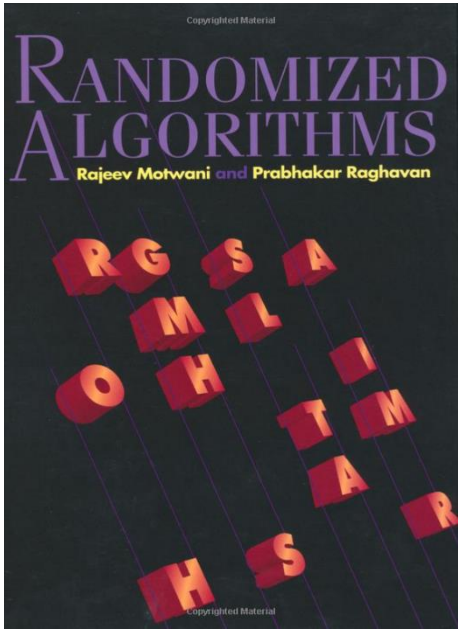
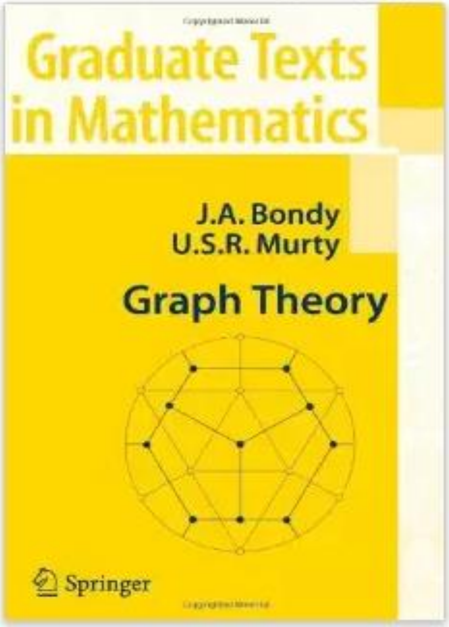
The Probabilistic Method

prepared and Instructed by
Shmuel Wimer
Eng. Faculty, Bar-Ilan University





The probabilistic method comprises two ideas:

- Any random variable assumes at least one value not smaller than its expectation.
- If an object chosen randomly from the universe satisfies a property with positive probability, there must be an object of the universe satisfying that property.

Theorem. For any undirected graph $G(V,E)$ with n vertices and m edges, there is a partition of V into A and B such that the edge cut-set has $m/2$ edges at least, namely $|\{(u,v) \in E | u \in A \text{ and } v \in B\}| \geq m/2$.

Proof. Consider the following experiment. Each vertex is independently and equiprobaly assigned to A or B.

The probability that the end points of an edge (u, v) are in different sets is ½.

By the linearity of expectation the expected number of edges in the cut is $m/2$.

It follows that there must a partition satisfying the theorem. ■

Consider the **satisfiability** problem. A set of m clauses is given in conjunctive (sum) normal form over n variables.

We have to decide whether there is a truth assignment of the n variables satisfying all the clauses (POS).

There is an optimization version called MAX-SAT where we seek for a truth assignment maximizing the number of satisfied clauses. This problem is *NP*-hard.

We subsequently show that there is always a truth assignment satisfying at least $m/2$ clauses. This is the best possible universal guarantee (consider x and $\bar{x}$).

Theorem: For any set of m clauses, there is a truth assignment satisfying at least $m/2$ clauses.

Proof: Suppose that every variable is set to TRUE or FALSE independently and equiprobaly.

For $1 \leq i \leq m$, let $Z_i = 1$ if the clause is satisfied, and $Z_i = 0$ otherwise.

Due to the **conjunctive** form, the probability that a clause containing k literals is not satisfied is $2^{-k} \leq 1/2$, or $1 - 2^{-k} \geq 1/2$ that it is satisfied, implying $\mathbf{E}[Z_i] \geq 1/2$.

The expected number of satisfied clauses is therefore $\sum_{i=1}^{m} \mathbf{E}[Z_i] \geq m/2$, implying that there must be an assignment for which $\sum_{i=1}^{m} Z_i \geq m/2$. ■

An orientation of a complete graph is called **tournament**.

A **Hamiltonian path** is an $(n-1)$-arc uni-directed path.

Theorem: (Szele 1943). There is an n-vertex tournament having at least $n!/2^{n-1}$ Hamiltonian paths.

Proof: for each vertex pair we chose an arc $v_i \to v_j$ or $v_j \to v_i$ with equal probability, generating a random tournament.

Let X be count the number of Hamiltonian paths in the tournament. X is a sum of $n!$ indicator random variables for the possibility that a path is Hamiltonian.

A Hamiltonian path occurs with probability $1/2^{n-1}$, hence $\mathbf{E}[X] = n!/2^{n-1}$, and there must be a graph with at least $n!/2^{n-1}$ Hamiltonian paths. ■

Expanding Graphs

$G(V,E)$ is called an **expanding graph** if there is a $c > 0$ such that for any $S \subseteq V$ there is $|\Gamma(S)| > c|S|$, where $\Gamma(S)$ is the set of S's neighbors. .

A particular type of expanding graph is a bipartite multi graph $G(L,R,E)$ called an **OR-concentrator**.

It is defined by a quadruple (n,d,α,c), where $|L| = |R| = n$, such that

1. $\deg(v) \leq d \; \forall v \in L$, and
2. $\forall S \subseteq L$ such that $|S| \leq \alpha n$ there is $|\Gamma(S)| > c|S|$.

In most applications it is desired to have d as small as possible and c as large as possible.

Of particular interest are those graph where α, c and d are constants independent of n and $c > 1$.

These are strict requirements and it is not trivial to construct such graphs. We rather show that such graphs exist.

We show that a random graph chosen from a suitable probability space has a positive probability of being $(n, d, \alpha, c) = (n, 18, ⅓, 2)$ OR-concentrator. (Constants are arbitrary, other combinations are possible.)

Theorem: There is an integer n_0 such that for all $n > n_0$ there is an $(n, 18, ⅓, 2)$ OR-concentrator.

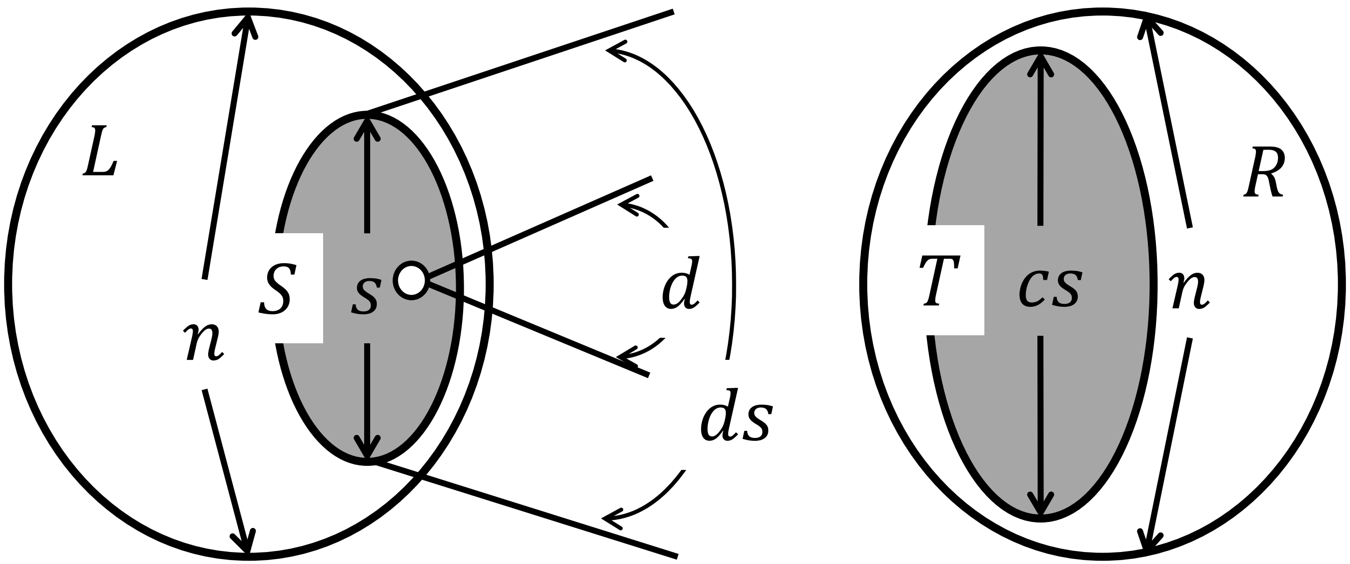
Proof: The proof is carried out in terms of d, c, and α, while the constants are pinned at the end of the proof.

Consider a random $G(L, R, E)$, where $v \in L$ choses its d neighbors $\Gamma(v) \subseteq R$ independently and uniformly with replacements, and avoid multi edges.

Let ε_s be the event that for $S \subseteq L, |S| = s$ there is $|\Gamma(S)| \leq cs$, namely, an OR-concentrator **does not exist**.

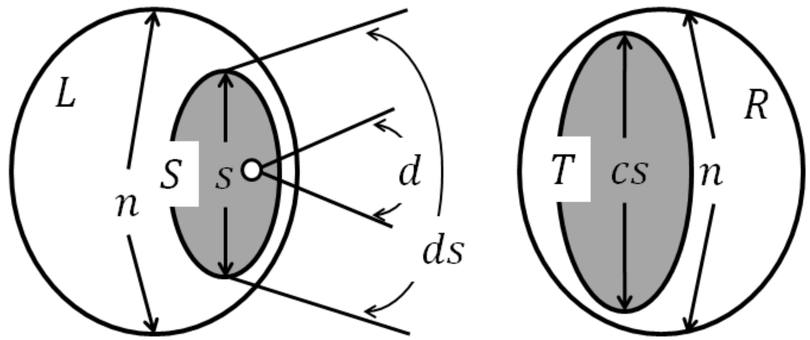
Plan: We shall first bound $\mathbf{Pr}[\varepsilon_s]$, and then sum over all the values of $s \leq \alpha n$. We thus obtain an upper bound on the probability that the random G **fails** to be an OR-concentrator with the parameters we seek.

Consider $S \subseteq L$, $|S| = s$ and any $T \subseteq R, |T| = cs$. There are $\binom{n}{s}$ ways to choose S and $\binom{n}{cs}$ ways to choose T.



There is $ds \geq |\Gamma(S)|$. The probability that $\Gamma(S) \subseteq T$ is $(cs/n)^{|\Gamma(S)|} \geq (cs/n)^{ds}$.

$|\Gamma(S)| \leq c|S|$ means **not having OR connector** (ε_s).



The number of possibilities to choose s vertices from L and cs from R is $\binom{n}{s}\binom{n}{cs}$.

The probability that all the ds edges emanating from some s vertices of L fall within any cs vertices of R is **bounded** by

$$\mathbf{Pr}[\varepsilon_s] \leq \binom{n}{s}\binom{n}{cs}\left(\frac{cs}{n}\right)^{ds}$$

Substituition of the inequality $\binom{n}{k} \le \left(\frac{ne}{k}\right)^k$ obtains

$$\mathbf{Pr}[\varepsilon_s] \le \left(\frac{ne}{s}\right)^s \left(\frac{ne}{cs}\right)^{cs} \left(\frac{cs}{n}\right)^{ds} = \left[\left(\frac{s}{n}\right)^{d-c-1} e^{1+c} c^{d-c}\right]^s$$

Using $\alpha = 1/3$ and s $\le \alpha n$, there is

$$\mathbf{Pr}[\varepsilon_s] \le \left[\left(\frac{1}{3}\right)^{d-c-1} e^{1+c} c^{d-c}\right]^s \le \left[\left(\frac{c}{3}\right)^{d} (3e)^{c+1}\right]^s$$

Using $c = 2$ and $d = 18$, there is

$$\mathbf{Pr}[\varepsilon_s] \le \left[\left(\frac{2}{3}\right)^{18} (3e)^3\right]^s = r^s,$$

where $r = (2/3)^{18}(3e)^3$, so that $r < \frac{1}{2}$.

Summing over all $1 \le s \le \alpha n = n/3$ there is

$$\sum_{n/3 \ge s \ge 1} \mathbf{Pr}[\varepsilon_s] \le \sum_{s \ge 1} r^s = \frac{r}{1-r} < 1,$$

showing that the desired **OR-concentrator exists**. ■

Crossing Number

The **crossing number** $cr(G)$ of a graph G is the smallest number of edge crossings in a planar embedding of G.

In VLSI it is the number of jumpers (via) required to layout a circuit.

For a planar graph $G(V,E)$, $|V| = n$, $|E| = m$ there is $cr(G) = 0$.

Euler formula for planar graph states $n - m + f = 2$.

Since a face comprises a least 3 edges, and each edge is shared by two faces, there is

$$0 = n - m + f - 2 \leq n - m/3 - 2.$$

Since $cr(G) = 0$ for a planar G, for any G there is

$$cr(G) \geq m - 3n + 6 \quad \text{for } n \geq 3.$$

Stronger lower bound can be derived with the aid of expectation.

Lemma: (The Crossing Lemma, proof by N. Alon). Let G be a simple graph with $m \geq 4n$. Then

$$cr(G) \geq \frac{1}{64}\frac{m^3}{n^2}.$$

Proof: Let $\tilde{G}$ be a planar embedding of G yielding $cr(G)$.

Let $S \subseteq V$ be obtained by choosing $v \in V$ randomly with probability $p := 4n/m$. Let $H := G[S]$ and $\widetilde{H} := \tilde{G}[S]$.

$\widetilde{H}$ is a planar embedding of H imposed by $\tilde{G}$.

Let X, Y and Z be the random variables of the number of vertices, number of edges and the number of crossings in $\widetilde{H}$, respectively.

It follows from the trivial lower bound that $Z := cr(\widetilde{H}) \geq cr(H) \geq Y - 3X + 6$. By linearity of expectation there is $E[Z] \geq E[Y] - 3E[X]$.

There is $E[X] = pn$ and $E[Y] = p^2m$ (an edge is defined by its two end vertices).

Since a crossing is defined by four vertices, there is $E[Z] = p^4 cr(\tilde{G}) = p^4 cr(G)$.

All in all there is

$$p^4 cr(G) \geq p^2 m - 3pn.$$

Dividing by p^4 yields

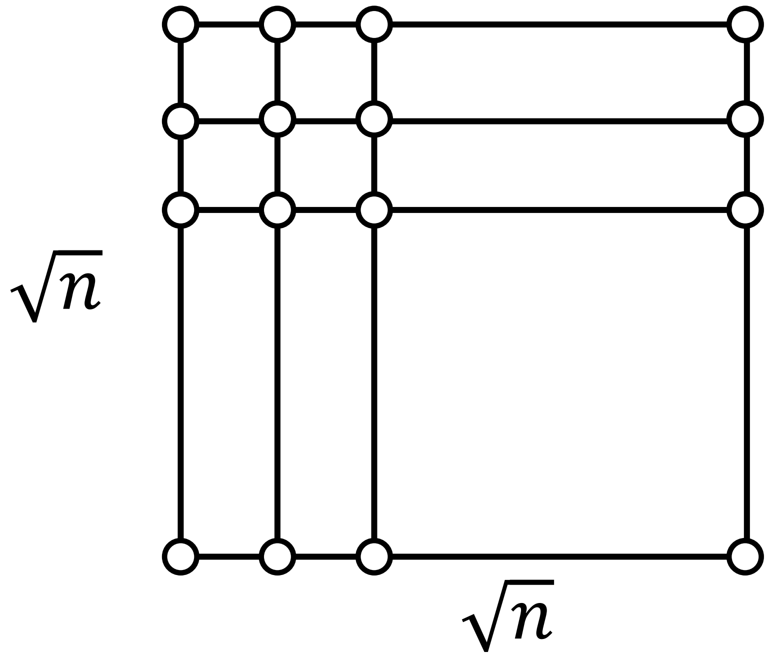
$$cr(G) \geq \frac{pm-3n}{p^3} = \frac{n}{(4n/m)^3} = \frac{1}{64}\frac{m^3}{n^2}. \blacksquare$$

The Crossing Lemma is useful in combinatorial geometry. Consider n points in the plane and lines passing through each pair of points.

Some of these $\binom{n}{2}$ at most distinct lines might pass through more than two points.

Given $k \geq 2$, how many lines can pass through at least k points?

If n is a perfect square and the point are on a $\sqrt{n} \times \sqrt{n}$ grid, there are $2\sqrt{n} + 2$ lines passing through $\sqrt{n}$ points.



Is there a configuration of n points in the plane yielding more lines passing through at least $\sqrt{n}$ points?

Theorem: (Szemerédi and Trotter 1983). Let P be a set of n points in the plane, and let l be the number of lines passing through at least $k+1$ points of P, $1 \leq k \leq 2\sqrt{n}$. Then $l < 32n^2/k^3$.

Proof: Form a graph G with vertex set P.

G's edges are the segments between consecutive points of the l lines. G has therefore at least kl edges and its crossing number is at most $\binom{l}{2}$.

If it happens that $kl < 4n$, because $1 \leq k \leq 2\sqrt{n}$, there is $l < 4n/k \leq 16n^2/k^3 < 32n^2/k^3$.

Otherwise $kl \geq 4n$, and the Crossing Lemma applies ($m = kl$).

It follows from the lemma that $l^2/2 > \binom{l}{2} \geq cr(G) \geq (kl)^3/64n^2$, yielding again $l \leq 32n^2/k^3$. ■

Properties of Almost All Graphs

Theorem: (Gilbert 1959). Let G be a random graph whose edges have constant probability p. Almost every such graph is connected.

Proof: Let us denote the graph by G^p, having n vertices. G^p can get disconnected by vertex bipartition followed by deletion of the two-sided edges.

Plan: We obtain an upper bound the probability q_n that G^p is **disconnected**, by choosing $S \subseteq V$ and summing the probabilities $P\left(\left[S,\overline{S}\right]=\emptyset\right)$ over all $\left[S,\overline{S}\right]$ partitions .

Let $|S| = k$. There are $k(n-k)$ possible edges in $[S, \overline{S}]$, so $P([S, \overline{S}] = \emptyset) = (1-p)^{k(n-k)}$. By considering all S, there is $q_n \leq \frac{1}{2}\sum_{k=1}^{n-1} \binom{n}{k} (1-p)^{k(n-k)}$.

This inequality is symmetric in k and $n-k$, so there is $q_n \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)}$.

There is $\binom{n}{k} < n^k$. Also, since in the above summation there is $n-k \geq \lceil n/2 \rceil$ and $1-p < 1$, there is $(1-p)^{k(n-k)} \leq (1-p)^{k\lceil n/2 \rceil}$.

All in all there is $q_n \leq \sum_{k=1}^{\lfloor n/2 \rfloor} \left[n(1-p)^{n/2}\right]^k$.

For sufficiently large n there is $n(1-p)^{n/2}$<1, so $q_n < \sum_{k=1}^{\infty} \left[n(1-p)^{n/2}\right]^k = \frac{n(1-p)^{n/2}}{1-n(1-p)^{n/2}}$.

We conclude that with $n \to \infty$, there is $q_n \to 0$, which means that for large enough graphs with constant edge probability the graphs is almost surely connected. ■

Markov's Inequality and Random Graphs

Let (Ω_n, P_n), $n \geq 1$ be a **probability space**, Ω_n is a sample space and $P_n : \Omega_n \to [0,1]$ a probability function satisfying $\sum_{\omega \in \Omega_n} P_n(\omega) = 1$.

We subsequently explore the existence of few properties in random large graphs.

Large means $|V[G]| = n \to \infty$, whereas the probability p of an edge depends on n and satisfies $p(n) \to 0$.

$\mathbf{G}_{n,p}$ denotes the probability space of such graphs.

Markov's Inequality states that if X is a nonnegative random variable and $t \in \mathbb{R}, t > 0$, then

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Markov's Inequality is applied to show that $G \in \mathbf{G}_{n,p}$ almost surly has a particular property for a certain p.

It is obtained by setting $X = X_n$ and $t = 1$.

Corollary: Let $X_n \in \mathbb{N}$ be a nonnegative random variable in a probability space $(\Omega_n, P_n),\ n \geq 1$. If $E(X_n) \to 0$ as $n \to \infty$, then $P(X_n = 0) \to 1$ as $n \to \infty$.

Asymptotic Behavior of Graphs

Example: We are interested in the number X of triangles in $G \in \mathbf{G}_{n,p}$.

X can be expressed as the sum

$$X = \sum\{X_S : S \subseteq V, |S| = 3\},$$

where X_S is the indicator random variable for the event A_S that $G[S]$ is a triangle.

$X_S = 1$ if S imposes a triangle and $X_S = 0$ otherwise. By the expectation definition there is

$$E(X_S) = P(X_S = 1).$$

There is $P(A_S) = p^3$.

By linearity of expectation, there is

$$E(X) = \sum\{E(X_S) : S \subseteq V,\ |S| = 3\} = \binom{n}{3} p^3 < (pn)^3.$$

Thus if $pn \to 0$ as $n \to \infty$, then$E(X) \to 0$, so $P(X = 1)$ $\to 0$ and $P(X = 0) \to 1$.

It means that if $pn \to 0$ as $n \to \infty$, G will almost surly be triangle-free. ■

Consider the probability of having the independent sets in a graph of n vertices and edge probability p, not exceeding a certain size, which of course depends on n.

Theorem: (Erdös 1961). The size of maximal independent set in a random grap $\alpha\left(G \in \mathbf{G}_{n,p}\right)$ is almost surely no larger than $\lceil 2p^{-1}\log n \rceil$.

The theorem states that if the probability of an edge is fixed, it is very difficult to find an independent set of size that grows with n, even very slowly as $\log n$.

Proof: Let $S \subset V[G]$, $|S| = k+1$, $k \in \mathbb{N}$. k is pinned down later.

The probability that S is an independent set is the probability that none of the vertex pairs has a connecting edge, namely, $(1-p)^{\binom{k+1}{2}}$.

Let A_S be the event that S is an independent set and let X_S be the corresponding indicator random variable.

There is $E(X_S) = P(X_S = 1) = P(A_S) = (1-p)^{\binom{k+1}{2}}$.

Let Z be the number of independent sets of size $k+1$. Then

$$Z = \sum\{X_S : S \subset V,\ |S| = k+1\}.$$

By linearity of expectation there is

$$E(Z) = \sum\{E(X_S) : S \subset V,\ |S| = k+1\} =$$

$$\binom{n}{k+1}(1-p)^{\binom{k+1}{2}}.$$

There is $\binom{n}{k+1} \leq \frac{n^{k+1}}{(k+1)!}$ and $1 - p < e^{-p}$.

Substitution in $E(Z)$ yields

$$E(Z) \leq \frac{n^{k+1} e^{-p\binom{k+1}{2}}}{(k+1)!} = \frac{\left(ne^{-pk/2}\right)^{k+1}}{(k+1)!}$$

Let us now pin down k, supposing $k = \lceil 2p^{-1} \log n \rceil$.

Then $k \geq 2p^{-1} \log n$, and by exponentiation there is $ne^{-pk/2} \leq 1$, hence

$$E(Z) \leq \frac{1}{(k+1)!}$$

Since $k \geq 2p^{-1} \log n$, k grows at least as fast as $\log n$, hence $E(Z) \to 0$ as $n \to \infty$.

Recall the corollary stating that if $E(Z) \to 0$ as $n \to \infty$, then $P(Z = 0) \to 1$ as $n \to \infty$.

It means that $\alpha\left(G \in \mathbf{G}_{n,p}\right) \leq 2p^{-1} \log n$ with probability$\to$ 1 as $n \to \infty$, so $\alpha\left(G \in \mathbf{G}_{n,p}\right) \geq 2p^{-1} \log n$ with probability$\to$ 0 as $n \to \infty$. ■

The **distance** between two vertices is defined as the edge length of the shortest path connecting them.

The **diameter** of a graph is the maximum of the distance over all vertex pairs.

Theorem. If p is a constant then almost every G^p has diameter 2 (and hence connected).

Proof. Let $X(G^p)$ count the number of unordered vertex pairs which distance is larger than 2, hence having no common neighboring vertex.

If there are none such pairs, then G^p is connected and has diameter 2.

$X(G^p)$ is a random variable. If it would happen that $E(X) \to 0$ as $|V| = n \to \infty$ then it follows by Markov's Inequality that the theorem holds.

For two vertices $\{v_i, v_j\} \in V$ let X_{ij} be an indicator random variable specifying that they do not share a common neighboring vertex.

$X_{ij} = 1$ would happen if there is no common neighboring vertex.

For each of the other $n-2$ vertices the probability it does not connect to either of $\{v_i, v_j\}$ is $1 - p^2$. Hence $P\left(X_{ij} = 1\right) = (1 - p^2)^{n-2}$.

There are $\binom{n}{2}$ distinct vertex pairs. X is bounded by the sum of the $\binom{n}{2}$ random variables X_{ij}.

If follows from the linearity of expectation that $E(X) \leq \binom{n}{2}(1 - p^2)^{n-2}$.

Since p is constant while $n \to \infty$, there is $E(X) \to 0$. Consequently, almost every G^p has diameter 2, and is also connected. ■

This theorem is stronger than Gilbert's theorem. While the latter states that almost every G^p is connected, this one provides also the diameter.

Problem

A graph G is planar if and only if for any $H \subseteq G$, there is $H \neq K_5$ and $H \neq K_{3,3}$.

Let $G[U,V]$ be bipartite random graph with $|U| = |V| = n$, whose edges have probability $p(n)$ (non constant!).

Find the largest function $f(n)$ such that if $p(n) = o[f(n)]$ then almost every $G[U,V]$ is planar as $n \to \infty$.

Proof: We should find what probability $f(n)$ ensures that there is almost surely no $K_{3,3} \subset G[U,V]$.

Let X be the number of $K_{3,3}$ in $G[U,V]$.

There are $\binom{n}{3}^2$ distinct subgraphs $G[W,Z]$, where $W \subset U,\ Z \subset V,$ and $|W| = |Z| = 3$.

Let $X_{W,Z}$ be an indicator random variable of the event $G[W,Z] = K_{3,3}$. There is

$$E\left(X_{W,Z}\right) = P\left(G[W,Z] = K_{3,3}\right) = p(n)^9.$$

By linearity of expectation, there is

$$E(X) = \sum\{E(X_{W,Z}) : W \subset U, Z \subset V, |W| = |Z| = 3\}$$
$$= \binom{n}{3}^2 p(n)^9 < n^6 p(n)^9.$$

Thus if $n^6 p(n)^9 \to 0$ as $n \to \infty$, then $E(X) \to 0$, so $P(X = 1) \to 0$ and $P(X = 0) \to 1$.

Consequently

$$n^6 f(n)^9 = O(1) \Rightarrow f(n) = n^{-2/3}. \blacksquare$$