Large Deviations and Random Graphs

S.R.S. Varadhan
Courant Institute, NYU

UC Irvine
March 25, 2011

- Joint work with Sourav Chatterjee.

- Joint work with Sourav Chatterjee.
- G is a graph

- Joint work with Sourav Chatterjee.
- G is a graph
- $V(G)$ is the set of its vertices

- Joint work with Sourav Chatterjee.
- G is a graph
- $V(G)$ is the set of its vertices
- $E(G)$ is the set of its edges.

- Joint work with Sourav Chatterjee.
- G is a graph
- $V(G)$ is the set of its vertices
- $E(G)$ is the set of its edges.
- Not oriented.

- Joint work with Sourav Chatterjee.
- G is a graph
- $V(G)$ is the set of its vertices
- $E(G)$ is the set of its edges.
- Not oriented.
- $|V(G)|$ is the number of vertices in G

- Joint work with Sourav Chatterjee.
- G is a graph
- $V(G)$ is the set of its vertices
- $E(G)$ is the set of its edges.
- Not oriented.
- $|V(G)|$ is the number of vertices in G
- $|E(G)|$ are the number of edges in G.

- $\mathcal{G}_N$ is the set of all graphs G with N vertices.

- $\mathcal{G}_N$ is the set of all graphs G with N vertices.
- We think of $V(G)$ as $\{1, 2, \ldots, N\}$

- $\mathcal{G}_N$ is the set of all graphs G with N vertices.
- We think of $V(G)$ as $\{1, 2, \ldots, N\}$
- The set $\mathcal{E}$ is all unordered pairs (i, j), i.e. the full set of edges.

- $\mathcal{G}_N$ is the set of all graphs G with N vertices.
- We think of $V(G)$ as $\{1, 2, \ldots, N\}$
- The set $\mathcal{E}$ is all unordered pairs (i, j), i.e. the full set of edges.
- $E(G) \subset \mathcal{E}$

- G is determined by its edge set $E(G) \subset \mathcal{E}$

- G is determined by its edge set $E(G) \subset \mathcal{E}$
- $|\mathcal{G}_N| = 2^{\binom{N}{2}}$

- G is determined by its edge set $E(G) \subset \mathcal{E}$
- $|\mathcal{G}_N| = 2^{\binom{N}{2}}$
- A random graph with N vertices is just a probability measure on $\mathcal{G}_N$, i.e. a collection of weights $\{p(G)\}$ with $\sum_{G \in \mathcal{G}_N} p(G) = 1$

- What types of random graphs do we consider?

- What types of random graphs do we consider?
- A pair of vertices are connected by an edge with probability p

- What types of random graphs do we consider?
- A pair of vertices are connected by an edge with probability p
- Different edges are independent.

- What types of random graphs do we consider?
- A pair of vertices are connected by an edge with probability p
- Different edges are independent.
- Th probability $P_{N,p}(G)$ of a graph $G \in \mathcal{G}_N$ is given by

- What types of random graphs do we consider?
- A pair of vertices are connected by an edge with probability p
- Different edges are independent.
- Th probability $P_{N,p}(G)$ of a graph $G \in \mathcal{G}_N$ is given by
- $p^{|E(G)|}(1-p)^{\binom{N}{2}-|E(G)|}$

- What types of random graphs do we consider?
- A pair of vertices are connected by an edge with probability p
- Different edges are independent.
- Th probability $P_{N,p}(G)$ of a graph $G \in \mathcal{G}_N$ is given by
- $p^{|E(G)|}(1-p)^{\binom{N}{2}-|E(G)|}$
- If $p = \frac{1}{2}$, the distribution is uniform and we are essentially counting the number of graphs.

- We want to calculate the probabilities of the following types of events.

- We want to calculate the probabilities of the following types of events.
- Let Γ be a finite graph.

- We want to calculate the probabilities of the following types of events.
- Let Γ be a finite graph.
- The number of different occurrences of Γ in G is $\#_G(\Gamma)$

- We want to calculate the probabilities of the following types of events.
- Let Γ be a finite graph.
- The number of different occurrences of Γ in G is $\#_G(\Gamma)$
- The number of different occurrences of Γ in a complete graph with N vertices $\#_N(\Gamma)$

- We want to calculate the probabilities of the following types of events.
- Let Γ be a finite graph.
- The number of different occurrences of Γ in G is $\#_G(\Gamma)$
- The number of different occurrences of Γ in a complete graph with N vertices $\#_N(\Gamma)$
- The ratio $r_G(\Gamma) = \frac{\#_G(\Gamma)}{\#_N(\Gamma)}$

- We consider maps $V(\Gamma) \to V(G)$ that are one to one.

- We consider maps $V(\Gamma) \to V(G)$ that are one to one.
- If $|V(G)| = N$ and $|V(\Gamma)| = k$ there are $p(N,k) = N(N-1)\cdots(N-k+1)$ of them.

- We consider maps $V(\Gamma) \to V(G)$ that are one to one.
- If $|V(G)| = N$ and $|V(\Gamma)| = k$ there are $p(N,k) = N(N-1)\cdots(N-k+1)$ of them.
- Of these a certain number $p(G,N,k)$ will map a connected pair of vertices in Γ to connected ones in G.

- We consider maps $V(\Gamma) \to V(G)$ that are one to one.
- If $|V(G)| = N$ and $|V(\Gamma)| = k$ there are $p(N, k) = N(N-1)\cdots(N-k+1)$ of them.
- Of these a certain number $p(G, N, k)$ will map a connected pair of vertices in Γ to connected ones in G.
- $r_G(\Gamma) = \frac{p(G,N,k)}{p(N,k)}$.

- There is some ambiguity here due to possible multiple counting.

- There is some ambiguity here due to possible multiple counting.
- It is a multiple that depends only on Γ and will cancel out when we take the ratio.

- There is some ambiguity here due to possible multiple counting.
- It is a multiple that depends only on Γ and will cancel out when we take the ratio.
- What probabilities do we want to estimate ?

- There is some ambiguity here due to possible multiple counting.
- It is a multiple that depends only on Γ and will cancel out when we take the ratio.
- What probabilities do we want to estimate ?
- For each $j \in \{1, 2, \ldots, k\}$ we are given a finite graph Γ_j and a number $r_j \in [0, 1]$.

- There is some ambiguity here due to possible multiple counting.
- It is a multiple that depends only on Γ and will cancel out when we take the ratio.
- What probabilities do we want to estimate ?
- For each $j \in \{1, 2, \ldots, k\}$ we are given a finite graph Γ_j and a number $r_j \in [0, 1]$.
- We are interested in estimating the probability

$$P_{N,p}\big[\forall j,\ |r_G(\Gamma_j) - r_j| \leq \epsilon\big]$$

- More precisely we are interested in calculating the function

- More precisely we are interested in calculating the function
- $\psi_p(\{\Gamma_j, r_j\})$ given by

$$-\lim_{\epsilon\to 0}\lim_{N\to\infty}\frac{1}{\binom{N}{2}}\log P_{N,p}\big[\forall j,\ |r_G(\Gamma_j)-r_j|\leq\epsilon\big]$$

- More precisely we are interested in calculating the function
- $\psi_p(\{\Gamma_j, r_j\})$ given by

$$-\lim_{\epsilon\to 0}\lim_{N\to\infty}\frac{1}{\binom{N}{2}}\log P_{N,p}\big[\forall j,\ |r_G(\Gamma_j)-r_j|\leq\epsilon\big]$$

- $\psi_p(\{\Gamma_j, r_j\}) = 0$ if and only if $r_j = p^{E(\Gamma_j)}$ for $j = 1, 2, \ldots, k$.

- Let us consider the set

$$\mathcal{K} = \{f : [0,1]^2 \to [0,1]; f(x,y) = f(y,x)\}$$

- Let us consider the set

$$\mathcal{K} = \{f : [0,1]^2 \to [0,1]; f(x,y) = f(y,x)\}$$

- Define

$$H_p(f) = \int_0^1 \int_0^1 h_p(f(x,y))dxdy$$

- Let us consider the set

$$\mathcal{K} = \{f : [0,1]^2 \to [0,1]; f(x,y) = f(y,x)\}$$

- Define

$$H_p(f) = \int_0^1 \int_0^1 h_p(f(x,y))dxdy$$

- Where

$$h_p(f) = f \log \frac{f}{p} + (1-f) \log \frac{1-f}{1-p}$$

- For any $f \in \mathcal{K}$, finite graph Γ with vertices $\{1, 2, \ldots, k\}$, and edge set $E(\Gamma)$ we define

$$r^{\Gamma}(f) = \int_{[0,1]^{|V(\Gamma)|}} \Pi_{(i,j) \in E(\Gamma)} f(x_i, x_j) \Pi_{i=1}^{k} dx_i$$

- For any $f \in \mathcal{K}$, finite graph Γ with vertices $\{1, 2, \ldots, k\}$, and edge set $E(\Gamma)$ we define

$$r^{\Gamma}(f) = \int_{[0,1]^{|V(\Gamma)|}} \Pi_{(i,j)\in E(\Gamma)} f(x_i, x_j) \Pi_{i=1}^{k} dx_i$$

- For example if Γ is the triangle, then

$$r^{\Delta}(f) = \int_{[0,1]^3} f(x_1, x_2) f(x_2, x_3) f(x_3, x_1) dx_1 dx_2 dx_3$$

- For a k cycle it is

$$\int_{[0,1]^k} f(x_1, x_2) f(x_2, x_3) \cdots f(x_k, x_1) dx_1 dx_2 \cdots dx_k$$

- For a k cycle it is

$$\int_{[0,1]^k} f(x_1, x_2) f(x_2, x_3) \cdots f(x_k, x_1) dx_1 dx_2 \cdots dx_k$$

- The main result is.

$$\psi_p(\{\Gamma_j, r_j\}) = \inf_{\{f : \forall j,\ r^{\Gamma_j}(f) = r_j\}} H_p(f)$$

- What is it good for?

- What is it good for?
- Let us analyze $\psi_p(\Gamma^\Delta, c)$.

- What is it good for?
- Let us analyze $\psi_p(\Gamma^\Delta, c)$.
- Calculus of variations. The infimum is attained. Proof later. Compactness and continuity.

- What is it good for?
- Let us analyze $\psi_p(\Gamma^\Delta, c)$.
- Calculus of variations. The infimum is attained. Proof later. Compactness and continuity.
- Euler equation

$$\log \frac{f(x,y)}{1-f(x,y)} - \log \frac{p}{1-p} = \beta \int_0^1 f(x,z) f(y,z) dx$$

- Subject to

$$\int_{[0,1]^3} f(x,y)f(y,z)f(z,x)dxdydz = c$$

- Subject to

$$\int_{[0,1]^3} f(x,y)f(y,z)f(z,x)dxdydz = c$$

- If $|c - p^3| << 1$, then $f(x,y) = c^{\frac{1}{3}}$ is the only solution and so is optimal.

- Subject to

$$\int_{[0,1]^3} f(x,y)f(y,z)f(z,x)dxdydz = c$$

- If $|c - p^3| << 1$, then $f(x,y) = c^{\frac{1}{3}}$ is the only solution and so is optimal.
- For any c, if $p << 1$, then a clique

$$f = \mathbf{1}_{[0,c^{\frac{1}{3}}]}(x)\mathbf{1}_{[0,c^{\frac{1}{3}}]}(y)$$

is a better option than $f \equiv c^{\frac{1}{3}}$.

- A slight increase or decrease from p^3 in the proportion of triangles is explained by a corresponding deviation in the number of edges from p to $c^{\frac{1}{3}}$.

- A slight increase or decrease from p^3 in the proportion of triangles is explained by a corresponding deviation in the number of edges from p to $c^{\frac{1}{3}}$.
- This is not the case if p is small but c is not.

- A slight increase or decrease from p^3 in the proportion of triangles is explained by a corresponding deviation in the number of edges from p to $c^{\frac{1}{3}}$.
- This is not the case if p is small but c is not.
- A similar story when c is small but p is not.

- A slight increase or decrease from p^3 in the proportion of triangles is explained by a corresponding deviation in the number of edges from p to $c^{\frac{1}{3}}$.
- This is not the case if p is small but c is not.
- A similar story when c is small but p is not.
- A bipartite graph is a better option.

- What is the general Large Deviations setup and how do we apply it here?

- What is the general Large Deviations setup and how do we apply it here?
- A metric space $\mathcal{X}$ and a sequence P_n of probability distributions.

- What is the general Large Deviations setup and how do we apply it here?
- A metric space $\mathcal{X}$ and a sequence P_n of probability distributions.
- $P_n \to \delta_{x_0}$.

- What is the general Large Deviations setup and how do we apply it here?
- A metric space $\mathcal{X}$ and a sequence P_n of probability distributions.
- $P_n \to \delta_{x_0}$.
- If A is such that $d(x_0, A) > 0$ $P_n(A) \to 0$ as $n \to \infty$.

- Want a lower semi continuos function $I(x)$ such that

$$\frac{1}{n}\log P_n[S(x,\epsilon)] = -I(x) + o(\epsilon) + o_\epsilon(n)$$

- Want a lower semi continuos function $I(x)$ such that

$$\frac{1}{n}\log P_n[S(x,\epsilon)] = -I(x) + o(\epsilon) + o_\epsilon(n)$$

-

$$\limsup_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P_n(S(x,\epsilon)) \leq -I(x)$$

- Want a lower semi continuos function $I(x)$ such that

$$\frac{1}{n}\log P_n[S(x,\epsilon)] = -I(x) + o(\epsilon) + o_\epsilon(n)$$

- $$\limsup_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P_n(S(x,\epsilon)) \leq -I(x)$$

- $$\liminf_{\epsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log P_n(S(x,\epsilon)) \geq -I(x)$$

- If K is compact

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(K) \leq - \inf_{x\in K} I(x)$$

- If K is compact

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(K) \leq - \inf_{x\in K} I(x)$$

- If G is open

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq - \inf_{x\in G} I(x)$$

- For any ℓ the set $\{x : I(x) \leq \ell\}$ is compact.

- For any ℓ the set $\{x : I(x) \leq \ell\}$ is compact.
- For any $\ell < \infty$, there is a set K_ℓ such that $C \cap K_\ell = \emptyset$ implies

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -\ell$$

- For any ℓ the set $\{x : I(x) \leq \ell\}$ is compact.
- For any $\ell < \infty$, there is a set K_ℓ such that $C \cap K_\ell = \emptyset$ implies

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -\ell$$

- It now follows that for any closed set C

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n[C] \leq -\inf_{x\in C} I(x)$$

- Contraction Principle.

- Contraction Principle.
- $\{P_n\}$ satisfies LDP with rate $I(x)$ on $\mathcal{X}$,

- Contraction Principle.
- $\{P_n\}$ satisfies LDP with rate $I(x)$ on $\mathcal{X}$,
- $F : \mathcal{X} \to \mathcal{Y}$ is a continuous map.

- Contraction Principle.
- $\{P_n\}$ satisfies LDP with rate $I(x)$ on $\mathcal{X}$,
- $F : \mathcal{X} \to \mathcal{Y}$ is a continuous map.
- $Q_n = P_n F^{-1}$ satisfies an LDP on $\mathcal{Y}$

- Contraction Principle.
- $\{P_n\}$ satisfies LDP with rate $I(x)$ on $\mathcal{X}$,
- $F : \mathcal{X} \to \mathcal{Y}$ is a continuous map.
- $Q_n = P_n F^{-1}$ satisfies an LDP on $\mathcal{Y}$
- With rate function

$$J(y) = \inf_{x:F(x)=y} I(x)$$

- Let us turn to our case. The probability measures are on graphs with N vertices.

- Let us turn to our case. The probability measures are on graphs with N vertices.
- The space keeps changing.

- Let us turn to our case. The probability measures are on graphs with N vertices.
- The space keeps changing.
- Need to put them all on the same space.

- Let us turn to our case. The probability measures are on graphs with N vertices.
- The space keeps changing.
- Need to put them all on the same space.
- Every graph is an adjacency matrix.

- Let us turn to our case. The probability measures are on graphs with N vertices.
- The space keeps changing.
- Need to put them all on the same space.
- Every graph is an adjacency matrix.
- Random graph is a random symmetric matrix.
  $X = \{x_{i,j}\}, x_{i,i} = 0, x_{i,j} \in \{0, 1\}$

$$\begin{array}{cccc} 0 & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & 0 & \cdots & x_{2,N} \\ \ldots & \cdots & \cdots & \cdots \\ x_{N,1} & x_{N,2} & \cdots & 0 \end{array}$$

- Imbed in $\mathcal{K}$. Simple functions constant on small squares.

0	$x_{1,2}$	$\cdots$	$x_{1,N}$
$x_{2,1}$	0	$\cdots$	$x_{2,N}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$x_{N,1}$	$x_{N,2}$	$\cdots$	0

- Measures $\{Q_{N,p}\}$ on $\mathcal{K}$.

- Measures $\{Q_{N,p}\}$ on $\mathcal{K}$.
- The space $\mathcal{K}$ needs a topology. Weak is good. Nice compact space.

- Measures $\{Q_{N,p}\}$ on $\mathcal{K}$.
- The space $\mathcal{K}$ needs a topology. Weak is good. Nice compact space.
- $Q_{N,p} \Rightarrow \delta_p$

- Measures $\{Q_{N,p}\}$ on $\mathcal{K}$.
- The space $\mathcal{K}$ needs a topology. Weak is good. Nice compact space.
- $Q_{N,p} \Rightarrow \delta_p$
- Lower Bound. Let f be a nice function in $\mathcal{K}$.

- Measures $\{Q_{N,p}\}$ on $\mathcal{K}$.
- The space $\mathcal{K}$ needs a topology. Weak is good. Nice compact space.
- $Q_{N,p} \Rightarrow \delta_p$
- Lower Bound. Let f be a nice function in $\mathcal{K}$.
- Create a random graph with probability $f(\frac{i}{N}, \frac{j}{N})$ of connecting i and j

- By law of large numbers

$$Q_N^f \Rightarrow \delta_f$$

in the weak topology on $\mathcal{K}$.

- By law of large numbers

$$Q_N^f \Rightarrow \delta_f$$

in the weak topology on $\mathcal{K}$.

- The new measure Q_N^f on $\mathcal{K}$ has entropy

$$H(Q_N^f, Q_{N,p}) \simeq \binom{N}{2} H_p(f)$$

- Standard tilting argument

$$
\begin{aligned}
P(A) &= \int_A \frac{dP}{dQ} dQ \\
&= Q(A) \frac{1}{Q(A)} \int_A e^{-\log \frac{dQ}{dP}} dQ \\
&\geq Q(A) \exp[-\frac{1}{Q(A)} \int_A \log \frac{dQ}{dP} dQ] \\
&= \exp[-H(Q;P) + o(H(Q,P))]
\end{aligned}
$$

- Upper Bound. Cramér.

$$\frac{2}{N^2}\log E^{Q_{N,p}}[\frac{N^2}{2}\int J(x,y)f(x,y)dxdy]$$

$$\to \int_0^1\int_0^1 \log[pe^{J(x,y)} + (1-p)]dxdy$$

- Upper Bound. Cramér.

$$\frac{2}{N^2}\log E^{Q_{N,p}}[\frac{N^2}{2}\int J(x,y)f(x,y)dxdy]$$
$$\to \int_0^1\int_0^1 \log[pe^{J(x,y)} + (1-p)]dxdy$$

- Tchebychev. Half-plane. For small balls, optimize.

- Upper Bound. Cramér.

$$\frac{2}{N^2}\log E^{Q_{N,p}}[\frac{N^2}{2}\int J(x,y)f(x,y)dxdy]$$
$$\to \int_0^1\int_0^1 \log[pe^{J(x,y)} + (1-p)]dxdy$$

- Tchebychev. Half-plane. For small balls, optimize.
- $I(f) = H_p(f)$

- Upper Bound. Cramér.

$$\frac{2}{N^2}\log E^{Q_{N,p}}[\frac{N^2}{2}\int J(x,y)f(x,y)dxdy]$$
$$\to \int_0^1\int_0^1 \log[pe^{J(x,y)} + (1-p)]dxdy$$

- Tchebychev. Half-plane. For small balls, optimize.
- $I(f) = H_p(f)$
- Are we done!

- NO!, Why?

- NO!, Why?
- The object of interest is the map

$$F = \{r^{\Gamma_j}(f)\};\ \mathcal{K} \to [0,1]^k$$

- NO!, Why?
- The object of interest is the map

$$F = \{r^{\Gamma_j}(f)\};\ \mathcal{K} \to [0,1]^k$$

- They are not continuous unless no two edges in consists Γ share a common vertex.

- NO!, Why?
- The object of interest is the map

$$F = \{r^{\Gamma_j}(f)\};\ \mathcal{K} \to [0,1]^k$$

- They are not continuous unless no two edges in consists Γ share a common vertex.
- Well. Change the topology to L_1

- NO!, Why?
- The object of interest is the map

$$F = \{r^{\Gamma_j}(f)\};\ \mathcal{K} \to [0,1]^k$$

- They are not continuous unless no two edges in consists Γ share a common vertex.
- Well. Change the topology to L_1
- No chance. Even the Law of large numbers fails.

- NO!, Why?
- The object of interest is the map

$$F = \{r^{\Gamma_j}(f)\};\ \mathcal{K} \to [0,1]^k$$

- They are not continuous unless no two edges in consists Γ share a common vertex.
- Well. Change the topology to L_1
- No chance. Even the Law of large numbers fails.
- In between topology! Cut topology.

$$d(f,g) = \sup_{\substack{\|a\|\leq 1 \\ \|b\|\leq 1}} \int\int [f(x,y) - g(x,y]a(x)b(y)dxdy$$

$$= \sup_{A,B} \int_A \int_B [f(x,y) - g(x,y]dxdy$$

$$d(f,g) = \sup_{\substack{\|a\|\leq 1 \\ \|b\|\leq 1}} \int\int [f(x,y) - g(x,y]a(x)b(y)dxdy$$

$$= \sup_{A,B} \int_A \int_B [f(x,y) - g(x,y]dxdy$$

- In the cut topology F is continuous. Half the battle!

- $$d(f,g) = \sup_{\substack{\|a\|\le 1\\ \|b\|\le 1}} \int\int [f(x,y) - g(x,y]a(x)b(y)dxdy$$
$$= \sup_{A,B} \int_A \int_B [f(x,y) - g(x,y]dxdy$$
- In the cut topology F is continuous. Half the battle!
- Law of large numbers?

- Enough to take A and B as unions of intervals of the form $[\frac{j}{N}, \frac{j+1}{N}]$

- Enough to take A and B as unions of intervals of the form $[\frac{j}{N}, \frac{j+1}{N}]$
- For each $A \times B$ it is only the ordinary LLN for independent random variables.

- Enough to take A and B as unions of intervals of the form $[\frac{j}{N}, \frac{j+1}{N}]$
- For each $A \times B$ it is only the ordinary LLN for independent random variables.
- Error Bounds e^{-cN^2}

- Enough to take A and B as unions of intervals of the form $[\frac{j}{N}, \frac{j+1}{N}]$
- For each $A \times B$ it is only the ordinary LLN for independent random variables.
- Error Bounds e^{-cN^2}
- Number of rectangles $2^n \times 2^n$. That is good. LLN Holds.

- Enough to take A and B as unions of intervals of the form $[\frac{j}{N}, \frac{j+1}{N}]$
- For each $A \times B$ it is only the ordinary LLN for independent random variables.
- Error Bounds e^{-cN^2}
- Number of rectangles $2^n \times 2^n$. That is good. LLN Holds.
- Three fourths of the battle!

- If $\mathcal{K}$ were compact in the cut topology we would be done.

- If $\mathcal{K}$ were compact in the cut topology we would be done.
- But it is not. The projection $f \to \int f(x, y)dy$ is continuous. The image is L_1 and not compact.

- If $\mathcal{K}$ were compact in the cut topology we would be done.
- But it is not. The projection $f \to \int f(x, y)dy$ is continuous. The image is L_1 and not compact.
- The problem is invariant under a huge group. The permutation group Π_N.

- If $\mathcal{K}$ were compact in the cut topology we would be done.
- But it is not. The projection $f \to \int f(x,y)dy$ is continuous. The image is L_1 and not compact.
- The problem is invariant under a huge group. The permutation group Π_N.
- The function $H_p(f)$, $r^\Gamma(f)$ are invariant under the group $\sigma \in \Sigma$ of measure preserving transformations of $[0,1]$.

- Go to $\mathcal{K}/\Sigma$.

- Go to $\mathcal{K}/\Sigma$.
- It is compact! (Lovász-Szegedy)

- Go to $\mathcal{K}/\Sigma$.
- It is compact! (Lovász-Szegedy)
- But what is it?

- Go to $\mathcal{K}/\Sigma$.
- It is compact! (Lovász-Szegedy)
- But what is it?
- It is the space of "graphons"

- Go to $\mathcal{K}/\Sigma$.
- It is compact! (Lovász-Szegedy)
- But what is it?
- It is the space of "graphons"
- What is a graphon?

- G_n a sequence of graphs. Becoming infinite.

- G_n a sequence of graphs. Becoming infinite.
- We say G_n has limit if

$$\lim_{n\to\infty} r_{G_n}(\Gamma) = r(\Gamma)$$

exists for every finite graph Γ

- G_n a sequence of graphs. Becoming infinite.
- We say G_n has limit if

$$\lim_{n\to\infty} r_{G_n}(\Gamma) = r(\Gamma)$$

exists for every finite graph Γ

- Graphon is the map $\Gamma \to r(\Gamma)$

- G_n a sequence of graphs. Becoming infinite.
- We say G_n has limit if

$$\lim_{n\to\infty} r_{G_n}(\Gamma) = r(\Gamma)$$

  exists for every finite graph Γ
- Graphon is the map $\Gamma \to r(\Gamma)$
- It has a representation as $r^{\Gamma}(f)$ for some f in $\mathcal{K}$

- f is not unique. But $r^\Gamma(f) \equiv r^\Gamma(g)$ if and only if $f(x,y) = g(\sigma x, \sigma y)$ for some $\sigma \in \Sigma$

- f is not unique. But $r^\Gamma(f) \equiv r^\Gamma(g)$ if and only if $f(x, y) = g(\sigma x, \sigma y)$ for some $\sigma \in \Sigma$
- In other words $r(\cdot) \in \mathcal{K}/\Sigma$

- Since $\mathcal{K}/\Sigma$ is compact it is enough to prove the upper bound for balls $B(\tilde{f}, \epsilon)$ in $\mathcal{K}/\Sigma$

- Since $\mathcal{K}/\Sigma$ is compact it is enough to prove the upper bound for balls $B(\tilde{f}, \epsilon)$ in $\mathcal{K}/\Sigma$
- This means estimating

$$Q_{N,p}[\cup_{\sigma\in\Sigma} B(\sigma f, \epsilon)]$$

- Since $\mathcal{K}/\Sigma$ is compact it is enough to prove the upper bound for balls $B(\tilde{f}, \epsilon)$ in $\mathcal{K}/\Sigma$
- This means estimating

$$Q_{N,p}[\cup_{\sigma \in \Sigma} B(\sigma f, \epsilon)]$$

- Szemerédi's regularity lemma.

- Since $\mathcal{K}/\Sigma$ is compact it is enough to prove the upper bound for balls $B(\tilde{f}, \epsilon)$ in $\mathcal{K}/\Sigma$
- This means estimating

$$Q_{N,p}[\cup_{\sigma \in \Sigma} B(\sigma f, \epsilon)]$$

- Szemerédi's regularity lemma.
- The permutation group $\Pi_N \subset \Sigma$ by permuting intervals of length $\frac{1}{N}$.

- Since $\mathcal{K}/\Sigma$ is compact it is enough to prove the upper bound for balls $B(\tilde{f}, \epsilon)$ in $\mathcal{K}/\Sigma$
- This means estimating

$$Q_{N,p}[\cup_{\sigma \in \Sigma} B(\sigma f, \epsilon)]$$

- Szemerédi's regularity lemma.
- The permutation group $\Pi_N \subset \Sigma$ by permuting intervals of length $\frac{1}{N}$.
- Given $\epsilon > 0$, there is a finite set $\{g_j\} \subset \mathcal{K}$ such that for sufficiently large N,

$$\mathcal{K}_N = \cup_j \cup_{\sigma \in \Pi_N} B(\sigma g_j, \epsilon)$$

- It is therefore enough to estimate the probability

$$Q_{N,p}[\cup_j \cup_{\sigma \in \Pi_N} B(\sigma g_j, \epsilon)) \cap [\cup_{\sigma \in \Sigma} B(\sigma f, \epsilon)]]$$

- It is therefore enough to estimate the probability

$$Q_{N,p}[\cup_j \cup_{\sigma\in\Pi_N} B(\sigma g_j, \epsilon)) \cap [\cup_{\sigma\in\Sigma} B(\sigma f, \epsilon)]]$$

- Since j only varies over a finite set, it is enough to estimate for any g

$$Q_{N,p}\Big[[\cup_{\sigma\in\pi(N)} B(\sigma g, \epsilon)] \cap [\cup_{\sigma\in\Sigma} B(\sigma f, \epsilon)]\Big]$$

- It is therefore enough to estimate the probability

$$Q_{N,p}[\cup_j \cup_{\sigma\in\Pi_N} B(\sigma g_j,\epsilon)) \cap [\cup_{\sigma\in\Sigma} B(\sigma f,\epsilon)]]$$

- Since j only varies over a finite set, it is enough to estimate for any g

$$Q_{N,p}\Big[[\cup_{\sigma\in\pi(N)} B(\sigma g,\epsilon)] \cap [\cup_{\sigma\in\Sigma} B(\sigma f,\epsilon)]\Big]$$

- This is the same as

$$Q_{N,p}\Big[\cup_{\sigma\in\pi(N)} [B(\sigma g,\epsilon) \cap \cup_{\sigma\in\Sigma} B(\sigma f,\epsilon)]\Big]$$

- $Q_{N,p}$ is Π_N invariant. $N! << e^{cN^2}$.

$$Q_{N,p}\Big[B(g,\epsilon)\cap\cup_{\sigma\in\Sigma}B(\sigma f,\epsilon)\Big]$$

- $Q_{N,p}$ is Π_N invariant. $N! << e^{cN^2}$.

$$Q_{N,p}\Big[B(g,\epsilon) \cap \cup_{\sigma\in\Sigma} B(\sigma f,\epsilon)\Big]$$

- If the intersection is nonempty, then it is contained in $B(\sigma f, 3\epsilon)$ for some $\sigma \in \Sigma$.

- $Q_{N,p}$ is Π_N invariant. $N! << e^{cN^2}$.

$$Q_{N,p}\Big[B(g,\epsilon)\cap \cup_{\sigma\in\Sigma}B(\sigma f,\epsilon)\Big]$$

- If the intersection is nonempty, then it is contained in $B(\sigma f, 3\epsilon)$ for some $\sigma \in \Sigma$.
- The choice of σ does not depend on N. Only on ϵ.

- $Q_{N,p}$ is Π_N invariant. $N! << e^{cN^2}$.

$$Q_{N,p}\Big[B(g,\epsilon)\cap\cup_{\sigma\in\Sigma}B(\sigma f,\epsilon)\Big]$$

- If the intersection is nonempty, then it is contained in $B(\sigma f, 3\epsilon)$ for some $\sigma\in\Sigma$.
- The choice of σ does not depend on N. Only on ϵ.
- Balls are weakly closed.

- $Q_{N,p}$ is Π_N invariant. $N! << e^{cN^2}$.

$$Q_{N,p}\Big[B(g,\epsilon) \cap \cup_{\sigma\in\Sigma} B(\sigma f,\epsilon)\Big]$$

- If the intersection is nonempty, then it is contained in $B(\sigma f, 3\epsilon)$ for some $\sigma \in \Sigma$.
- The choice of σ does not depend on N. Only on ϵ.
- Balls are weakly closed.
- We have upper bounds. $H_p(\sigma f) = H_p(f)$

- $Q_{N,p}$ is Π_N invariant. $N! << e^{cN^2}$.

$$Q_{N,p}\Big[B(g,\epsilon) \cap \cup_{\sigma\in\Sigma} B(\sigma f,\epsilon)\Big]$$

- If the intersection is nonempty, then it is contained in $B(\sigma f, 3\epsilon)$ for some $\sigma \in \Sigma$.
- The choice of σ does not depend on N. Only on ϵ.
- Balls are weakly closed.
- We have upper bounds. $H_p(\sigma f) = H_p(f)$
- Done!

- With $p = \frac{1}{2}$, we have done the counting.

- With $p = \frac{1}{2}$, we have done the counting.
- The quantity

$$D(N,\epsilon) = \#|\{G : |r_G(\Gamma_j) - r_j| \leq \epsilon \, for \; j = 1, 2, \ldots, k\}|$$

- With $p = \frac{1}{2}$, we have done the counting.
- The quantity

$$D(N,\epsilon) = \#|\{G : |r_G(\Gamma_j) - r_j| \leq \epsilon \, for \; j = 1, 2, \ldots, k\}|$$

- Satisfies

$$\lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{2}{N^2} \log D(N,\epsilon) = \log 2 - \psi_{\frac{1}{2}}(\{\Gamma_j, r_j\})$$

Thank You.