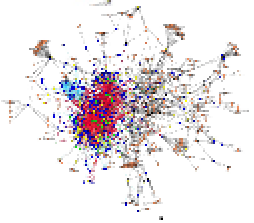


Graphons and Graph Limits

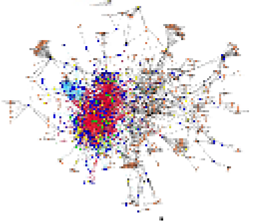
Parts 2 (dense graphs)

Christian Borgs, UC Berkeley

Part 3 (sparse graphs)

Jennifer Chayes, UC Berkeley

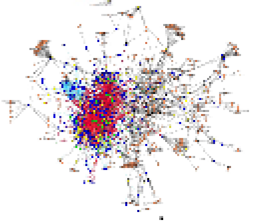
Simons Institute for the Theory of Computing
Fall 2022



Motivation: Three Related Problems

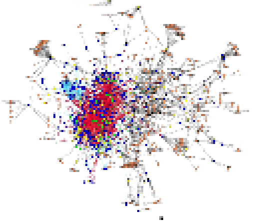
Questions:

- When should we consider two large graphs to be similar?
- What is the “correct notion” of a limit of graphs (preserving “essential” properties of the finite graphs in the sequence)?
- How do I non-parametrically model massive real-world networks data?



1) Modeling large random graphs

- A graphon is a symmetric 2-variable function over a probability space (Ω, μ), $W: \Omega \times \Omega \to [0,1]: (x, y) \mapsto W(x, y)$
- It generates inhomogeneous random graph $G_n(W)$ on by
  - assigning i.i.d. features $x_i \in \Omega$ according to μ to the vertices
  - connected $i < j$ independently with probability $P_{ij} = W(x_i, x_j)$
- By Aldous- Hoover, any exchangeable family of random graphs $(G_n)_{n \geq 1}$ can be generated by a (possibly random) graphon W



2) Notions of Similarity

<u>Subgraph frequencies</u>: Given a graph $G = (V, E)$ with adjacency matrix A and a graph H on k nodes, define

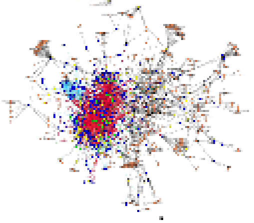
$$t_0(H, G) = \frac{1}{|V|^k} \sum_{v_1,\dots,v_k \in V} \prod_{ij \in E(H)} A_{v_i v_j} \prod_{ij \notin E(H)} (1 - A_{v_i v_j})$$

<u>Sampling</u>: Choose $x_1, \dots, x_k \in V$ uniformly at random and output $Smpl_k(G)$, the k-node graph with edge set $\{ij : x(i)x(j) \in E\}$

<u>Remark</u>: $\Pr(Smpl_k(G) = H) = t_0(H, G)$

which shows that similarity with respect to these two notions is equivalent

$$d_{TV}\big(Smpl_k(G), Smpl_k(G')\big) = \frac{1}{2} \sum_H |t_0(H, G) - t_0(H, G')|$$



2) Notions of Similarity

Two other notions:

- generalized min-cuts, $MinCut_{J,\alpha}(G)$
- micro-canonical free energies, $F_{J,\alpha}(G)$

both defined in terms of weighted cuts between different color classes for a coloring of G

2) Notions of Similarity

a) Multiway-Mincuts:

$$MinCut_{J,\alpha}(G) = \min_{\sigma} E_{G,J}(\sigma)$$

where $\alpha \in \Delta_k$ and the minimum is over the colorings with $|\sigma^{-1}(\{i\})| = n\alpha_i$ for all $i \in [k]$

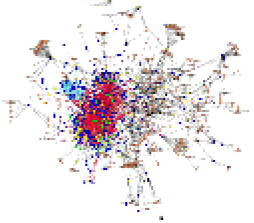
b) Micro-canonical free energy

$$F_{J,\alpha}(G) = -\frac{1}{n} \log Z_{J,\alpha}(G)$$

where

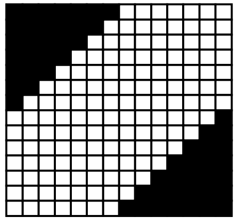
$$Z_{J,\alpha}(G) = \sum_{\sigma : V \to [k]} e^{-nE_{G,J}(\sigma)}$$

where the sum is over the colorings with $|\sigma^{-1}(\{i\})| = n\alpha_i$ for all $i \in [k]$



3) Cut-Metric

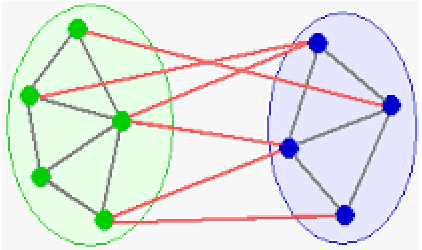
Empirical Graphon of a Graph G on n nodes



- Replace $[n]$ by n disjoint intervals $I_1, \ldots, I_n$ of width $1/n$ and divide $[0,1]^2$ into n^2 squares $I_i \times I_j$ of side length $1/n$
- Set W_G to 1 on the square ij if ij is an edge in G and to 0 if not

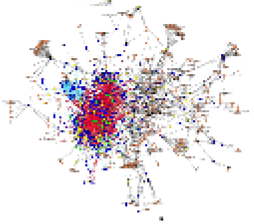
Cut norm* of a function $W: [0,1]^2 \to \mathbb{R}$

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



*) Equivalently, we can define $\|W\|_\square$ by

$$\|W\|_\square = \max_{f,g:[0,1]\to[0,1]} \left| \int f(x) W(x,y) g(y)\,dxdy \right|$$



3) Cut-Metric

Cut distance of two graphons $W_1, W_2 \colon [0,1]^2 \to [0,1]$

$$\delta_\square(W_1, W_2) = \inf_\phi \left\| W_1^\phi - W_2 \right\|_\square$$

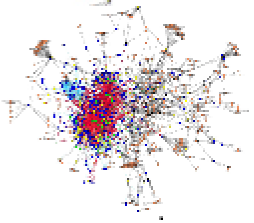
where the inf is over measure preserving bijections, and

$$W_1^\phi(x,y) = W_1(\phi(x), \phi(y))$$

Cut distance of two finite graphs G_1, G_2 we set

$$\delta_\square(G_1, G_2) := \delta_\square\left(W_{G_1}, W_{G_2}\right)$$

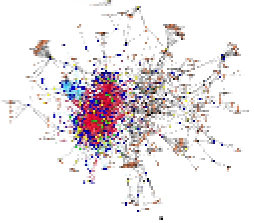
$$= \inf_\phi \max_{S,T \subset [0,1]} \left| \int_{S\times T} \left(W_{G_1}(\phi(x), \phi(y)) - W_{G_2}(x,y) \right) dxdy \right|$$



4) All these notions are equivalent!

<u>Thm:</u> Let G_n be a sequence of graphs. Then the following are equivalent

1) For all finite graphs H, the subgraph frequencies $t_0(H, G_n)$ converge
2) For all $k \geq 1$, the distributions of $Smpl_k(G_n)$ converge
3) For all $k \geq 1$, $J \in \mathbb{R}^{k\times k}$ and $\alpha \in \Delta_k$, the multi-way cuts $MinCut_{J,\alpha}(G_n)$ converge
4) For all $k \geq 1$, $J \in \mathbb{R}^{k\times k}$ and $\alpha \in \Delta_k$, the micro-canonical free energies $F_{J,\alpha}(G_n)$ converge
5) The sequence is a Cauchy sequence w.r.t. the cut metric



4) All these notions are equivalent!

a) Proof Idea:

I) Prove that if $\delta_{\square}(G, G') \leq \epsilon$, the other properties differ by at most a constant times ϵ (the constant you will get will be moderate, roughly proportional to k^2, and the norm of J). These proof are relatively elementary

II) The other direction is more difficult, and often will require k to be exponentially large in $1/\epsilon^2$

4b) Bounding subgraph counts in term of $\delta_\square$

Lemma: If H is a graph on k nodes and G, G' are two finite graphs, then

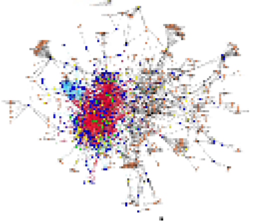
$$|t_0(H,G) - t_0(H,G')| \le \binom{k}{2} \delta_\square(G,G')$$

Step 1: Use empirical graphons: Define

$$t_0(H,W) = \int \prod_{i \in V(H)} dx_i \prod_{ij \in E(H)} W(x_i,x_j) \prod_{ij \notin E(H)} (1 - W(x_i,x_j))$$

Then $t_0(H,G) = t_0(H,W_G)$.

Proof: On the squares $I_s \times I_t$ the function W_G is constant (and equal to A_{st}). Thus the integral $\int \prod_{i \in V(H)} dx_i$ becomes n^{-k} times a sum over $v_1, \dots, v_k \in V(G)$, and $W(x_i,x_j)$ becomes $A_{v_i v_j}$



4b) Bounding subgraph counts in term of $\delta_\square$

Step 2: Prove that if $\phi: [0,1] \to [0,1]$ is measure preserving,

$$t_0(H, W) = t_0(H, W^\phi)$$

Step 3: Prove

$$|t_0(H, W) - t_0(H, U)| \le \binom{k}{2} \|W - U\|_\square$$

Putting things together:

$$|t_0(H, G) - t_0(H, G')| = |t_0(H, W_G) - t_0(H, W_{G'})| \le \binom{k}{2} \left\|W_G^\phi - W_{G'}\right\|_\square$$

Take infimum over all ϕ

$$|t_0(H, G) - t_0(H, G')| \le \delta_\square(G, G')$$

4b) Bounding subgraph counts in term of $\delta_\square$

Step 2: Prove that $t_0(H,W) = t_0(H,W^\phi)$ if ϕ is measure preserving

Proof: By inspection

$$t_0(H,W) = \int \prod_{i\in V(H)} dx_i \prod_{ij\in E(H)} W(x_i,x_j) \prod_{ij\notin E(H)} (1 - W(x_i,x_j))$$

Indeed, if we replace $W(x_i,x_j)$ by

$$W^\phi(x_i,x_j) = W(\phi(x_i),\phi(x_j))$$

this transforms the uniform random variable x_i into $\phi(x_i)$ which is again uniform

Step 3: We write

$$t_0(H,W)=\int\prod_{ij\in E(H)}W(x_i,x_j)\prod_{ij\notin E(H)}(1-W(x_i,x_j))=\int\prod_{i<j}W_{ij}(x_i,x_j)$$

where

$$W_{ij}=W_{ij}(x_i,x_j)=W(x_i,x_j)\qquad \text{if } ij\in E(F)$$

$$W_{ij}=W_{ij}(x_i,x_j)=1-W(x_i,x_j)\ \text{if } ij\notin E(F)$$

Thus

$$t_0(H,W)-t_0(H,U)=\int\prod_{i<j}W_{ij}-\prod_{i<j}U_{ij}$$

Changing W_{ij} to U_{ij} one at a time, we need to estimate $\binom{n}{2}$ integrals of the form

$$\int(W_{ij}(x_i,x_j)-U_{ij}(x_i,x_j))\prod_{st\neq ij}V_{st}(x_s,x_t)$$

where V_{st} is either equal to W_{st} or U_{st}

- Thus we need to estimate integrals of the form

$$\int (W_{ij}(x_i, x_j) - U_{ij}(x_i, x_j)) \prod_{st \neq ij} V_{st}(x_s, x_t)$$

where V_{st} is either equal to W_{st} or U_{st}

- Each V_{st} depends either on x_i or x_j or maybe neither of them, but not on both (since these terms appear already in the difference).
- If we fix all other x_l's, we therefore get an integral of the form

$$\int \left(W_{ij}(x_i, x_j) - U_{ij}(x_i, x_j)\right) f(x_i) g(x_j) dx_i dx_j$$

with f and g being functions with values in $[0,1]$ (depending implicitly on the other variables, of course).

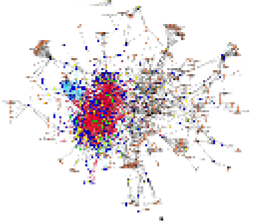
- By the equivalent definition of the cut-norm the absolute value of the integral is bounded by

$$\left\|W_{ij} - U_{ij}\right\|_{\square} = \|W - U\|_{\square}$$

uniformly in the remaining variables x_l

- Integrating over the remaining variables and taking into account that we have to do this $\binom{k}{2}$ times to change all W_{ij} to U_{ij}, this gives

$$|t_0(H,W) - t_0(H,U)| \leq \sum_{i<j} \int \prod_{l \notin \{i,j\}} dx_l \, \|W - U\|_\square = \binom{k}{2} \|W - U\|_\square$$



3c) Bounding $\delta_\square$ in terms of subgraph counts

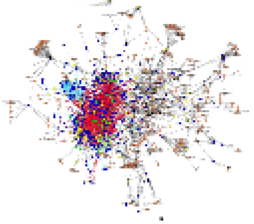
Thm If there exists a k s.th. $d_{TV}\left(Smpl_k(G), Smpl_k(G')\right) \leq \frac{10}{\sqrt{\log_2 k}}$ then

$$\delta_\square(G, G') \leq \frac{20}{\sqrt{\log_2 k}}$$

Proof: it is not hard to show that for all graphons W, $\delta_\square(\mathrm{G_k}(\mathrm{W}), \mathrm{W}) \to_{\mathbb{P}} 0$

Here we use a more difficult quantitative result on the expectation from [BCLSV'08], which says that

$$\mathbb{E}[\delta_\square(G_k(W), W)] \leq \frac{5}{\sqrt{log_2 k}}$$



3c) Bounding $\delta_\square$ in terms of subgraph counts

$$\delta_\square(G,G') \le \delta_\square(G,G_k(W_G)) + \delta_\square\left(G_k(W_G),G_k(W_{G'})\right) + \delta_\square\left(G',G_k(W_{G'})\right)$$

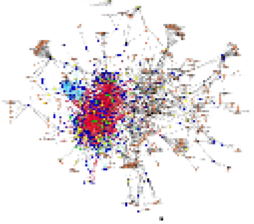
Taking expectations in the above bound we get

$$\delta_\square(G,G') \le \frac{10}{\sqrt{log_2 k}} + \mathbb{E}\left[\delta_\square\left(G_k(W_G),G_k\left(W_{G'}\right)\right)\right]$$

$$\le \frac{10}{\sqrt{log_2 k}} + Pr\left(G_k(W_G) \ne G_k\left(W_{G'}\right)\right)$$

$$= \frac{10}{\sqrt{log_2 k}} + Pr\left(Smpl_k(G) \ne Smpl_k(G')\right) \le \frac{20}{\sqrt{log_2 k}}$$

where we used that $G_k(W_G)$ has the same distribution as $Smpl_k(G)$



Summary so far

- Graphons are functions W of two variables lying in some feature space
- Given a probability distribution over features, Graphons give a natural random graph model $G_n(W)$ by connecting vertices with features x, y with probability $W(x, y)$
- If two graphs are close in the cut-metric, they have similar subgraph counts, distribution of sampled subgraphs, multi-way cuts, and micro canonical free energies, and vice versa

Outlook Graphs and Graphons

Graphs

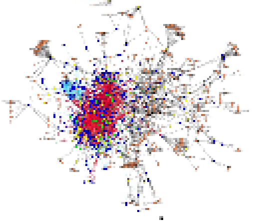
- Vertex set V
- Adjacency matrix $A: V \times V \rightarrow \{0,1\}$

↓ Graph Limits

↑ Non-Parametric Random Graph Models

Graphons

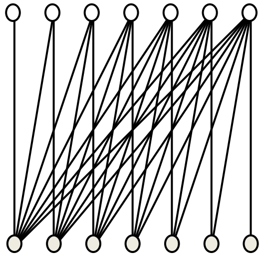
- Probability space $(\Omega, \mathcal{F}, \mu)$
- Symmetric, measurable function $W: \Omega \times \Omega \rightarrow [0,1]$



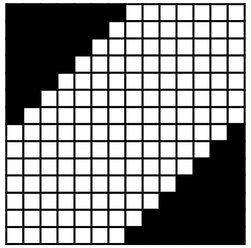
5) Graphons as Limits in the Cut Metric

Heuristically, the limit of black/white pattern is a grey picture on $[0,1]^2$.

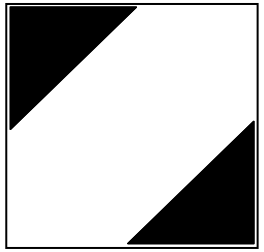
Half graph



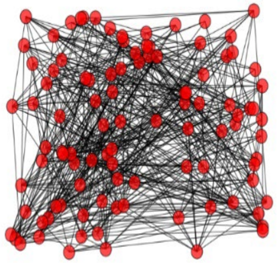
~



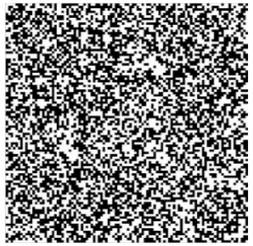
→



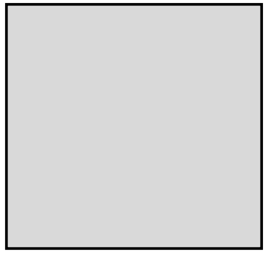
Random graph $G_{n,p}$ with $p = {}^{1}/_{2}$



~

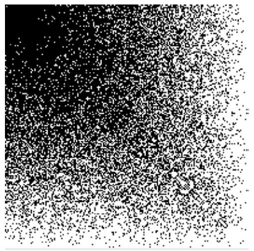


→



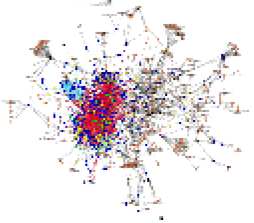
$W \equiv {}^{1}/_{2}$

Randomly grown uniform attachment graph, ordered by degrees



→

$W(x,y) = 1 - \max(x,y)$



5) Graphons as Limits in the Cut Metric

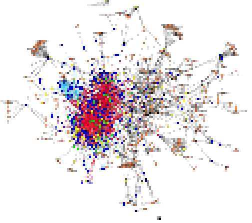
Def A sequence of graphs G_n converges to a graphon $W: [0,1]^2 \to [0,1]$ in the cut metric iff

$$\delta_\square(W_{G_n}, W) \to 0$$

Thm [BCLSV '08,'12]: Let G_n be a sequence of graphs with $V(G_n) \to \infty$. Then the following are equivalent

1) For all finite graphs H, $t_0(H, G_n) \to t_0(H, W)$
2) For all $k \geq 1$, $Smpl_k(G_n) \to G_k(W)$ in distribution
3) $\delta_\square(W_{G_n}, W) \to 0$

The limits of $MinCut_{J,\alpha}(G)$ and $F_{J,\alpha}(G)$ can also expressed in terms of W, and convergence to these is also equivalent



5) Graphons as Limits in the Cut Metric

Def A sequence of graphs G_n converges to a graphon $W: [0,1]^2 \to [0,1]$ in the cut metric iff

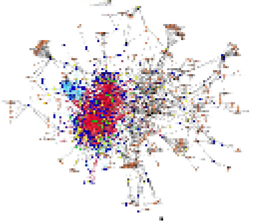
$$\delta_\square(W_{G_n}, W) \to 0$$

Questions:

1) Is there any growing sequence that converges to a graphon?
2) Can an arbitrary graphon $W: [0,1]^2 \to [0,1]$ be obtained as a limit of a sequence of graphs
3) Given a sequence G_n, is there a subsequence that converges to a graphon W?

Answer 1 + 2: For any graphon $W: [0,1]^2 \to [0,1]$, the sequence of inhomogeneous random graphs $G_n(W)$ converges to W in the cut metric

Answer 3: Yes. This follows from the weak regularity lemma.



6) Graphons as Limits in the Cut Metric

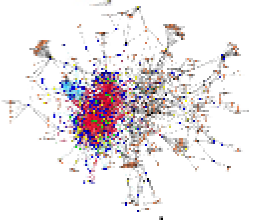
Weak Regularity Lemma:

For any graphon W and any k, there exists a $k \times k$ matrix B such that

$$\delta_{\square}(W_B, W) \leq \frac{5}{\sqrt{\log_2 k}}$$

What does this mean?

- Any graphon can be approximated by a block graphon
- For all ϵ, we can cover the space of graphons with a finite ϵ-net of block graphons
- With some extra work, this implies that every sequence of graphons and hence of graphs has a subsequence converging to some graphon



7) Proof of the Weak Regularity Lemma

Partitions: $P = \{Y_1, Y_2, \dots, Y_k\}$ where $\bigcup_i Y_i = [0,1]$ is a partition of $[0,1]$ into disjoint subsets. $|P| = k$ is called the size of P

Averaging over partitions: W_P is the block graphon that is constant on $Y_i \times Y_j$ that is obtained by averaging

$$W_P(x,y) = \frac{1}{\lambda(Y_i)\lambda(Y_j)} \int_{Y_i \times Y_j} W(x',y')dx'dy'$$

Lemma [Frieze-Kannan '99]: For all $W: [0.1]^2 \to [-1,1]$ there exists a partition P of $[0,1]$ into at most $4^{\lceil 1/\epsilon^2 \rceil}$ many classes s.th.

$$\|W - W_P\|_\square \le \epsilon$$

7) Proof of the Weak Regularity Lemma

Proof of the Frieze-Kannan Lemma:

Starting with the trivial partition P_0 consisting of just one class $Y_1 = [0,1]$, we will successively construct partitions P_t s.th. latest at $t = \lfloor 1/\epsilon^2 \rfloor$

$$\left\| W - W_{P_t} \right\|_\square \leq \epsilon$$

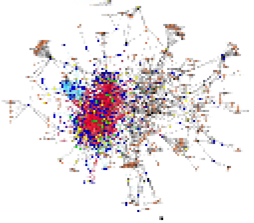
Assume this has not happened up to step t, i.e., assume that $\left\| W - W_{P_t} \right\|_\square > \epsilon$.

$\Longrightarrow$ there exists $S, T \subset [0,1]$ s.th. $\left| \int_{S\times T} W - W_{P_t} \right| > \epsilon$

Let P_{t+1} be the common refinement of P_t, $\{S, S^c\}$ and $\{T, T^c\}$

$$\Longrightarrow \epsilon < \left| \int_{S\times T} W - W_{P_t} \right| = \left| \int 1_{S\times T} (W_{P_{t+1}} - W_{P_t}) \right| \leq \left\| (W_{P_{t+1}} - W_{P_t}) \right\|_2$$

where the second step uses that P_{t+1} is a refinement of $\{S, S^c\}$ and $\{T, T^c\}$, and the last one uses Cauchy-Schwarz.



7) Proof of the Weak Regularity Lemma

Next we use that P_{t+1} is a refinement of P_t to get

$$\left\|(W_{P_{t+1}} - W_{P_t})\right\|_2^2 = \left\|W_{P_{t+1}}\right\|_2^2 + \left\|W_{P_t}\right\|_2^2 - 2\int W_{P_{t+1}} W_{P_t}$$

$$= \left\|W_{P_{t+1}}\right\|_2^2 + \left\|W_{P_t}\right\|_2^2 - 2\int W_{P_t} W_{P_t} = \left\|W_{P_{t+1}}\right\|_2^2 - \left\|W_{P_t}\right\|_2^2$$

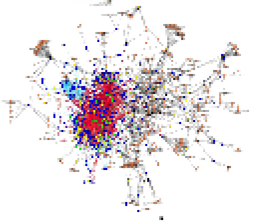
We therefore have shown

$$\left\|W_{P_{t+1}}\right\|_2^2 > \left\|W_{P_t}\right\|_2^2 + \epsilon^2 \geq \cdots \geq \left\|W_{P_0}\right\|_2^2 + (t+1)\epsilon^2 \geq (t+1)\epsilon^2$$

Since the l.h.s. is bounded by 1, this produces a contradiction if $(t+1)\epsilon^2 \geq 1$, showing that latest when $t = \lfloor 1/\epsilon^2 \rfloor$,

$$\left\|W - W_{P_t}\right\|_\square \leq \epsilon,$$

as claimed



7) Proof of the Weak Regularity Lemma

<u>Remarks:</u>

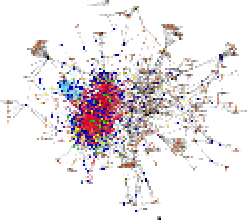
1. It is easy to transform a partition into an equi-parititition, i.e., a partition whose classes have all the same measure, by subdividing the given partition into smaller pieces. All this will do is change the constants involved. Being generous with these constants, one gets a bound on the number of classes of $2^{25/\epsilon^2}$, or equivalently, an error of $5/\sqrt{\log_2 k}$ in terms of the number of classes k
2. This shows that for each $W: [0.1]^2 \to [-1,1]$ there exists a partition P into k classes such that

$$\|W - W_P\|_\square \leq \frac{5}{\sqrt{\log_2 k}}$$

Applying measure preserving transformations, this then will give a partition into successive intervals of length $1/k$, which means that W_P becomes a standard block Graphon W_B where B is a $k \times k$ matrix, as claimed earlier.

Summary

- Graphons are functions W of two variables lying in some feature space
- Given a probability distribution over features, Graphons give a natural random graph model $G_n(W)$ by connecting vertices with features x, y with probability $W(x, y)$
- If two graphs are close in the cut-metric, they have similar subgraph counts, distribution of sampled subgraphs, multi-way cuts, and micro canonical free energies, and vice versa
- A Cauchy sequence in the cut metric converges to a graphon, and the limiting subgraph counts, distribution of sampled subgraphs, multi-way cuts, and micro canonical free energies can be expressed in terms of the limiting graphon



Thank you!