Random Graphs

ISI, Bangalore, 25/1/17

- What is a Graph?

- What is a Graph?
- It is a set $\mathcal{X}$ of points called vertices.

- What is a Graph?
- It is a set $\mathcal{X}$ of points called vertices.
- Pairs (x, y) of points of called edges. (x, y) and (y, x) are the same edge.

- What is a Graph?
- It is a set $\mathcal{X}$ of points called vertices.
- Pairs (x, y) of points of called edges. (x, y) and (y, x) are the same edge.
- If there are n vertices there are $\frac{n(n-1)}{2}$ possible edges.

- What is a Graph?
- It is a set $\mathcal{X}$ of points called vertices.
- Pairs (x, y) of points of called edges. (x, y) and (y, x) are the same edge.
- If there are n vertices there are $\frac{n(n-1)}{2}$ possible edges.
- A graph specifies a sub set $\mathcal{E}$ from all possible edges as being present.

- What is a Graph?
- It is a set $\mathcal{X}$ of points called vertices.
- Pairs (x, y) of points of called edges. (x, y) and (y, x) are the same edge.
- If there are n vertices there are $\frac{n(n-1)}{2}$ possible edges.
- A graph specifies a sub set $\mathcal{E}$ from all possible edges as being present.
- A graph $\mathcal{G}$ is $\{\mathcal{X}, \mathcal{E}\}$ vertices and subset of edges.

- A random graph is simply one in which the edge set $\mathcal{E}$ is random.

- A random graph is simply one in which the edge set $\mathcal{E}$ is random.
- In Erdös-Renyi graphs $P[(x, y) \in \mathcal{E}] = p$ and different edges belong to $\mathcal{E}$ independently of one another.

- A random graph is simply one in which the edge set $\mathcal{E}$ is random.
- In Erdös-Renyi graphs $P[(x,y) \in \mathcal{E}] = p$ and different edges belong to $\mathcal{E}$ independently of one another.
- These are dense graphs with np edges coming out of each vertex.

- A random graph is simply one in which the edge set $\mathcal{E}$ is random.
- In Erdös-Renyi graphs $P[(x,y) \in \mathcal{E}] = p$ and different edges belong to $\mathcal{E}$ independently of one another.
- These are dense graphs with np edges coming out of each vertex.
- One can think of the graph of n vertices as a symmetric $n \times n$ matrix A of 0's and 1's with 0's on the diagonal.

- A random graph is simply one in which the edge set $\mathcal{E}$ is random.
- In Erdös-Renyi graphs $P[(x, y) \in \mathcal{E}] = p$ and different edges belong to $\mathcal{E}$ independently of one another.
- These are dense graphs with np edges coming out of each vertex.
- One can think of the graph of n vertices as a symmetric $n \times n$ matrix A of 0's and 1's with 0's on the diagonal.
- The matrix notation is useful.

- A random graph is simply one in which the edge set $\mathcal{E}$ is random.
- In Erdös-Renyi graphs $P[(x,y) \in \mathcal{E}] = p$ and different edges belong to $\mathcal{E}$ independently of one another.
- These are dense graphs with np edges coming out of each vertex.
- One can think of the graph of n vertices as a symmetric $n \times n$ matrix A of 0's and 1's with 0's on the diagonal.
- The matrix notation is useful.
- Trace A^2 is $2|\mathcal{E}|$ and trace A^3 is $6|\Delta|$,

- Laws of large numbers.

- Laws of large numbers.
- $\frac{2|\mathcal{E}|}{n^2} \to p$ and $\frac{6|\Delta|}{n^3} \to p^3$.

- Laws of large numbers.
- $\frac{2|\mathcal{E}|}{n^2} \to p$ and $\frac{6|\Delta|}{n^3} \to p^3$.
- How many homomorphisms from a finite graph $\mathcal{H} = (H, E)$ to $\mathcal{G} = (\mathcal{X}, \mathcal{E})$.

- Laws of large numbers.
- $\frac{2|\mathcal{E}|}{n^2} \to p$ and $\frac{6|\Delta|}{n^3} \to p^3$.
- How many homomorphisms from a finite graph $\mathcal{H} = (H, E)$ to $\mathcal{G} = (\mathcal{X}, \mathcal{E})$.
- $t(\mathcal{H}, \mathcal{G})$ is the number of Homomorphisms of $\mathcal{H} \to \mathcal{G}$.

- Laws of large numbers.
- $\frac{2|\mathcal{E}|}{n^2} \to p$ and $\frac{6|\Delta|}{n^3} \to p^3$.
- How many homomorphisms from a finite graph $\mathcal{H} = (H, E)$ to $\mathcal{G} = (\mathcal{X}, \mathcal{E})$.
- $t(\mathcal{H}, \mathcal{G})$ is the number of Homomorphisms of $\mathcal{H} \to \mathcal{G}$.
- For the Erdös-Renyi random graph with probability p for an edge,

- Laws of large numbers.
- $\frac{2|\mathcal{E}|}{n^2} \to p$ and $\frac{6|\Delta|}{n^3} \to p^3$.
- How many homomorphisms from a finite graph $\mathcal{H} = (H, E)$ to $\mathcal{G} = (\mathcal{X}, \mathcal{E})$.
- $t(\mathcal{H}, \mathcal{G})$ is the number of Homomorphisms of $\mathcal{H} \to \mathcal{G}$.
- For the Erdös-Renyi random graph with probability p for an edge,
- $\frac{t(\mathcal{H},\mathcal{G})}{n^{|H|}} \to p^{|\mathcal{S}|}$

- $\mathcal{G}_n$ is a sequence of graphs.

- $\mathcal{G}_n$ is a sequence of graphs.
- $\frac{t(\mathcal{H},\mathcal{G}_n)}{n^{|H|}} \to \sigma(\mathcal{H})$. The graph limit.

- $\mathcal{G}_n$ is a sequence of graphs.
- $\frac{t(\mathcal{H},\mathcal{G}_n)}{n^{|H|}} \to \sigma(\mathcal{H})$. The graph limit.
- What is $\sigma(\mathcal{H})$?

- $\mathcal{G}_n$ is a sequence of graphs.
- $\frac{t(\mathcal{H},\mathcal{G}_n)}{n^{|H|}} \to \sigma(\mathcal{H})$. The graph limit.
- What is $\sigma(\mathcal{H})$?
- Graphon. Vertex set $H = \{x_1, \ldots, x_k\}$. Edges $e \in E$

- $\mathcal{G}_n$ is a sequence of graphs.
- $\frac{t(\mathcal{H},\mathcal{G}_n)}{n^{|H|}} \to \sigma(\mathcal{H})$. The graph limit.
- What is $\sigma(\mathcal{H})$?
- Graphon. Vertex set $H = \{x_1, \ldots, x_k\}$. Edges $e \in E$
- There is a symmetric f, $0 \le f \le 1$ on $[0,1]^2$ with

$$\sigma(\mathcal{H}) = \int_{[0,1]^k} \prod_{(x_i,x_j)=e\in E} f(x_i, x_j) \Pi_{x_i \in H} dx_i$$

- Large Deviations

- Large Deviations
- $(X, \mathcal{B}, P_n)$

- Large Deviations
- $(X, \mathcal{B}, P_n)$
- For C closed

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

- Large Deviations
- $(X, \mathcal{B}, P_n)$
- For C closed

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

- For G open

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

- Large Deviations
- $(X, \mathcal{B}, P_n)$
- For C closed

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

- For G open

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

- $I(x) \geq 0$ is lower semicontinuous and has compact level sets $K_\ell = \{x : I(x) \leq \ell\}$

- Bernoulli $P[X = 1] = p$.

- Bernoulli $P[X = 1] = p$.
- $$\frac{1}{n}\sum \delta_{\frac{i}{N}} X_i$$

- Bernoulli $P[X = 1] = p$.
- $$\frac{1}{n} \sum \delta_{\frac{i}{N}} X_i$$
- LDP on $\mathcal{M}([0, 1])$

- Bernoulli $P[X = 1] = p$.
- $$\frac{1}{n} \sum \delta_{\frac{i}{N}} X_i$$
- LDP on $\mathcal{M}([0,1])$
- $I(\rho(\cdot)) =$

$$\int_0^1 [\rho(x) \log \frac{\rho(x)}{p} + (1 - \rho(x)) \log \frac{1 - \rho(x)}{1 - p} dx]$$

- Bernoulli $P[X=1]=p$.
- $$\frac{1}{n}\sum \delta_{\frac{i}{N}} X_i$$
- LDP on $\mathcal{M}([0,1])$
- $I(\rho(\cdot)) =$

$$\int_0^1 [\rho(x)\log\frac{\rho(x)}{p} + (1-\rho(x))\log\frac{1-\rho(x)}{1-p} dx]$$

- $$\lim_{n\to\infty}\frac{1}{n}\log E[\exp[\sum_i J(\frac{i}{n})X_i]] = \int_0^1 \psi(J(x))dx$$

- $$\psi(v) = \log E[e^{vX}] = \log[pe^v + (1-p)]$$

- $$\psi(v) = \log E[e^{vX}] = \log[pe^v + (1-p)]$$

- $$\sup_v [v\rho - \psi(v)] = \rho \log \frac{\rho}{p} + (1-\rho) \log \frac{1-\rho}{1-p}$$

- $$\psi(v) = \log E[e^{vX}] = \log[pe^v + (1-p)]$$
- $$\sup_v [v\rho - \psi(v)] = \rho \log \frac{\rho}{p} + (1-\rho) \log \frac{1-\rho}{1-p}$$
- Chebychev's inequality for half planes.

- $$\psi(v) = \log E[e^{vX}] = \log[pe^v + (1-p)]$$

- $$\sup_v [v\rho - \psi(v)] = \rho \log \frac{\rho}{p} + (1-\rho) \log \frac{1-\rho}{1-p}$$

- Chebychev's inequality for half planes.
- Optimize

- $$\psi(v) = \log E[e^{vX}] = \log[pe^v + (1-p)]$$

- $$\sup_v [v\rho - \psi(v)] = \rho \log \frac{\rho}{p} + (1-\rho) \log \frac{1-\rho}{1-p}$$

- Chebychev's inequality for half planes.
- Optimize
- We have local upper bounds in the weak topology. Space is compact we get global upper bounds for closed sets.

- Our problem is different

- Our problem is different
- The number of i.i.d variables is $\frac{n(n-1)}{2} \simeq \frac{n^2}{2}$.

- Our problem is different
- The number of i.i.d variables is $\frac{n(n-1)}{2} \simeq \frac{n^2}{2}$.
- The space is symmetric functions f on $[0,1]^2$. Or a function on $D = \{0 \le x \le y \le 1\}$.

- Our problem is different
- The number of i.i.d variables is $\frac{n(n-1)}{2} \simeq \frac{n^2}{2}$.
- The space is symmetric functions f on $[0,1]^2$. Or a function on $D = \{0 \le x \le y \le 1\}$.
- The rate function when normalized by n^2 is $\frac{1}{2}\int_{[0,1]^2} h_\rho(f(x,y))dxdy$ where

$$h_\rho(f) = f \log \frac{f}{\rho} + (1-f) \log \frac{1-f}{1-\rho}$$

- The matrix A can be viewed as a function $a(x, y)$on $[0, 1]^2$

- The matrix A can be viewed as a function $a(x, y)$on $[0, 1]^2$
- $a(x, y) = a_{i,j}$ on $\frac{i-1}{n} < x < \frac{i}{n}, \frac{j-1}{n} < y < \frac{j}{n}$

- The matrix A can be viewed as a function $a(x, y)$on $[0, 1]^2$
- $a(x, y) = a_{i,j}$ on $\frac{i-1}{n} < x < \frac{i}{n}, \frac{j-1}{n} < y < \frac{j}{n}$
- We have measures $P_{n,\rho}$ on the space $\mathcal{A}$ of symmetric functions a on $[0, 1]^2$ with $0 \leq a \leq 1$.

- The matrix A can be viewed as a function $a(x, y)$on $[0, 1]^2$
- $a(x, y) = a_{i,j}$ on $\frac{i-1}{n} < x < \frac{i}{n}, \frac{j-1}{n} < y < \frac{j}{n}$
- We have measures $P_{n,\rho}$ on the space $\mathcal{A}$ of symmetric functions a on $[0, 1]^2$ with $0 \leq a \leq 1$.
- The LLN says $P_{n,\rho} \to \delta_\rho$

- The matrix A can be viewed as a function $a(x, y)$on $[0, 1]^2$
- $a(x, y) = a_{i,j}$ on $\frac{i-1}{n} < x < \frac{i}{n}, \frac{j-1}{n} < y < \frac{j}{n}$
- We have measures $P_{n,\rho}$ on the space $\mathcal{A}$ of symmetric functions a on $[0, 1]^2$ with $0 \le a \le 1$.
- The LLN says $P_{n,\rho} \to \delta_\rho$
- What is the LDP?

- The matrix A can be viewed as a function $a(x, y)$on $[0, 1]^2$
- $a(x, y) = a_{i,j}$ on $\frac{i-1}{n} < x < \frac{i}{n}, \frac{j-1}{n} < y < \frac{j}{n}$
- We have measures $P_{n,\rho}$ on the space $\mathcal{A}$ of symmetric functions a on $[0, 1]^2$ with $0 \leq a \leq 1$.
- The LLN says $P_{n,\rho} \to \delta_\rho$
- What is the LDP?
- We are interested in quantities like number of triangles.

- The matrix A can be viewed as a function $a(x, y)$on $[0, 1]^2$
- $a(x, y) = a_{i,j}$ on $\frac{i-1}{n} < x < \frac{i}{n}, \frac{j-1}{n} < y < \frac{j}{n}$
- We have measures $P_{n,\rho}$ on the space $\mathcal{A}$ of symmetric functions a on $[0, 1]^2$ with $0 \leq a \leq 1$.
- The LLN says $P_{n,\rho} \to \delta_\rho$
- What is the LDP?
- We are interested in quantities like number of triangles.
- $\int a(x, y)a(y, z)a(z, x)dxdydz$

- $a_{ii}^n = 0$, $a_{2i,2j}^n = a_{2i+1,2j+1}^n = 0$,

- $a^n_{ii} = 0$, $a^n_{2i,2j} = a^n_{2i+1,2j+1} = 0$,
- $a^n_{2i,2j+1} = a^n_{2i+1.2j} = 1$

- $a^n_{ii} = 0$, $a^n_{2i,2j} = a^n_{2i+1,2j+1} = 0$,
- $a^n_{2i,2j+1} = a^n_{2i+1.2j} = 1$
- No of triangles is
  $0 = \int a(x,y)a(y,z)a(z,x)dxdydz$

- $a^n_{ii} = 0$, $a^n_{2i,2j} = a^n_{2i+1,2j+1} = 0$,
- $a^n_{2i,2j+1} = a^n_{2i+1.2j} = 1$
- No of triangles is $0 = \int a(x,y)a(y,z)a(z,x)dxdydz$
- Weak limit of a^n is $a(x,y) = \frac{1}{2}$.

- $a^n_{ii} = 0$, $a^n_{2i,2j} = a^n_{2i+1,2j+1} = 0$,
- $a^n_{2i,2j+1} = a^n_{2i+1.2j} = 1$
- No of triangles is $0 = \int a(x,y)a(y,z)a(z,x)dxdydz$
- Weak limit of a^n is $a(x,y) = \frac{1}{2}$.
- $\int a(x,y)(y,z)a(z,x)dxdydz = \frac{1}{8}$

- $a_{ii}^n = 0$, $a_{2i,2j}^n = a_{2i+1,2j+1}^n = 0$,
- $a_{2i,2j+1}^n = a_{2i+1.2j}^n = 1$
- No of triangles is $0 = \int a(x,y)a(y,z)a(z,x)dxdydz$
- Weak limit of a^n is $a(x,y) = \frac{1}{2}$.
- $\int a(x,y)(y,z)a(z,x)dxdydz = \frac{1}{8}$
- Expect $\frac{n^3}{8}$ triangles!

- Rearrange the vertices. Relabel them.

- Rearrange the vertices. Relabel them.
- $a^n_{i,j} = 1$ if $1 \le i \le \frac{n}{2}, j > \frac{n}{2}$ or $1 \le j \le \frac{n}{2}, i > \frac{n}{2}$

- Rearrange the vertices. Relabel them.
- $a^n_{i,j} = 1$ if $1 \leq i \leq \frac{n}{2}, j > \frac{n}{2}$ or $1 \leq j \leq \frac{n}{2}, i > \frac{n}{2}$
- Otherwise 0

- Rearrange the vertices. Relabel them.
- $a_{i,j}^n = 1$ if $1 \le i \le \frac{n}{2}, j > \frac{n}{2}$ or $1 \le j \le \frac{n}{2}, i > \frac{n}{2}$
- Otherwise 0
- $a(x,y) = 1$ if $0 \le x \le \frac{1}{2}, \frac{1}{2} \le y \le 1$ or $0 \le y \le \frac{1}{2}, \frac{1}{2} \le x \le 1$

- Rearrange the vertices. Relabel them.
- $a^n_{i,j} = 1$ if $1 \le i \le \frac{n}{2}, j > \frac{n}{2}$ or $1 \le j \le \frac{n}{2}, i > \frac{n}{2}$
- Otherwise 0
- $a(x,y) = 1$ if $0 \le x \le \frac{1}{2}, \frac{1}{2} \le y \le 1$ or $0 \le y \le \frac{1}{2}, \frac{1}{2} \le x \le 1$
- $\int a(x,y)a(y,z)a(z,x)dxdydz = 0$

- weak topology is no good. Too weak.

- weak topology is no good. Too weak.
- Strong or L_1 topology is too strong.

- weak topology is no good. Too weak.
- Strong or L_1 topology is too strong.
- LLN is not valid.

- weak topology is no good. Too weak.
- Strong or L_1 topology is too strong.
- LLN is not valid.
- We need some thing in between.

- $$d(f,g) = \sup_{h:\|h\|_\infty \le 1} \left| \int [f-g] h(x,y)\, dx dy \right|$$

- $$d(f,g) = \sup_{h:\|h\|_\infty \le 1} \left| \int [f-g]h(x,y)dxdy \right|$$

- $$d(f,g) = \sup_E \left| \int_E [f-g]dxdy \right|$$

- $$d(f,g) = \sup_{h:\|h\|_\infty \le 1} \left| \int [f-g] h(x,y)\,dx\,dy \right|$$

- $$d(f,g) = \sup_E \left| \int_E [f-g]\,dx\,dy \right|$$

- $$d_\square(f,g) = \sup_{\substack{h,k:\|h\|_\infty \le 1 \\ \|k\|_\infty \le 1}} \left| \int [f-g] h(x) k(y)\,dx\,dy \right|$$

- $$d(f,g) = \sup_{h:\|h\|_\infty \le 1} \left| \int [f-g] h(x,y) \, dx dy \right|$$

- $$d(f,g) = \sup_E \left| \int_E [f-g] \, dx dy \right|$$

- $$d_\square(f,g) = \sup_{\substack{h,k:\|h\|_\infty \le 1 \\ \|k\|_\infty \le 1}} \left| \int [f-g] h(x) k(y) \, dx dy \right|$$

- $$d_\square(f,g) = \sup_{E,F} \left| \int_{E \times F} [f-g] \, dx dy \right|$$

$$\sigma(\mathcal{H}, f) = \int_{[0,1]^k} \prod_{(i,j)\in E} f(x_i, x_j)\,dx_1 \cdots dx_k$$

- $$\sigma(\mathcal{H}, f) = \int_{[0,1]^k} \prod_{(i,j)\in E} f(x_i, x_j) dx_1 \cdots dx_k$$
- Continuous in the $d_\square$ metric.

- $$\sigma(\mathcal{H}, f) = \int_{[0,1]^k} \prod_{(i,j)\in E} f(x_i, x_j)dx_1 \cdots dx_k$$
- Continuous in the $d_\Box$ metric.
- Replace f_n by f in one edge at a time.

- $$\sigma(\mathcal{H}, f) = \int_{[0,1]^k} \prod_{(i,j)\in E} f(x_i, x_j) dx_1 \cdots dx_k$$
- Continuous in the $d_\square$ metric.
- Replace f_n by f in one edge at a time.
- $\int F_n(x_i) f_n(x_i, x_j) G_n(x_j) \simeq \int F_n(x_i) f(x_i, x_j) G_n(x_j)$

- Is the LLN valid in the cut or box topology?

- Is the LLN valid in the cut or box topology?
- Let E, F be subsets of $1, 2, \ldots, n$.

- Is the LLN valid in the cut or box topology?
- Let E, F be subsets of $1, 2, \ldots, n$.
- Uniformly in E and F

$$P[| [\sum_{\substack{i \in E \\ j \in F}} X_{ij} - p|E||F|] \mid \geq \epsilon n^2] \leq e^{-c(\epsilon)n^2}$$

- Is the LLN valid in the cut or box topology?
- Let E, F be subsets of $1, 2, \ldots, n$.
- Uniformly in E and F

$$P[|\,[\sum_{\substack{i \in E \\ j \in F}} X_{ij} - p|E||F|]\,| \geq \epsilon n^2] \leq e^{-c(\epsilon)n^2}$$

- The number of such pairs is at most $2^n \times 2^n = e^{O(n)}$.

- Is the LLN valid in the cut or box topology?
- Let E, F be subsets of $1, 2, \ldots, n$.
- Uniformly in E and F

$$P[|\,[\sum_{\substack{i\in E\\ j\in F}} X_{ij} - p|E||F|]\,| \geq \epsilon n^2] \leq e^{-c(\epsilon)n^2}$$

- The number of such pairs is at most $2^n \times 2^n = e^{O(n)}$.
- LLN holds in the cut metric.

- Lower bound is easy.

- Lower bound is easy.
- Assume $\rho(x, y)$ is continuous.

- Lower bound is easy.
- Assume $\rho(x, y)$ is continuous.
- Tilt. Make $X_{i,j}$ Bernoulli with $Q_n[X_{i,j} = 1] = \rho(\frac{i}{n}, \frac{j}{n})$

- Lower bound is easy.
- Assume $\rho(x, y)$ is continuous.
- Tilt. Make $X_{i,j}$ Bernoulli with $Q_n[X_{i,j} = 1] = \rho(\frac{i}{n}, \frac{j}{n})$
- P gets replaced by Q_n and the law of large numbers for Q_n provides the limit $\rho(x)$ in the cut metric.

- Lower bound is easy.
- Assume $\rho(x, y)$ is continuous.
- Tilt. Make $X_{i,j}$ Bernoulli with $Q_n[X_{i,j} = 1] = \rho(\frac{i}{n}, \frac{j}{n})$
- P gets replaced by Q_n and the law of large numbers for Q_n provides the limit $\rho(x)$ in the cut metric.
- A is a neighborhood of ρ and $Q_n(A) \to 1$.

- $$P_n(A) = \int_A \frac{dP_n}{dQ_n} dQ_n$$

- $$P_n(A) = \int_A \frac{dP_n}{dQ_n} dQ_n$$

- $$= Q_n(A) \frac{1}{Q_n(A)} \int_A \exp[-\log \frac{dQ_n}{dP_n}] dQ_n$$

- $$P_n(A) = \int_A \frac{dP_n}{dQ_n} dQ_n$$
- $$= Q_n(A) \frac{1}{Q_n(A)} \int_A \exp[-\log \frac{dQ_n}{dP_n}] dQ_n$$
- $$\geq Q_n(A) \exp[-\frac{1}{Q_n(A)} \int_A \log \frac{dQ_n}{dP_n} dQ_n]$$

- $$P_n(A) = \int_A \frac{dP_n}{dQ_n} dQ_n$$

- $$= Q_n(A) \frac{1}{Q_n(A)} \int_A \exp[-\log \frac{dQ_n}{dP_n}] dQ_n$$

- $$\geq Q_n(A) \exp[-\frac{1}{Q_n(A)} \int_A \log \frac{dQ_n}{dP_n} dQ_n]$$

- $$\liminf \frac{1}{n^2} \log P_n(A) \geq -\lim \frac{1}{n^2} \int \log \frac{dQ_n}{dP_n} dQ_n$$

$$P_n(A) = \int_A \frac{dP_n}{dQ_n} dQ_n$$

$$= Q_n(A) \frac{1}{Q_n(A)} \int_A \exp[-\log \frac{dQ_n}{dP_n}] dQ_n$$

$$\geq Q_n(A) \exp[-\frac{1}{Q_n(A)} \int_A \log \frac{dQ_n}{dP_n} dQ_n]$$

$$\liminf \frac{1}{n^2} \log P_n(A) \geq -\lim \frac{1}{n^2} \int \log \frac{dQ_n}{dP_n} dQ_n$$

$$\geq -I(\rho)$$

- Upper bound needs Szemeredi’s Regularity Theorem.

- Upper bound needs Szemeredi's Regularity Theorem.
- http://www.math.uchicago.edu/ may/VIGRE/VIGRE2011/REUPapers/LeeG.pdf

- Upper bound needs Szemeredi's Regularity Theorem.
- http://www.math.uchicago.edu/may/VIGRE/VIGRE2011/REUPapers/LeeG.pdf
- $\mathcal{G}$ is a graph. Its vertices are $\mathcal{X}$ and its edges are $\mathcal{E}$.

- Upper bound needs Szemeredi's Regularity Theorem.
- http://www.math.uchicago.edu/ may/VIGRE/VIGRE2011/REUPapers/LeeG.pdf
- $\mathcal{G}$ is a graph. Its vertices are $\mathcal{X}$ and its edges are $\mathcal{E}$.
- If A and B are disjoint subsets of $\mathcal{X}$ then $e(A, B)$ is the number of edges connecting A and B. $|A|$ and $|B|$ are the size or the number of vertices in $|A|$ and $|B|$.

- Upper bound needs Szemeredi's Regularity Theorem.
- http://www.math.uchicago.edu/may/VIGRE/VIGRE2011/REUPapers/LeeG.pdf
- $\mathcal{G}$ is a graph. Its vertices are $\mathcal{X}$ and its edges are $\mathcal{E}$.
- If A and B are disjoint subsets of $\mathcal{X}$ then $e(A, B)$ is the number of edges connecting A and B. $|A|$ and $|B|$ are the size or the number of vertices in $|A|$ and $|B|$.
- $r(A, B) = \frac{e(A,B)}{|A||B|} \leq 1$

- If $\mathcal{P}$ is a partition of $\mathcal{X}$ into disjoint subsets $A_1, \ldots, A_k$, then $g(\mathcal{P})$ is defined as

- If $\mathcal{P}$ is a partition of $\mathcal{X}$ into disjoint subsets $A_1, \ldots, A_k$, then $g(\mathcal{P})$ is defined as
- $$g(\mathcal{P}) = \sum_{i<j} [r(A_i, A_j)]^2 \frac{|A_i||A_j|}{n^2} \leq \sum_{i<j} \frac{|A_i||A_j|}{n^2} \leq 1$$

- If $\mathcal{P}$ is a partition of $\mathcal{X}$ into disjoint subsets $A_1, \ldots, A_k$, then $g(\mathcal{P})$ is defined as
- $$g(\mathcal{P}) = \sum_{i<j} [r(A_i, A_j)]^2 \frac{|A_i||A_j|}{n^2} \leq \sum_{i<j} \frac{|A_i||A_j|}{n^2} \leq 1$$
- We will consider partitions into $k+1$ sets where A_0 is special, in which case we define

- If $\mathcal{P}$ is a partition of $\mathcal{X}$ into disjoint subsets $A_1, \ldots, A_k$, then $g(\mathcal{P})$ is defined as
- $$g(\mathcal{P}) = \sum_{i<j} [r(A_i, A_j)]^2 \frac{|A_i||A_j|}{n^2} \leq \sum_{i<j} \frac{|A_i||A_j|}{n^2} \leq 1$$
- We will consider partitions into $k+1$ sets where A_0 is special, in which case we define
- $h(\mathcal{P}) =$
$$\sum_{1 \leq i < j \leq k} [r(A_i, A_j)]^2 \frac{|A_i||A_j|}{n^2} + \sum_{a \in A_0} \sum_{1 \leq i \leq k} [r(\{a\}, A_i)]^2 \frac{|A_i|}{n^2}$$

- A pair (A_i, A_j), $i \neq j \neq 0$ of the partition $\mathcal{P}$ is ϵ regular if

- A pair (A_i, A_j), $i \neq j \neq 0$ of the partition $\mathcal{P}$ is ϵ regular if
- For any two subsets $B_i \subset A_i$ and $B_j \subset A_j$ with $|B_i| \geq \epsilon |A_i|$ and $|B_j| \geq \epsilon |A_j|$ we have

- A pair (A_i, A_j), $i \neq j \neq 0$ of the partition $\mathcal{P}$ is ϵ regular if
- For any two subsets $B_i \subset A_i$ and $B_j \subset A_j$ with $|B_i| \geq \epsilon |A_i|$ and $|B_j| \geq \epsilon |A_j|$ we have
- $$|r(B_i, B_j) - r(A_i, A_j)| \leq \epsilon$$

- A partition $\mathcal{P}$ of the set $\mathcal{X}$ of n vertices of a graph into $k+1$ subsets $A_0, A_1, \ldots, A_k$ is called ϵ regular if

- A partition $\mathcal{P}$ of the set $\mathcal{X}$ of n vertices of a graph into $k+1$ subsets $A_0, A_1, \ldots, A_k$ is called ϵ regular if
- $|A_0| \leq \epsilon n$

- A partition $\mathcal{P}$ of the set $\mathcal{X}$ of n vertices of a graph into $k+1$ subsets $A_0, A_1, \ldots, A_k$ is called ϵ regular if
- $|A_0| \leq \epsilon n$
- $|A_1| = |A_2| = \cdots = |A_k| = d$

- A partition $\mathcal{P}$ of the set $\mathcal{X}$ of n vertices of a graph into $k+1$ subsets $A_0, A_1, \ldots, A_k$ is called ϵ regular if
- $|A_0| \leq \epsilon n$
- $|A_1| = |A_2| = \cdots = |A_k| = d$
- And out of all possible pairs A_i, A_j with $1 \leq i < j \leq k$ at most ϵk^2 are not regular.

- **Lemma.** Given $\epsilon > 0$ there is an $n_0(\epsilon)$ that satisfies the following. For any integer q there is an integer $q'(\epsilon, q) > q$ with the property that if $n \geq n_0(\epsilon)$ and $n \geq q$, for any graph with n vertices there is an ϵ regular partition of its vertices $\mathcal{X}$ into $\ell + 1$ sets $A_0, A_1, \ldots, A_\ell$ for some ℓ with $q \leq \ell \leq q'(\epsilon, q)$.

- **Lemma.** Given $\epsilon > 0$ there is an $n_0(\epsilon)$ that satisfies the following. For any integer q there is an integer $q'(\epsilon, q) > q$ with the property that if $n \geq n_0(\epsilon)$ and $n \geq q$, for any graph with n vertices there is an ϵ regular partition of its vertices $\mathcal{X}$ into $\ell + 1$ sets $A_0, A_1, \ldots, A_\ell$ for some ℓ with $q \leq \ell \leq q'(\epsilon, q)$.
- **Idea of proof. Step 1.** Suppose we have a partition $A_0, A_1, \ldots, A_k$ with $|A_1| = |A_2| = \cdots = |A_k| = d$ and $|A_0| \leq \delta n$ with $\delta < \frac{1}{4}$ and ϵk^2 pairs of A_i, A_j that are not regular.

- **Lemma.** Given $\epsilon > 0$ there is an $n_0(\epsilon)$ that satisfies the following. For any integer q there is an integer $q'(\epsilon, q) > q$ with the property that if $n \geq n_0(\epsilon)$ and $n \geq q$, for any graph with n vertices there is an ϵ regular partition of its vertices $\mathcal{X}$ into $\ell + 1$ sets $A_0, A_1, \ldots, A_\ell$ for some ℓ with $q \leq \ell \leq q'(\epsilon, q)$.
- **Idea of proof. Step 1.** Suppose we have a partition $A_0, A_1, \ldots, A_k$ with $|A_1| = |A_2| = \cdots = |A_k| = d$ and $|A_0| \leq \delta n$ with $\delta < \frac{1}{4}$ and ϵk^2 pairs of A_i, A_j that are not regular.
- We notice that the regularity condition has two parts. The size of A_0 and the regularity of all but at most ϵk^2 of the pairs in $A_1, A_2, \ldots, A_k$.

- Suppose we have a partition that is not regular and it is not because of the size of A_0. We can assume without loss of generality that $\epsilon < \frac{1}{4}$. There are at least ϵk^2 pairs of sets A_i, A_j from the collection that are not regular

- Suppose we have a partition that is not regular and it is not because of the size of A_0. We can assume without loss of generality that $\epsilon < \frac{1}{4}$. There are at least ϵk^2 pairs of sets A_i, A_j from the collection that are not regular
- Let us take one such pair A_i, A_j, with subsets $B_i \subset A_i, B_j \subset A_j$ with the property $|B_i| \geq \epsilon |A_i|, |B_j| \geq \epsilon |A_j|$ and

$$|r(B_i, B_j) - r(A_i, A_j)| \geq \epsilon$$

- Suppose we have a partition that is not regular and it is not because of the size of A_0. We can assume without loss of generality that $\epsilon < \frac{1}{4}$. There are at least ϵk^2 pairs of sets A_i, A_j from the collection that are not regular
- Let us take one such pair A_i, A_j, with subsets $B_i \subset A_i, B_j \subset A_j$ with the property $|B_i| \geq \epsilon |A_i|, |B_j| \geq \epsilon |A_j|$ and

$$|r(B_i, B_j) - r(A_i, A_j)| \geq \epsilon$$

- We refine the partition by replacing A_i, A_j by $B_i, A_i \cap B_i^c$ and $B_j, A_j \cap B_j^c$

- **Step 2.** Any refinement increases $g(\mathcal{P})$.

- **Step 2.** Any refinement increases $g(\mathcal{P})$.
- We can think of $x_{i,j} = 1$ or 0 depending on whether there is an edge or not as random variables and the ratio $r(A_i, A_j)$ as the conditional expectation given a sub σ-field. The measure is the product measure $\frac{1}{n^2}$ on any pair (i, j).

- **Step 2.** Any refinement increases $g(\mathcal{P})$.
- We can think of $x_{i,j} = 1$ or 0 depending on whether there is an edge or not as random variables and the ratio $r(A_i, A_j)$ as the conditional expectation given a sub σ-field. The measure is the product measure $\frac{1}{n^2}$ on any pair (i, j).
- $$E[|E[X|\Sigma]|^2]$$
  is increasing in Σ.

- But refining a pair that is not regular increases $g(\mathcal{P})$ by $\frac{\epsilon^4 d^2}{n^2}$

- But refining a pair that is not regular increases $g(\mathcal{P})$ by $\frac{\epsilon^4 d^2}{n^2}$
- (A, B) is not regular. There is A_1, A_2 and B_1, B_2 that make up A and B.

$$x_{ij} = e(A_i, B_j)$$

- But refining a pair that is not regular increases $g(\mathcal{P})$ by $\frac{\epsilon^4 d^2}{n^2}$
- (A, B) is not regular. There is A_1, A_2 and B_1, B_2 that make up A and B.

$$x_{ij} = e(A_i, B_j)$$

- $x = x_{11} + x_{12} + x_{21} + x_{22}$, $y_{i,j} = \frac{x_{ij}}{|A_i||B_j|}$

- But refining a pair that is not regular increases $g(\mathcal{P})$ by $\frac{\epsilon^4 d^2}{n^2}$
- (A, B) is not regular. There is A_1, A_2 and B_1, B_2 that make up A and B.

$$x_{ij} = e(A_i, B_j)$$

- $x = x_{11} + x_{12} + x_{21} + x_{22}$, $y_{i,j} = \frac{x_{ij}}{|A_i||B_j|}$
-

$$\sum_{i,j} \frac{|A_i||B_j|}{|A||B|} y_{i,j}^2 - [\frac{x}{|A||B|}]^2 =$$

- $$\sum_{i,j} \frac{|A_i||B_j|}{|A||B|}[y_{i,j} - \frac{x}{|A||B|}]^2 \geq \epsilon^4$$

- $$\sum_{i,j} \frac{|A_i||B_j|}{|A||B|}[y_{i,j} - \frac{x}{|A||B|}]^2 \geq \epsilon^4$$
- Since we can repeat this for ϵk^2 pairs $g(\mathcal{P})$ goes up by at least $\frac{\epsilon^5 k^2 d^2}{n^2}$.

- $$\sum_{i,j} \frac{|A_i||B_j|}{|A||B|}[y_{i,j} - \frac{x}{|A||B|}]^2 \geq \epsilon^4$$
- Since we can repeat this for ϵk^2 pairs $g(\mathcal{P})$ goes up by at least $\frac{\epsilon^5 k^2 d^2}{n^2}$.
- Since $n = kd + |A_0|$, $k^2 d^2 \geq \frac{1}{2} n^2$ and $g(\mathcal{P})$ goes up by $\frac{1}{2}\epsilon^5$

- $$\sum_{i,j} \frac{|A_i||B_j|}{|A||B|}[y_{i,j} - \frac{x}{|A||B|}]^2 \geq \epsilon^4$$
- Since we can repeat this for ϵk^2 pairs $g(\mathcal{P})$ goes up by at least $\frac{\epsilon^5 k^2 d^2}{n^2}$.
- Since $n = kd + |A_0|$, $k^2 d^2 \geq \frac{1}{2} n^2$ and $g(\mathcal{P})$ goes up by $\frac{1}{2}\epsilon^5$
- This can only happen a finite number of times. In fact at most $2\epsilon^{-5}$ times.

- In addition to A_0 we have at most $k2^{k-1}$ subsets of unequal size. Let us subdivide them into equal subsets of size $d' = [\frac{d}{4^k}]$

- In addition to A_0 we have at most $k2^{k-1}$ subsets of unequal size. Let us subdivide them into equal subsets of size $d' = [\frac{d}{4^k}]$
- We will have at most $k4^k$ subsets of size d'. The sets may not divide evenly and the leftover ones are thrown in A_0.

- In addition to A_0 we have at most $k2^{k-1}$ subsets of unequal size. Let us subdivide them into equal subsets of size $d' = [\frac{d}{4^k}]$
- We will have at most $k4^k$ subsets of size d'. The sets may not divide evenly and the leftover ones are thrown in A_0.
- This procedure can be repeated and will have to end after at most $2\epsilon^{-5}$ steps.

- In addition to A_0 we have at most $k2^{k-1}$ subsets of unequal size. Let us subdivide them into equal subsets of size $d' = [\frac{d}{4^k}]$
- We will have at most $k4^k$ subsets of size d'. The sets may not divide evenly and the leftover ones are thrown in A_0.
- This procedure can be repeated and will have to end after at most $2\epsilon^{-5}$ steps.
- We need to keep track of vertices piled into A_0 and estimate the size. Each step adds at most $kd'2^{k-1}$ vertices.

$$kd'2^{k-1} \le \frac{kd2^{k-1}}{4^k} \le \frac{n}{2^{k+1}}$$

- $$kd'2^{k-1} \le \frac{kd2^{k-1}}{4^k} \le \frac{n}{2^{k+1}}$$
- We repeat the subdivision at most $2\epsilon^{-5}$ times.

- $$kd'2^{k-1} \le \frac{kd2^{k-1}}{4^k} \le \frac{n}{2^{k+1}}$$
- We repeat the subdivision at most $2\epsilon^{-5}$ times.
- Let $q_\epsilon(k_0)$ be the result of iteration of the map $k \to k4^k$ repeated $2\epsilon^{-5}$ times starting from k_0. It is the largest number of sets in the partition we can end up with.

- $$kd'2^{k-1} \le \frac{kd2^{k-1}}{4^k} \le \frac{n}{2^{k+1}}$$
- We repeat the subdivision at most $2\epsilon^{-5}$ times.
- Let $q_\epsilon(k_0)$ be the result of iteration of the map $k \to k4^k$ repeated $2\epsilon^{-5}$ times starting from k_0. It is the largest number of sets in the partition we can end up with.
- If we we can control the size of the exceptional set we would be done.

- The increase in the exceptional set when we have k sets of size d is at most

$$k2^{k-1}\frac{d}{4^k} = \frac{kd}{2^{k+1}} \le \frac{n}{2^{k+1}}$$

- The increase in the exceptional set when we have k sets of size d is at most

$$k2^{k-1}\frac{d}{4^k} = \frac{kd}{2^{k+1}} \le \frac{n}{2^{k+1}}$$

- If the initial step is k_0 then at every stage $k \ge k_0$, and in $2\epsilon^{-5}$ steps it goes up by $n\epsilon^{-5}\frac{k_0}{2^{k_0}}$.

- The increase in the exceptional set when we have k sets of size d is at most

$$k2^{k-1}\frac{d}{4^k} = \frac{kd}{2^{k+1}} \leq \frac{n}{2^{k+1}}$$

- If the initial step is k_0 then at every stage $k \geq k_0$, and in $2\epsilon^{-5}$ steps it goes up by $n\epsilon^{-5}\frac{k_0}{2^{k_0}}$.
- It is less than $\frac{n\epsilon}{2}$, provided $\frac{k_0}{2^{k_0}} < 2\epsilon^6$

- The increase in the exceptional set when we have k sets of size d is at most

$$k2^{k-1}\frac{d}{4^k} = \frac{kd}{2^{k+1}} \leq \frac{n}{2^{k+1}}$$

- If the initial step is k_0 then at every stage $k \geq k_0$, and in $2\epsilon^{-5}$ steps it goes up by $n\epsilon^{-5}\frac{k_0}{2^{k_0}}$.
- It is less than $\frac{n\epsilon}{2}$, provided $\frac{k_0}{2^{k_0}} < 2\epsilon^6$
- The initial size of A_0 is at most k_0 and $k_0 < \frac{n\epsilon}{2}$ if n is large enough

- The increase in the exceptional set when we have k sets of size d is at most

$$k2^{k-1}\frac{d}{4^k} = \frac{kd}{2^{k+1}} \leq \frac{n}{2^{k+1}}$$

- If the initial step is k_0 then at every stage $k \geq k_0$, and in $2\epsilon^{-5}$ steps it goes up by $n\epsilon^{-5}\frac{k_0}{2^{k_0}}$.
- It is less than $\frac{n\epsilon}{2}$, provided $\frac{k_0}{2^{k_0}} < 2\epsilon^6$
- The initial size of A_0 is at most k_0 and $k_0 < \frac{n\epsilon}{2}$ if n is large enough
- We are done!