yields a different edge of $\mathbf{H}^+$. Hence, this procedure generates a random edge of $\mathbf{H}^+$ with an orientation. As $\sigma$ is involution invariant, this distribution on the edges of $\mathbf{H}$ is invariant under flipping the orientation.

For two measurable sets $A$ and $B$, $\int_A \deg_B(x) d\lambda(x)$ measures the probability of choosing a weighted rooted graph $(H, v, \alpha)$ in $A$ with an oriented edge towards $B$ and $\int_B \deg_A(x) d\lambda(x)$ measures the probability of choosing a weighted rooted graph $(H, u, \alpha)$ in $B$ with an oriented edge towards $A$. As the distribution on the edges of $\mathbf{H}$ is invariant under flipping the orientation, we have

$$\int_A \deg_B(x) d\lambda(x) = \int_B \deg_A(x) d\lambda(x)$$

Thus $(\mathbf{H}^+, \sigma^+)$ is a graphing.

Let $(H, v, \alpha)$ be an rooted graph from $\sigma^+$ with no automorphism. We claim that the connected component of $\mathbf{H}^+$ containing $(H, v, \alpha)$ as a root is isomorphic to $(H, v)$. Indeed, assinging the role of the root to different nodes of $H$ gives non-isomorphic rooted graphs, and as we get an injection of $V(H)$ into $\mathfrak{B}^\bullet$. From the definition of adjacency in $\mathbf{H}$, this injection preserves adjacncies and nonadjacencies, and $\mathfrak{B}^\bullet$ is the set of connected components of $\mathbf{H}^+$. This proves $(\mathbf{H}^+, \sigma^+)$ respresents $\sigma$. $\square$

## References

[1] L. Lovász, *Large Networks and Graph Limits*, 2012, Providence American Mathematical Society
[2] O. Pikhurko, *Descriptive combinatorics Lecture on 2019 at the University of Warwick.*
[3] Yufei Zhao, *Graph Theory and Additive Combinatorics*, scribed lecture note.