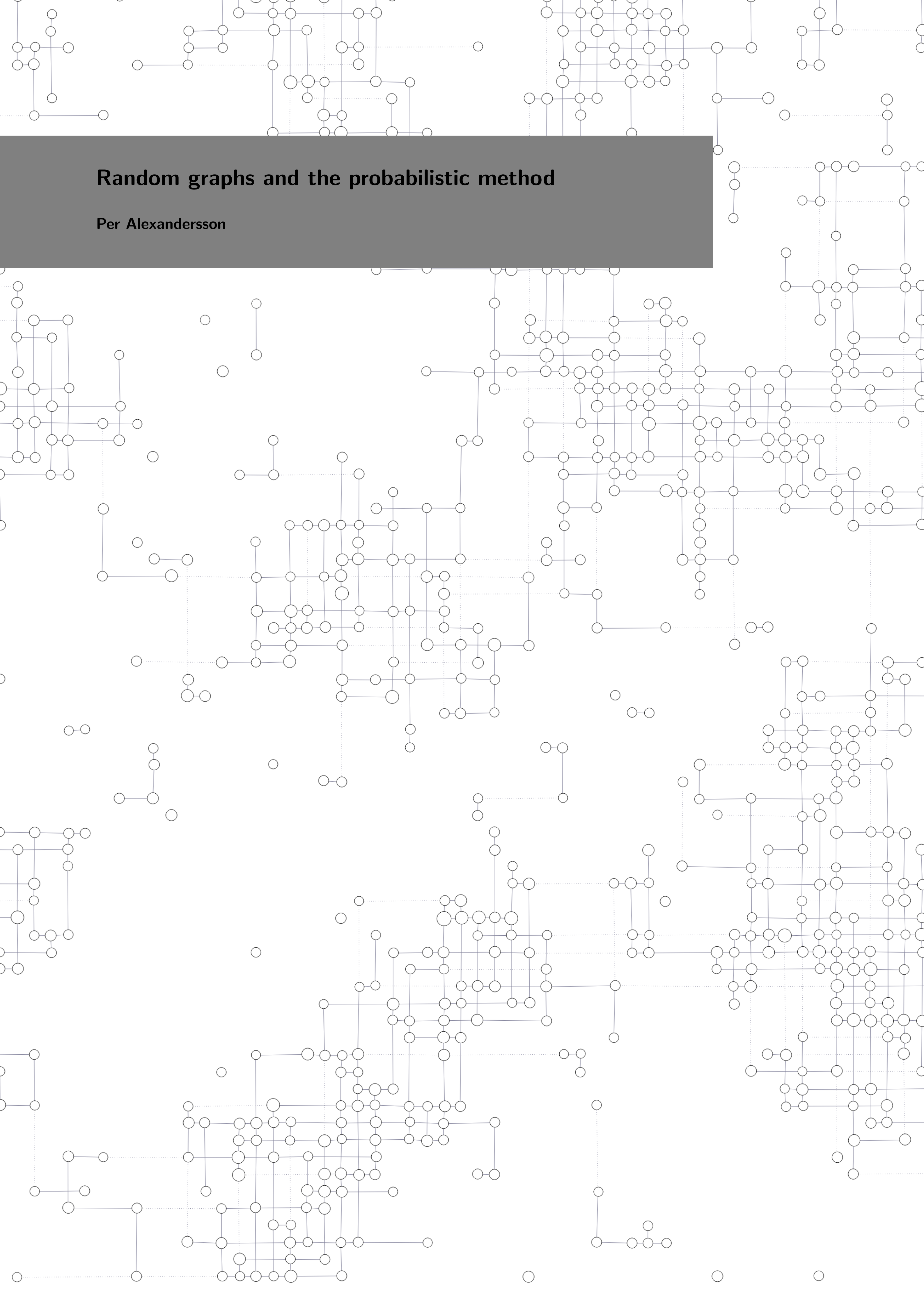


Random graphs and the probabilistic method

Per Alexandersson

Introduction

This is a brief overview of random graphs and the first and second moment methods. Questions and suggestions are welcome at `per.w.alexandersson@gmail.com`.

Version: 9th November 2023,21:51

RECALL THE STANDARD NOTATION where $\mathbb{E}[X]$ denotes the expected value of a random variable, and the variance is defined as

$$\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{1}$$

Recall that if X is a *discrete* random variable taking values in $\mathbb{N}$, then

$$\mathbb{E}[X] = \sum_{k\in\mathbb{N}} k \cdot \mathbb{P}[X = k]. \tag{2}$$

WE FIRST REVIEW some of the useful inequalities when working with the probabilistic method. We let X denote a *non-negative* random variable. In many cases, we will consider a sequence $\{X_n\}_{n=1}^{\infty}$ of such random variables. We also impose the condition that $\mathbb{E}[X]$ is finite.

CAUCHY–SCHWARZ INEQUALITY: For random variables X and Y, we have

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]. \tag{3}$$

MARKOV'S INEQUALITY: For any $a > 0$,

Markov's inequality is what is known as the *first moment method*.

$$\mathbb{P}[X > 0] \leq \mathbb{E}[X] \quad \text{and} \quad \mathbb{P}[X > a] \leq \frac{\mathbb{E}[X]}{a}. \tag{4}$$

In particular, if $\mathbb{E}[X_n] \to 0$, then $X_n = 0$ almost always, as $n \to \infty$.

CHEBYSHEV'S INEQUALITY: For all random variables X,

$$\mathbb{P}[|X - \mathbb{E}[X]| > \lambda] \leq \frac{\mathrm{Var}[X]}{\lambda^2} \text{ for all } \lambda > 0. \tag{5}$$

PALEY–ZYGMUND INEQUALITY:

$$\mathbb{P}[X \geq \alpha\, \mathbb{E}[X]] \geq (1-\alpha)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \qquad \text{for all } 0 < \alpha < 1. \tag{6}$$

SECOND MOMENT METHOD (I):

$$\mathbb{P}[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X^2]} \qquad \text{assuming } \mathbb{E}[X^2] \neq 0. \tag{7}$$

SECOND MOMENT METHOD (II):

$$\mathbb{P}[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \qquad \text{assuming } \mathbb{E}[X^2] \neq 0. \tag{8}$$

CHERNOFF'S BOUND: Let $X_1, \ldots, X_n$ be *independent* unbiased random variables with values in $\{-1, 1\}$, $X = X_1 + \cdots + X_n$. Then

$$\mathbb{P}[X \geq t] \leq e^{-\frac{t^2}{2n}}. \tag{9}$$

Proofs of the statements in the introduction

Proof of Cauchy–Schwarz inequality. As the random variable $(X - aY)^2$ is non-negative for all $a \in \mathbb{R}$, we have

$$0 \leq \mathbb{E}[(X - aY)^2] = \mathbb{E}[X^2] + a^2\mathbb{E}[Y^2] - 2a\mathbb{E}[XY].$$

We rearrange and obtain

$$2a\mathbb{E}[XY] - a^2\mathbb{E}[Y^2] \leq \mathbb{E}[X^2]. \tag{10}$$

By choosing $a = \mathbb{E}[XY]/\mathbb{E}[Y^2]$, we can easily deduce the inequality.

Substituting the value of a in (10):

$$2\frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]^2}\mathbb{E}[Y^2] \leq \mathbb{E}[X^2]$$

$$\iff$$

$$2\frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} - \frac{\mathbb{E}[XY]^2}{\mathbb{E}[Y^2]} \leq \mathbb{E}[X^2]$$

Proof of Markov's inequality. Let $\mathbf{1}_{\{X\geq k\}}$ denote the random variable that is 1 if $X \geq k$ and 0 otherwise. We have that

$$\mathbb{E}[X] = \mathbb{E}[X \cdot \mathbf{1}_{\{X\geq k\}} + X \cdot \mathbf{1}_{\{X<k\}}] \geq k \cdot \mathbb{E}[\mathbf{1}_{\{X\geq k\}}] = k \cdot \mathbb{P}[X \geq k].$$

This implies Markov's inequality.

Proof of Chebyshev's inequality. The random variable $Y = (X - \mathbb{E}[X])^2$ is non-negative, so we can apply Markov's inequality. In other words,

$$\mathbb{P}[Y \geq k] \leq \frac{\mathbb{E}[Y]}{k}.$$

By letting $k = \lambda^2$ and substituting, we have

$$\mathbb{P}[(X - \mathbb{E}[X])^2 \geq \lambda^2] \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\lambda^2}.$$

This statement is now equivalent with Chebyshev's inequality.

In the right hand side, recall that $\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Proof of Paley–Zygmund inequality. Let $\mu := \mathbb{E}[X]$ and note that

$$\mathbb{E}[X] = \mathbb{E}[X \cdot \mathbf{1}_{X<\alpha\mu}] + \mathbb{E}[X \cdot \mathbf{1}_{X\geq\alpha\mu}].$$

The product of the random variables $X \cdot \mathbf{1}_{X<\alpha\mu}$ never exceeds $\alpha\mu$, so $\mathbb{E}[X \cdot \mathbf{1}_{X<\alpha\mu}] \leq \alpha\mu$. In the second term, we apply Cauchy–Schwarz's inequality and get

$$\mu \leq \alpha\mu + \sqrt{\mathbb{E}[X^2]\ \mathbb{E}[\mathbf{1}_{X\geq\alpha\mu}]}.$$

Rearranging the sides and squaring gives the final inequality

Note $\mathbb{E}[\mathbf{1}_{X\geq\alpha\mu}] = \mathbb{P}[X \geq \alpha\mu]$.

$$(1 - \alpha)^2\mu^2 \leq \mathbb{E}[X^2]\ \mathbb{P}[X \geq \alpha\mu]$$

which after division by $\mathbb{E}[X^2]$ gives exactly what we wished to prove.

Proof of Second moment methods. The first inequality follows immediately from the Paley–Zygmund inequality. For the second case, note that the first case implies

Take $\alpha \to 0$ in (6).

$$1 - \mathbb{P}[X = 0] \leq 1 - \frac{\mathrm{Var}[X]}{\mathbb{E}[X^2]}$$

Rewriting both sides now gives

$$\mathbb{P}[X > 0] \leq \frac{\mathbb{E}[X^2] - \mathrm{Var}[X]}{\mathbb{E}[X^2]}$$

which together with (1) implies the desired inequality.

PROOF OF THE CHERNOFF BOUND. Recall that $X_1, \dots, X_n$ are unbiased random variables[1] taking values in $\{-1, 1\}$. For any λ, we then have

$$\mathbb{E}[e^{\lambda X_i}] = \frac{1}{2}e^{\lambda} + \frac{1}{2}e^{-\lambda} \le e^{\frac{\lambda^2}{2}}. \tag{11}$$

[1] That is, 1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$.
The last inequality follows from comparing Taylor coefficients.

We now have

$$\mathbb{E}\left[e^{\lambda X}\right] = \mathbb{E}\left[e^{\lambda(X_1+\cdots+X_n)}\right] = \mathbb{E}\left[e^{\lambda X_1}e^{\lambda X_2}\cdots e^{\lambda X_n}\right].$$

Since the X_i are independent, we can interchange product and expectation. After using (11) we arrive at

Note $\mathbb{E}[e^{\lambda X_i}] \le e^{\frac{\lambda^2}{2}}$.

$$\mathbb{E}\left[e^{\lambda X}\right] \le e^{n\frac{\lambda^2}{2}}. \tag{12}$$

Markov's inequality applied to (12) now tells us that for any $a > 0$, we have

$$\mathbb{P}[e^{\lambda X} > a] \le \frac{e^{n\frac{\lambda^2}{2}}}{a}.$$

Putting $a = e^{\lambda t}$, we have

$$\mathbb{P}[e^{\lambda X} > e^{\lambda t}] = \mathbb{P}[X > t]$$

$$\mathbb{P}[X > t] \le e^{n\frac{\lambda^2}{2} - \lambda t}. \tag{13}$$

Note now this is true for any $\lambda > 0$, so we choose λ such that the right hand side is minimized. This gives the final inequality

Let $\lambda = t/n$.

$$\mathbb{P}[X > t] \le e^{-\frac{t^2}{2n}}. \tag{14}$$

More background

IN ORDER TO SHOW THAT a non-negative random variable X_n is almost always 0 as $n \to \infty$, it is enough to use the first moment method. That is, it is enough to show that $\mathbb{E}[X_n] \to 0$.

However, if we want to show that a random variable is *positive* with probability 1, *is is not enough* to show that $\mathbb{E}[X_n] \to \infty$. For example, the random variable

$$X_n := \begin{cases} 0 & \text{with probability } 1 - 1/n \\ n^2 & \text{with probability } 1/n \end{cases} \tag{15}$$

has the property that $\mathbb{E}[X_n] \to \infty$ but $\mathbb{P}[X_n = 0] \to 1$. The issue is in some sense that X_n has too much variance. This is what the second moment method allows us to control, so that X_n does not deviate too much from $\mathbb{E}[X_n]$.

$\mathbb{E}[X_n] = n$

Graphs with several properties

IN RAMSEY THEORY, we sometimes want to construct a graph with two or more properties, say P and P'. We have no real control over the dependence[2] between these properties. However, it is easy to show[3] that

[2] Knowing that G has property P might or might not increase the likelihood of G having property P'.

[3]

$$\begin{aligned}\mathbb{P}[A \wedge B] &= 1 - \mathbb{P}[A^c \vee B^c] \\ &\ge 1 - \mathbb{P}[A^c] - \mathbb{P}[B^c]\end{aligned}$$

$$\begin{aligned}\mathbb{P}[G \text{ has property } P \text{ and } P'] \geq 1 - \mathbb{P}[G \text{ has property } \neg P] \\ - \mathbb{P}[G \text{ has property } \neg P'].\end{aligned}$$

If we can show either of the following two (equivalent) inequalities

$$\mathbb{P}[G \text{ has property } \neg P] + \mathbb{P}[G \text{ has property } \neg P'] < 1 \tag{16}$$

$$\mathbb{P}[G \text{ has property } P] + \mathbb{P}[G \text{ has property } P'] > 1 \tag{17}$$

we can conclude that $\mathbb{P}[G$ has property P and $P'] > 0$. There must therefore exist at least one graph G with both P and P' as properties.

Threshold functions

We let $\mathcal{G}(n,p)$ be the Erdős–Rényi graph with n vertices and each edge appearing with probability p. A common problem in random graph theory is to consider a random variable describing some property of $G \in \mathcal{G}(n,p)$. A function $\tau(n)$ is said to be a *threshold function* for the property P, if we have

$$\lim_{n\to\infty} \mathbb{P}[G \in \mathcal{G}(n,p_n) \text{ has property } P] = \begin{cases} 0 & \text{if } p_n \ll \tau(n) \\ 1 & \text{if } p_n \gg \tau(n). \end{cases} \tag{18}$$

Ordo notation and asymptotic analysis

Exponential inequality: For $0 \leq p \leq 1$, we have that

$$1 - p \leq e^{-p} \qquad \text{and} \qquad (1-p)^n \leq e^{-pn}. \tag{19}$$

Stirling's formula:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1 + O(1/n)). \tag{20}$$

Basic binomial inequalities:

$$\left(\frac{n}{k}\right)^n \leq \binom{n}{k} \leq \frac{n^k}{k!}. \tag{21}$$

Central binomial coefficient inequality:

Theorem 1. *We have that* $\binom{2n}{n} \geq 4^n/\sqrt{4n+2}$.

Proof. We let $X = Y_1 + Y_2 + \cdots + Y_{2n}$ where the Y_i are independent fair[4] coin tosses.

Evidently $\mathbb{E}[X] = n$ and with some more work, one can show[5] that $\mathrm{Var}[X] = n/2$. By using Chebyshev's inequality we can prove that

$$\mathbb{P}[|X-n| < \sqrt{n}] \geq \frac{1}{2}.$$

However, we can explicitly compute $\mathbb{P}[|X-n| < \sqrt{n}]$, as X is a discrete random variable. We have

$$\mathbb{P}[|X-n| < \sqrt{n}] = \sum_{\substack{k \\ -\sqrt{n} \leq k \leq \sqrt{n}}} \mathbb{P}[X = k + n].$$

[4] $\mathbb{P}[Y_i = 1] = \mathbb{P}[Y_i = 0] = \frac{1}{2}$

[5] Note that

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_i \mathbb{E}[Y_i^2] + 2\sum_{i<j} \mathbb{E}[Y_i]\mathbb{E}[Y_j] \\ &= n + 2\binom{2n}{2} \cdot (1/2)^2 \\ &= \frac{n}{2} + n^2.\end{aligned}$$

We now have $\mathbb{P}[X = k+n] = \binom{2n}{n+k}2^{-2n}$ by deciding which of the $n+k$ random variables are equal to 1. The central binomial coefficient is largest, so $\binom{2n}{n+k} \leq \binom{2n}{n}$, for all k. The sum runs over $1+2\sqrt{n}$ values in total so we may conclude that

$$\frac{1}{2} \leq \mathbb{P}[|X-n| < \sqrt{n}] \leq (1+2\sqrt{n})\binom{2n}{n}2^{-2n}.$$

Finally, rearranging the inequality gives

$$\frac{4^n}{2(1+2\sqrt{n})} \leq \binom{2n}{n}$$

and this is what we wanted to prove. □

PROBABILITY GENERATING FUNCTION: Let X be a random variable, taking values in $\{0, 1, 2, \dots\}$. The *probability generating function* is defined as

$$P_X(u) = \sum_{k \geq 0} \mathbb{P}[X = k]u^k.$$

By definition, $P_X(1) = 1$. For example, let X be the random variable giving the number of coin flips needed to reach heads. Clearly, $\mathbb{P}[X = k] = 2^{-k}$, so

$$\begin{aligned} P_X(u) &= 2^{-1} + 2^{-2}u + 2^{-3}u^2 + \cdots \\ &= \frac{1}{2}(1 + (u/2) + (u/2)^2 + \cdots) \\ &= \frac{1}{2} \cdot \frac{1}{1 - \frac{u}{2}}. \end{aligned}$$

CHARACTERISTIC FUNCTION: Let X be a random variable from some probability distribution. The *characteristic function* $\phi_X(t)$ of X is defined as

$$\phi_X(t) := \mathbb{E}[e^{itX}].$$

The characteristic function completely determines the probability distribution. We note that the *Gaussian distribution* $N(0,1)$ has characteristic function

$$e^{-t^2/2}.$$

Lévy's continuity theorem: A sequence $X_1, X_2, \dots$ of random variables converge in distribution to X if and only if the corresponding characteristic functions converge pointwise to some $\phi_X(t)$, which is continuous in the origin.

One can show[6] that the characteristic function and the probability generating function are related:

$$P_X(e^{it}) = \phi_X(t). \tag{22}$$

[6] Let $Y = e^{itX}$. Then

$$\begin{aligned} P_X(e^{it}) &= \sum_{k \geq 0} \mathbb{P}[X = k]e^{itk} \\ &= \mathbb{E}[Y] \\ &= \mathbb{E}[e^{itX}] \\ &= \phi_X(t). \end{aligned}$$

First moment method examples

Another Chernoff-type inequality

Let $X_1, \dots, X_n$ be independent Bernoulli random variables, where $\mathbb{P}[X_i = 1] = 1/3$ for all i and let $\mu = \mathbb{E}[X_1 + \dots + X_n]$. Then

$$\mathbb{P}[X_1 + X_2 + \dots + X_n \geq \mu + t] \leq e^{-\frac{3t^2}{8n}}.$$

We prove this as follows. First note that $\mu = n/3$ so we can instead try to bound

$$\mathbb{P}\left[\left(X_1 - \frac{1}{3}\right) + \dots + \left(X_n - \frac{1}{3}\right) \geq t\right].$$

Introduce the random variables[7] $Y_i := X_i - 1/3$ (which are also pairwise independent). Following the same approach as in the Chernoff bound proof, we have

[7] Note $\mathbb{P}[Y_i = 2/3] = 1/3$ and $\mathbb{P}[Y_i = -1/3] = 2/3$.

$$\mathbb{E}[e^{\lambda Y_1}] = \frac{1}{3}e^{\frac{2\lambda}{3}} + \frac{2}{3}e^{-\frac{\lambda}{3}}.$$

Now there is some $C \geq 27/4$ such that for all $\lambda \in \mathbb{R}$ the inequality

$$\frac{1}{3}e^{\frac{2\lambda}{3}} + \frac{2}{3}e^{-\frac{\lambda}{3}} \leq e^{\frac{\lambda^2}{C}} \tag{23}$$

holds. We can prove this statement as follows. Note that (23) is equivalent with

$$e^{2\lambda} + 2e^{-\lambda} \leq 3e^{\frac{9\lambda^2}{C}}.$$

Since $2 \leq x + x^{-1}$ for all $x > 0$, we have that

$$e^{2\lambda} + 2e^{-\lambda} \leq e^{2\lambda} + (e^{\lambda} + e^{-\lambda})e^{-\lambda} = 1 + e^{2\lambda} + e^{-2\lambda}.$$

Hence, it suffices to find some C such that for all $\lambda > 0$,

$$1 + e^{\lambda} + e^{-\lambda} \leq 3e^{\frac{9\lambda^2}{4C}}.$$

Comparing coefficients in the Taylor expansion shows that $C = 27/4 = 6.75$ works. In fact, $C = 8$ seem[8] to be the optimal value in (23).

[8] See Hoeffding's inequality, `https://en.wikipedia.org/wiki/Hoeffding%27s_inequality`

A SIMPLE TRANSFORMATION SHOWS THAT for any $\lambda \geq 0$,

$$\mathbb{P}[Y_1 + \dots + Y_n \geq \lambda t] = \mathbb{P}\left[e^{\lambda(Y_1 + \dots + Y_n)} \geq e^{\lambda t}\right].$$

The right hand side can be bounded above by Markov's inequality. Furthermore, the fact that the random variables are independent implies that

$$\mathbb{P}\left[e^{\lambda(Y_1 + \dots + Y_n)} \geq e^{\lambda t}\right] \leq \frac{\mathbb{E}[e^{\lambda Y_1}]^n}{e^{\lambda t}}.$$

By now using (23) in the right hand side, we can then conclude that

$$\mathbb{P}[Y_1 + \dots + Y_n \geq t] \leq e^{-\lambda t + n\lambda^2/C}.$$

We take the derivative with respect to λ, and find that the minimum of the right hand side is attained at $\lambda = Ct/(2n)$. Inserting this value gives[9] the final inequality

$$\mathbb{P}\left[Y_1 + \cdots + Y_n \geq t\right] \leq e^{-C\frac{t^2}{4n}}.$$

[9]
$$-\frac{Ct}{2n}t + n/C\left(\frac{Ct}{2n}\right)^2 = -\frac{Ct^2}{2n} + \frac{Ct^2}{4n} = -\frac{Ct^2}{4n}$$

Since we know $C = 27/4$ is a value for which this is true, we have

$$\mathbb{P}\left[Y_1 + \cdots + Y_n \geq t\right] \leq e^{-\frac{27t^2}{16n}}.$$

A SLIGHLY DIFFERENT APPROACH gives the weaker bound

$$\mathbb{P}\left[Y_1 + \cdots + Y_n \geq t\right] \leq e^{t-(n/3+t)\log(1+3t/n)},$$

and this is not as explicit.

Vectors

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be arbitrary vectors of length 1 in $\mathbb{R}^n$. Show that one can find signs $b_i \in \{-1, 1\}$ such that

$$|b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n| \leq \sqrt{n}.$$

We let each b_i be a uniform $\{0, 1\}$ random variable X_i. Then the expected length of such a linear combination of unit vectors is

Recall that $|\mathbf{w}|^2$ is equal to the scalar product $\mathbf{w} \cdot \mathbf{w}$.

$$\mathbb{E}\left[|X_1\mathbf{v}_1 + X_2\mathbf{v}_2 + \cdots + X_n\mathbf{v}_n|^2\right] = \mathbb{E}\left[\sum_{i,j} X_iX_j\,\mathbf{v}_i \cdot \mathbf{v}_j\right].$$

By linearity of expectation, and using the fact that the X_i are independent, we get

$$\sum_{i \neq j} \mathbb{E}[X_i]\mathbb{E}[X_j]\,\mathbf{v}_i \cdot \mathbf{v}_j \; + \; \sum_i \mathbb{E}[X_i^2]\,|\mathbf{v}_i|^2.$$

Now, since $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = 1$, the first sum vanish and the second sum is simply n. Thus, the expected length is $\sqrt{n}$ and we conclude that there must[10] be some choice of signs for which the length is at most $\sqrt{n}$.

[10] This is simply the first moment principle.

Hamiltonian paths in tournaments

A *tournament* is a complete graph on n vertices, where each edge is directed. Hence, there are $2^{\binom{n}{2}}$ such tournaments. Show that there is a tournament which contains at least $n!/2^{n-1}$ Hamiltonian[11] path.

[11] Such a path must respect the edge directions

We first note that there are exactly $n!$ possible Hamiltonian paths, as we visit all vertices in some order. For each such path P, let Y_P the indicator variable that P is indeed Hamiltonian, i.e, all edges are directed in the appropriate manner. Since there are $n-1$ edges in every such path, the probability that P is a Hamiltonian path is 2^{n-1}. Hence, the expected number of Hamiltonian paths in a tournament is $n!2^{n-1}$. This implies the existence of a tournament with at least $n!2^{n-1}$ Hamiltonian paths.

Expected diameter

WE SHALL SHOW THAT THE EXPECTED DIAMETER in a random graph is 2 asymptotically. Let $G \in \mathcal{G}(n,p)$ where $0 < p < 1$.

The *diameter* of a graph is defined as $\max_{u,v \in G} |Path(u,v)|$ where $Path(u,v)$ is the length of the shortest path connecting u and v.

Let u and v be different vertices in G. A path of length 2 from u to v is of the form $u \to w \to v$, and there is a one-to-one correspondence between $w \in V \setminus \{u,v\}$ and such paths. The probability that $u \to w \to v$ is a path is p^2, so

$$\mathbb{P}[u \text{ and } v \text{ are not connected by a path of length } 2]$$

is equal to $(1-p^2)^{n-2}$. The expected number of vertices *not* connected by such a path in $G \in \mathcal{G}(n,p)$ is therefore $\binom{n}{2}(1-p^2)^{n-2}$, and this goes to 0 as $n \to \infty$. Now let X denote the number of pairs of vertices with distance at least 3. Clearly, $\mathbb{E}[X] \leq \binom{n}{2}(1-p^2)^{n-2}$ so $\mathbb{E}[X] \to 0$. Markov's inequality[12] now implies that almost surely there are no vertices in G with distance more than two apart, as $n \to \infty$. On the other hand, the probability that all pairs of vertices are at distance 1 from one another is $p^{n(n-1)/2}$, which goes to 0. Hence, the diameter is 2 almost surely.

[12] $\mathbb{P}[X > 0] \leq \mathbb{E}[X]$

We can do a bit better, and show that with $p = \sqrt{2}\sqrt{\ln(n)/n}$ is a threshold for the property of having diameter 2.

See https://www.cs.cmu.edu/~avrim/598/chap4only.pdf

Minimal cut

Let $G \in \mathcal{G}(n,p)$. Let us partition the vertices $V = S \cup T$ into non-empty sets with size k and $n-k$, respectively. The expected number of edges between S and T is then $k(n-k)p$, since there are $k(n-k)$ edges, and each of them appear with probability p. This is minimized when $k = 1$, so in order to have a minimal cut, we should take $S = \{s\}$, the source, or $S = V \setminus \{t\}$. As we wish to find the minimal cut, we should take

$$\mathbb{E}[\min(X,Y)]$$

the expected minimum of two random variables with $\mathbb{P}[X = k] = \binom{n-1}{k}p^k(1-p)^{n-1-k}$. Note that $X \sim N(n/2, np(1-p))$ so

$$\mathbb{E}[\min(X,Y)] \approx \frac{n}{2} - \sqrt{np(1-p)}\min(N_1,N_2)$$

where N_1, N_2 are normal distributions with mean 0 and variance 1. One can show that $\min(N_1,N_2) = \pi^{-1}$, so the answer should be

$$\frac{n}{2} - \frac{\sqrt{np(1-p)}}{\pi}.$$

Perfect matchings

WE LET X_n BE THE RANDOM VARIABLE which count the number of perfect matchings of $G \in \mathcal{G}(2n,p)$. Note that the number of vertices is even. Find the expected value for the number of matchings, and a threshold function for the existence of a perfect matching.

We can analyze the situation when we express X_n as a sum of indicator variables. We let Y_S be the indicator variable that the edges in S are present. We can compute the total number of sets S, for which the edges form a perfect matching. There are

$$\frac{1}{n!}\binom{2n}{2,2,\ldots,2} = (2n-1)(2n-3)\cdots 3\cdot 1 \tag{24}$$

such sets in total and each such indicator variable is 1 with probability p^n, as we force all n edges in S to be present. Evidently, $X_n = \sum_S Y_S$ and we therefore have that

$$\mathbb{E}[X_n] = p^n(2n-1)(2n-3)\cdots 3\cdot 1.$$

We first note[13] that $\mathbb{E}[X_n] < p^n 2^n n!$. By using Stirling's approximation, we have that

$$2^n n! < 2^n \cdot \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1+O(1/n)),$$

and may conclude that

$$\mathbb{E}[X_n] < \sqrt{2\pi n}\left(\frac{2pn}{e}\right)^n (1+O(1/n)).$$

Let us now put $p = 1/n$. This gives

$$\mathbb{E}[X_n] < \sqrt{2\pi n}\left(\frac{2}{e}\right)^n (1+O(1/n))$$

and this clearly approaches 0 as $n \to \infty$. Hence, perfect matchings require at least $p \gg 1/n$. A similar calculation shows that $\mathbb{E}[X_n] > 0$ whenever $p \geq \log(n)/n$, so one natural[14] guess is that $C\log(n)/n$ is a threshold, for some C.

In order to find the threshold function, we shall only consider bipartite matchings, where we first partition G into two sets of vertices, A and B, both of size n, and we only consider the edges between these sets.

We know from Hall's theorem[15] that for k vertices in A, we need at least k different neighbors in B.

We shall consider two cases, where we first consider $k \leq n/2$. Let $U \subset A$ and $W \subset B$, with $|U| = k$ and $|W| = k-1$. Hall's condition is *not* satisfied if all neighbors of U are in W. Hence, all edges between U and $B\setminus W$ must be non-edges. This situation happens with probability[16]

$$(1-p)^{k(n-k+1)} \leq e^{-pkn/2}.$$

There are $\binom{n}{k}\binom{n}{k-1}$ such choices of U and W, so the probability of Hall's condition to fail is less than

$$\binom{n}{k}^2 e^{-pkn/2} \leq n^{2k}e^{-pkn/2} = e^{-pkn/2+2k\log(n)}. \tag{25}$$

If $p > 4\log(n)/n$, we have that (25) approaches 0 as $n \to \infty$.

[13] $$(2n)(2n-2)\cdots 4\cdot 2 = 2^n n!$$

[14] Actually, this is in fact not very natural, but it is the correct one.

[15] We need for every $U \subseteq A$ that $|N(A)| \geq |A|$, and same on the other side.

[16] We have k edges in U which are not allowed to match with the $n-k+1$ edges in $B\setminus W$

For the second case when $k \geq n/2$, we again have sets $U \subseteq A$ and $W \subset B$ and we do not wish to have any edges between U and $B \setminus W$. The setup is the same as in previous case (just interchange the role of A and B). Hence, Hall's condition is violated with the same probability as before, so we can now conclude that if $p \gg \log(n)/n$ with probability approaching 1, Hall's condition is fulfilled. In other words, we almost surely have a perfect matching.

Note also that we must also have the obvious requirement that there are no isolated vertices. The threshold function for this is $p = \log(n)/n$, so we are ok. In conclusion, $C \log(n)/n$ is the threshold for existence of perfect matchings, for some appropriate value of C.

Warning: This is one of the rare instances where the calculations for $\mathbb{E}[X]$ does not give a clear hint about the threshold, since a few graphs might have a huge number of perfect matchings, making the expected number of perfect matchings to be positive, but the number of graphs with at least one perfect matching to be few.

Second moment method examples

Isolated vertices

Let $G \in \mathcal{G}(n,p)$. We shall show that G has isolated vertices whenever $p \ll \log(n)/n$.

First, we let Y_i be the indicator variable that vertex i is isolated. This means that $Y_i = q^{n-1}$, where $q := 1-p$. We let X be the number of isolated vertices, so that

$$X = Y_1 + Y_2 + \cdots + Y_n.$$

The expected number of isolated vertices is then

$$\mathbb{E}[X] = nq^{n-1}.$$

Furthermore[17],

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_i \mathbb{E}[Y_i^2] + 2\sum_{i<j} \mathbb{E}[Y_iY_j] \\ &= \mathbb{E}[X] + 2\sum_{i<j} q^{2n-1}.\end{aligned}$$

[17] We have $\mathbb{E}[Y_iY_j] = q^{2n-1}$.

Hence,

$$\mathbb{E}[X^2] < nq^{n-1} + n^2q^{2n-1}.$$

The second moment method tells us that

$$\mathbb{P}[X \geq 1] > \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]},$$

so

$$\mathbb{P}[X \geq 1] > \frac{n^2q^{2n-2}}{nq^{n-1} + n^2q^{2n-1}}. \tag{26}$$

Now, some algebra gives that

$$\frac{n^2q^{2n-2}}{nq^{n-1} + n^2q^{2n-1}} = \frac{q^{n-1}}{\frac{1}{n} + q^n} > \frac{q^n}{\frac{1}{n} + q^n} = \frac{1}{\frac{q^{-n}}{n} + 1}.$$

Finally, we have that with $p = \frac{\log(n)}{n}$,

$$\frac{(1-p)^{-n}}{n} = \frac{\left(1 - \frac{\log(n)}{n}\right)^{-n}}{n} \to 1$$

as $n \to \infty$, and it is then quite clear that if $p \ll \frac{\log(n)}{n}$, this limit is 0. This implies that the expression in the right hand side of (26) approaches 1 as $n \to \infty$, and we almost surely have an isolated vertex.

Triangles

Let P be the property G *has a triangle*. Show that $\tau(n) = n^{-1}$ is a threshold function. We let X_n to be the expected number of triangles of a $G \in \mathcal{G}(n,p)$. For each subset $S \subset \{1,2,\ldots,n\}$, of size 3, we let I_S be the indicator variable which is 1 if the vertices given by S form a triangle. We then have that $\mathbb{E}[I_S] = p^3$ for all such S. By definition,

We let $\binom{[n]}{3}$ denote the set of subsets of $\{1, \dots, n\}$ of size 3.

$$X_n = \sum_{S \subset \binom{[n]}{3}} I_S \quad \Longrightarrow \quad \mathbb{E}[X_n] = \sum_{S \subset \binom{[n]}{3}} \mathbb{E}[I_S] = \binom{n}{3} p^3.$$

Note that the random variables $\{I_S\}$ are not independent!

We now use the first moment method when considering the case $p_n \ll n^{-1}$. As $n \to \infty$, we have that

We write $f(n) \in \Theta(n^k)$ if $\lim_{n \to \infty} f(n)/n^k = 0$.

$$\mathbb{E}[X_n] = \binom{n}{3} p^3 \sim n^3 p_n^3 \in \Theta(1).$$

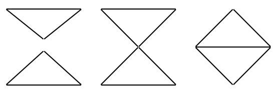
Hence, $\mathbb{E}[X_n] \to 0$ and by Markov's inequality (4) we can deduce that $\mathbb{P}[X_n > 0] \to 0$. In other words, if $p_n < \tau(n)$, then $G \in \mathcal{G}(n, p_n)$ is *triangle-free* almost always as $n \to \infty$.

The second moment method must be used to prove the second case. We first compute $\mathbb{E}[X_n^2]$. By using the indicator variables,

$$\mathbb{E}[X_n^2] = \mathbb{E}\left[\left(\sum_{S \subset \binom{[n]}{3}} I_S\right)^2\right] = \sum_{S \subset \binom{[n]}{3}} \mathbb{E}[I_S^2] + \sum_{\substack{S,T \subset \binom{[n]}{3} \\ S \neq T}} \mathbb{E}[I_S I_T].$$

We have that $\mathbb{E}[I_S^2] = \mathbb{E}[I_S]$, so the first sum is easy. The second sum can be partitioned into cases, depending on the size of $S \cap T$. We have that

$$\sum_{\substack{S,T \subset \binom{[n]}{3} \\ |S \cap T| = 0}} \mathbb{E}[I_S I_T] + \sum_{\substack{S,T \subset \binom{[n]}{3} \\ |S \cap T| = 1}} \mathbb{E}[I_S I_T] + \sum_{\substack{S,T \subset \binom{[n]}{3} \\ |S \cap T| = 2}} \mathbb{E}[I_S I_T]$$



(I) Pick 6 vertices, then choose which 3 is S.
(II) Pick 5 vertices, choose which 3 is S, then which of these is common with T.
(III) Pick 4 vertices, choose which 3 is S, then an edge of S common with T.

In each case, we know exactly how many edges which are involved. Hence, the above sum is precisely

$$\binom{n}{6}\binom{6}{3} p^6 + \binom{n}{5}\binom{5}{2} 3p^6 + \binom{n}{4}\binom{4}{3} 3p^5.$$

To conclude, for large n, we have

$$\mathbb{E}[X_n^2] \sim n^3 p^3 + n^6 p^6 + n^5 p^6 + n^4 p^5 \tag{27}$$
$$\mathbb{E}[X_n]^2 \sim n^6 p^6 \tag{28}$$
$$\mathrm{Var}[X_n] \sim n^3 p^3 + n^5 p^6 + n^4 p^5. \tag{29}$$

The second moment method (7) now gives that for large n,

$$\mathbb{P}(X_n = 0) \leq \frac{\mathrm{Var}[X_n]}{\mathbb{E}[X_n^2]} \sim \frac{n^3 p^3 + n^5 p^6 + n^4 p^5}{n^3 p^3 + n^6 p^6 + n^5 p^6 + n^4 p^5}.$$

Multiplying numerator and denominator with $(np)^{-6}$, we get

$$\frac{(np)^{-3} + n^{-1} + p(np)^{-2}}{1 + (np)^{-3} + n^{-1} + p(np)^{-2}}.$$

If now $p_n \gg n^{-1}$, we have that $(np) \to \infty$, so the numerator above approaches 0, while the denominator approaches 1. We can conclude that $\mathbb{P}(X_n = 0) \to 0$ and consequently, $X_n \geq 1$ almost always[18]. Hence, we almost always have a triangle.

[18] Since X_n is a discrete variable.

The second moment method guarantees that most graphs has at least one triangle, and it is not the case that there are only a very small number of graphs with a huge number of triangles.

Paths on a cycle

LET G BE A RANDOM SUBGRAPH of the cycle graph with n vertices, where each edge appear independently with probability $p \in (0,1)$. Let X denote the random variable counting the total number of paths in G. As a convention, if all n edges are present, there is no path.

(a) Find a formula for $\mu := \mathbb{E}[X]$.

(b) Compute $\mathbb{E}[X^2]$.

(c) Find all $\alpha > 0$ so that $\mathbb{P}\left[\, |X-\mu| \geq n^{-\alpha}\mu \,\right] \to 0$ as $n \to \infty$.

We let $Y_{r,k}$ be the indicator variable that starting from vertex r, there is a path of length exactly k clockwise. We have that

$$\mathbb{E}[Y_{r,k}] = \begin{cases} p^k(1-p)^2 & \text{if } k < n-1 \\ p^k(1-p) & \text{if } k = n-1 \\ 0 & \text{if } k = n \end{cases}$$

since in the first case, every edge in the path must appear, and the edge before and after are not allowed to be present. For $k = n-1$, there is exactly one path consisting of $n-1$ edges. We let X be the random variable counting total number of paths, so that $X = \sum_{r,k} Y_{r,k}$. Simple calculations give

$$\begin{aligned} \mathbb{E}[X] &= n(p + p^2 + \cdots + p^{n-2})(1-p)^2 + np^{n-1}(1-p) \\ &= np(1-p). \end{aligned}$$

WE WOULD NOW LIKE TO DO second moment analysis, so we need to compute $\mathbb{E}[X_n^2]$. In order to do this, we need the expected value of products of indicator variables. Note that if the corresponding paths overlap or are directly adjacent, the expected value of the product is 0. The only products with non-zero expectation are the following cases. In all cases, we need that $k + m \leq n-2$, or the paths would overlap.

In the first case, the paths are far apart and

$$\mathbb{E}[Y_{r,k} \cdot Y_{s,m}] = p^{k+m}(1-p)^4.$$

There are $n\binom{n-3}{3}$ such configurations[19] in total. Furthermore, for $2 \leq t \leq n-4$, there are[20]

$$n(n-3-t)(t-1)$$

configurations where the sum of the length of the paths is t. The total contribution of this case to $\mathbb{E}[X_n^2]$ is therefore

$$\sum_{t=2}^{n-4} n(n-3-t)(t-1)(1-p)^4 p^t,$$

[19] Pick three points, p_1, p_2, p_3 on the cycle, except vertices 1, $n-1$ and n. This gives $n-3$ points. The vertices $1, \ldots, p_1$, followed by $p_2+1, \ldots, p_3+1$ are now two such paths we seek.

[20] Verify that $\sum_{t=2}^{n-4} n(n-3-t)(t-1)$ is indeed $n\binom{n-3}{3}$.

which can be computed as

$$\frac{n(1-p)\left(np^n - np^{n+1} + 5p^{n+1} - 5p^4 + 3p^5 - 3p^n + np^4 - np^5\right)}{p^2}. \tag{30}$$

In the second case, the paths are adjacent[21] and $k+m < n-2$. We have

$$\mathbb{E}[Y_{r,k} \cdot Y_{s,m}] = p^{k+m}(1-p)^3,$$

and there are exactly[22] $n\binom{n-3}{2}$ such configurations, and $n(t-1)$ configurations where total path length is t, $2 \leq t \leq n-3$. This contributes with [23]

$$2\sum_{t=2}^{n-3} n(t-1)(1-p)^3 p^t,$$

which simplifies to

$$\frac{2n(1-p)\left(p^4 + 3p^n - np^n - 4p^{1+n} + np^{1+n}\right)}{p^2}.$$

The last case, there are exactly two paths which are adjacent on both sides $k+m = n-2$,

$$\mathbb{E}[Y_{r,k} \cdot Y_{s,m}] = p^{k+m}(1-p)^2.$$

There are $n(n-3)$ such configurations (pick a point on the circle, where the first path starts. There are $n-3$ remaining points where the second path may start so that it has length at least one. These choices uniquely determine the two paths, and we get

$$n(n-3)p^{n-2}(1-p)^2.$$

Note also that we have $\mathbb{E}[Y_{r,k}^2] = \mathbb{E}[Y_{r,k}]$ which also need to be accounted for. Adding all expressions together, we conclude that $\mathbb{E}[X_n^2]$ is equal to

$$\mathbb{E}[X_n^2] = np(1-p)\left(1 + (n-3)p(1-p)\right).$$

Note that this is symmetric in p and $1-p$, which we could expect. We can now compute the variance,

$$\begin{aligned}\mathrm{Var}(X) &= np(1-p)\left(1 + (n-3)p(1-p)\right) - n^2p^2(1-p)^2 \\ &= np(1-p)(1-3p(1-p)).\end{aligned}$$

We now apply Chebyshev's inequality with $\lambda = n^{1/2-\epsilon}$ for some $\epsilon > 0$. We then have

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \frac{\mathbb{E}[X]}{n^{1/2-\epsilon}}\right] < n\frac{\mathrm{Var}(X)}{n^{2\epsilon}\mathbb{E}[X]^2}.$$

Taking $n \to \infty$ will show that the right hand side converges to 0, so we see that all $\alpha < 1/2$ fulfill the requirements.

[21] $k+2=s$

[22] Choose starting point of first interval, and count pairs (m,k) of positive integers such that $k+m \leq n-1$.

[23] We have factor 2 from the expansion of the square.

AS AN ALTERNATIVE SOLUTION, we can see that $X = Z/2$ is half the number of vertices with degree exactly equal to 1. Let V_j be the indicator variable that vertex j has degree 1. There are two possible ways for a vertex to have degree exactly equal to one, so $\mathbb{E}[V_j] = 2p(1-p)$. It follows that $\mathbb{E}[Z] = 2np(1-p)$. Furthermore,

$$\mathbb{E}[Z^2] = \mathbb{E}[Z] + \sum_{i \neq j} \mathbb{E}[V_i V_j].$$

We now note that

$$\mathbb{E}[V_i V_j] = \begin{cases} p(1-p)^2 + p^2(1-p) = p(1-p) & \text{if } |i-j| = 1 \\ 4p^2(1-p)^2 & \text{otherwise.} \end{cases}$$

There are $2n$ terms of the first type, and $n^2 - 3n$ terms of the second type. We can therefore conclude that

$$\begin{aligned} \mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \frac{\mathbb{E}[Z^2] - \mathbb{E}[Z]^2}{4} \\ &= \frac{2np(1-p) + 2np(1-p) + (n^2-3n)4p^2(1-p)^2 - 4n^2p^2(1-p)^2}{4}. \end{aligned}$$

This simplifies to $np(1-p)(1-3p+3p^2)$ as we saw before.

WE SHOULD ALSO SHOW THAT for $\alpha > 1/2$ the expression

$$\mathbb{P}\left[\ |X - \mu| \ \geq \ n^{-\alpha}\mu\ \right]$$

does not converge to 0. It suffices to show[24] that the expression

$$\mathbb{P}\left[\ X \ \geq \ \mu(1 + n^{-\alpha})\ \right] \geq \mathbb{P}\left[\ X \ \geq \ (n + \sqrt{n})/4\ \right]$$

does not go to 0.

[24] The inequality follows from the fact that $p(1-p) \leq 1/4$ and $\mu = p(1-p)$

Expected clique size

We now analyze the clique size of a random graph.

Theorem 2. *Let $G \in \mathcal{G}(n, 1/2)$. Then G has no clique*[25] *of size $2\log_2(n)$ almost surely as $n \to \infty$.*

[25] A *clique* of size k is a complete subgraph with k vertices.

Proof. If S is a subset of vertices, we let X_S be the indicator variable, such that $X_S = 1$ if S is a clique. There are $\binom{n}{k}$ possible cliques of size k in G. Furthermore, if $|S| = k$, then

$$\mathbb{P}[X_S = 1] = 2^{-k(k-1)/2}$$

as there are $k(k-1)/2$ edges in a k-clique. Let X be the number of cliques of size k. We then have

By Markov's inequality and basic inequalities for binomials.

$$\mathbb{P}[X > 0] \leq \mathbb{E}[X] = \binom{n}{k} 2^{-k(k-1)/2} < \frac{n^k}{k!} \frac{1}{2^{k(k-1)/2}}.$$

Some more algebra gives that

$$\mathbb{P}[X > 0] \leq \frac{1}{k!} \left(\frac{n}{2^{(k-1)/2}} \right)^k. \tag{31}$$

If we now set $k := 2\log_2(n) + 1$, we have that

$$\frac{n}{2^{(k-1)/2}} = \frac{n}{2^{2\log_2(n)/2}} = 1.$$

Hence, from (31) we finally have that

$$\mathbb{P}[X > 0] \leq \frac{1}{(2\log_2(n)+1)!} < \frac{1}{\log_2(n)}$$

and the right hand side approaches 0 as $n \to \infty$. □

On the other hand, by using the second moment method one can prove that it is almost surely true that for any $\epsilon > 0$, there is a clique of size at least $(2-\epsilon)\log_2(n)$ as n goes to infinity.

See `https://theory.epfl.ch/courses/topicstcs/Lecture1.pdf` and `https://www.win.tue.nl/~nikhil/courses/2015/2WO08/probabilistic_method-2.pdf`

Squares in the ladder graph

Let L_n be the graph with vertex set $\{1, 2, \dots, n\} \times \{0, 1\}$ in $\mathbb{Z}^2$, where vertices are connected by an edge if they are a unit distance apart. We let G_n be a random subgraph of L_n, where each edge appear with probability p. A square is a 4-cycle in G, and we let X be the random variable which counts the number of 2-cycles.

Find a threshold function for the existence of a square.

LET Y_i BE THE INDICATOR VARIABLE that the square with corners $(i, 0)$, $(i+1, 1)$ is present. There are $n-1$ such possible squares. Since a square has four edges, $\mathbb{P}(Y_i = 1) = p^4$. We then have that $\mathbb{E}[X] = \mathbb{E}[Y_1 + \cdots + Y_{n-1}] = (n-1)p^4$. We shall also compute $\mathbb{E}[X^2]$. We have that

$$\mathbb{E}[X^2] = (n-1)\mathbb{E}[Y_i^2] + 2\sum_{i=1}^{n-2} \mathbb{E}[Y_i Y_{i+1}] + (n-2)(n-3)\mathbb{E}[Y_i Y_j],$$

as there are $(n-1)^2$ terms in total, and there are exactly $2(n-2)$ terms involving adjacent squares. It follows that

$$\mathbb{E}[X^2] = (n-1)p^4 + 2(n-2)p^7 + (n-2)(n-3)p^8.$$

The second moment method gives that

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = \frac{(n-1)^2 p^8}{(n-1)p^4 + 2(n-2)p^7 + (n-2)(n-3)p^8}.$$

The right hand is, for large n, approximately equal to

$$\frac{np^4}{1 + 2p^3 + np^4}$$

which approaches $1/2$ if $p = n^{-1/4}$. Hence, we can easily prove that if $p \gg n^{-1/4}$, then

$$\mathbb{P}(X > 0) \to 1 \text{ as } n \to \infty,$$

but $\mathbb{E}[X] \to 0$ if $p \ll n^{-1/4}$.

Threshold function for squares in a grid

Let G_n be the $n \times n$-grid graph with n^2 vertices. A random subgraph is chosen from G_n where each edge appear independently with probability p. We call a 4-cycle in G_n a *square*. Show that $\tau(n) = 1/\sqrt{n}$ is a threshold function for the property G contains a square.

In G_n, there are $(n-1)^2$ squares. Let S_{ij} be the indicator random variable that square (i,j) is present. Clearly, $S_{ij} = 1$ with probability p^4 and 0 with probability $1 - p^4$. We let X_n be the random variable counting the number of squares in a random graph. Note then that

$X_n = \sum_{ij} S_{ij}$

$$\mathbb{E}(X_n) = \sum_{i,j} \mathbb{E}(S_{ij}) = (n-1)^2 p^4. \tag{32}$$

We shall now compute $\mathbb{E}(X_n^2)$. It is given by

$$\mathbb{E}(X_n^2) = \sum_{i,j,i',j'} \mathbb{E}(S_{ij} S_{i'j'})$$

which then gives

$$\mathbb{E}(X_n^2) = \sum_{i,j} \mathbb{E}(S_{ij}^2) + 4 \sum_{i<i',j<j'} \mathbb{E}(S_{ij} S_{i'j'}) \tag{33}$$

$$+ 2 \sum_{i,j<j'} \mathbb{E}(S_{ij} S_{ij'}) + 2 \sum_{i<i',j} \mathbb{E}(S_{ij} S_{i'j}). \tag{34}$$

The first sum is simply $\mathbb{E}(X_n) = (n-1)^2 p^4$, and the second has terms which are products of independent variables, so

$$\sum_{i<i',j<j'} \mathbb{E}(S_{ij} S_{i'j'}) = \sum_{i<i',j<j'} \mathbb{E}(S_{ij}) \mathbb{E}(S_{i'j'}) = \binom{n-1}{2}^2 p^8.$$

The last two sums in (34) be separated further (they are equal due to symmetry), so we only compute the first. We get

$$\sum_{i<i',j} \mathbb{E}(S_{ij} S_{i'j}) = \sum_{i<i'-1,j} \mathbb{E}(S_{ij}) \mathbb{E}(S_{i'j}) + \sum_{i=i'-1,j} \mathbb{E}(S_{ij} S_{i'j}) \tag{35}$$

$$= \binom{n-2}{2} (n-1) p^8 + (n-2)(n-1) p^7. \tag{36}$$

Note that in the last sum, we do indeed get p^7, as two adjacent squares have only 7 edges. Putting it all together, we have that

$$\mathbb{E}(X_n^2) = (n-1)^2 p^4 + 4 \binom{n-1}{2}^2 p^8 + 4 \binom{n-2}{2} (n-1) p^8 + 4(n-2)(n-1) p^7. \tag{37}$$

WE NOW USE THE FIRST MOMENT METHOD in the case $p_n \ll 1/\sqrt{n}$. We have

$$\mathbb{E}[X_n] \sim n^2 p^4 \in \Theta(1),$$

so by Markov's inequality, $\mathbb{P}[X_n > 0] \to 0$ and the graph is square-free almost surely as $n \to \infty$.

The second moment method in the case $p_n \gg 1/\sqrt{n}$. Using the computations from (32) and (37), $\frac{\mathbb{E}[X_n]^2}{\mathbb{E}[X_n^2]}$ is equal to

$$\frac{(n-1)^4 p^8}{(n-1)^2 p^4 + 4\binom{n-1}{2}^2 p^8 + 4\binom{n-2}{2}(n-1)p^8 + 4(n-2)(n-1)p^7}.$$

As we are interested in the limit $n \to \infty$, we can replace $n = n+1$. We also expand the binomial coefficients:

$$\frac{n^4 p^8}{n^2 p^4 + n^2(n-1)^2 p^8 + 2n(n-1)(n-2)p^8 + 4(n-2)(n-1)p^7}.$$

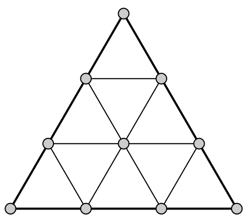
We can rewrite this as

$$\frac{1}{\left(\frac{n-1}{n}\right)^2 + \frac{1}{n^2p^4} + 2\frac{(n-1)(n-2)}{n^3} + 4\frac{(n-2)(n-1)}{n^4 p}},$$

and all terms in the denominator except the first one approach 0 as $n \to \infty$, assuming $p_n \gg 1/\sqrt{n}$. Thus, we can now conclude that $\mathbb{E}[X_n]^2/\mathbb{E}[X_n^2] \to 1$ and by the second moment method, $\mathbb{P}[X_n > 1]$ approaches 1. Hence, we can say that the random graph almost surely contains a square as $n \to \infty$.

Threshold function for triangular grid

Let T_n be the graph on $1+2+3+\cdots+n$ vertices, where the vertex set is given by $\{(i,j) \in \mathbb{Z}^2 : 1 \leq j \leq i \leq n\}$ and (i,j) is connected to (i',j') via an edge if and only if $|i'-i| \leq 1$ and $0 \leq j'-j \leq 1$. For $n=4$, the graph looks as follows:



(38)

A random subgraph G is chosen from T_n where each edge appear independently with probability p. Find the threshold function $\tau(n)$ for the property G *contains a* 3*-cycle.*

There are $1+3+5+\cdots+(2n-1) = n^2$ possible 3-cycles (triangles) in T_n. We let X_n be the random variable counting triangles and let S_k be the indicator variable for triangle k being present. Clearly, $\mathbb{E}[S_k] = p^3$ for each k, and $\mathbb{E}[X_n] = n^2p^3$.

Here it is reasonable to guess that $\tau(n) := n^{-2/3}$. We see that if $p_n \ll n^{-2/3}$, then $\mathbb{E}[X_n] \in \Theta(1)$ and by Markov's inequality, T_n is almost surely triangle-free.

In order to apply the second moment method, we must compute $\mathbb{E}[X_n^2]$. This is given by

$$\sum_k \mathbb{E}[S_k^2] + \sum_{k,k' \text{ adjacent}} \mathbb{E}[S_k S_{k'}] + \sum_{k,k' \text{ non-adjacent}} \mathbb{E}[S_k S_{k'}]. \tag{39}$$

Each pair of adjacent triangles is determined by an edge which is not adjacent to the unbounded region, see the figure in (38). There are $3(1+2+\cdots+(n-1)) = 3n(n-1)/2$ edges in total in T_n and thus $3n(n-1)/2 - 3(n-1) = 3(n-1)(n-2)/2$ possible pairs[26] of adjacent triangles.

[26] Unordered pairs

The expression in (39) is given by expanding $\mathbb{E}[(\sum_k S_k)^2]$, so it has $n^2 \cdot n^2 = n^4$ terms in total. The first sum in (39) has n^2 terms, the second sum has $3(n-1)(n-2)$ terms[27], and the last sum must therefore have

[27] Ordered pairs

$$n^4 - 3(n-1)(n-2) - n^2 = n^4 - 4n^2 + 9n - 6$$

terms. Hence,

$$\mathbb{E}[X_n^2] = n^2p^3 + 3(n-1)(n-2)p^5 + (n^4 - 4n^2 + 9n - 6)p^6.$$

Now, $\frac{\mathbb{E}[X_n]^2}{\mathbb{E}[X_n^2]}$ is given by

$$\frac{(n^2p^3)^2}{n^2p^3 + 3(n-1)(n-2)p^5 + (n^4 - 4n^2 + 9n - 6)p^6}.$$

We rewrite this as

$$\frac{1}{\frac{1}{n^2p^3} + \frac{3(n-2)(n-1)}{n^4p} + \frac{(n^3+n^2-3n+6)(n-1)}{n^4}}$$

and if $p_n \gg \tau(n) = n^{-2/3}$, then the first two fractions in the denominator approach 0, while the last term approaches 1, as n goes to infinity. Hence, $\frac{\mathbb{E}[X_n]^2}{\mathbb{E}[X_n^2]} \to 1$, and by the second moment method, we can be sure that if $p_n \gg n^{-2/3}$, then $\mathbb{P}[X_n > 1] \to 1$ as $n \to \infty$. In other words, there is a triangle almost surely in T_n.

Ramsey type questions

Maximal number of edges in graph without a path of length c

WHAT IS THE MAXIMAL NUMBER[28] of edges of a graph on n vertices that does not contain a path with c edges?

[28] See also the related Erdős–Sós conjecture.

We shall show below that $cn/2$ edges is an upper bound, meaning that the average degree is at most c.

For a lower bound, take m disjoint copies of the complete graph on c vertices. This has $n = mc$ vertices, $mc(c-1)/2$ edges and average degree $c-1$. Furthermore, there is no path of length c, so for sure $n(c-1)/2$ is a lower bound.

FOR THE UPPER BOUND, suppose that we are given a graph G on n vertices with $cn/2$ edges. Our goal is to show that G contains a path of length c. Note that G has average degree[29] c.

[29] Average degree is $2\frac{|E|}{|V|}$

From G we iteratively remove vertices whose degree is at most $c/2$. It can be checked[30] that this operation does not decrease the average degree, and we can be sure to end up with a non-empty graph where every vertex has degree at least $c/2$.

[30] If H has n vertices, with average degree $2k$ and nk edges, then removing one vertex with degree k results in a graph H' with $(n-1)$ vertices and $(n-1)k$ edges, thus preserving the average degree.

Lemma 1. *Suppose G is a graph where every vertex has degree at least d and average degree $2d$. Then*[31] *G contains a path of length $2d$.*

[31] Compare with Dirac's theorem, 10.1.1 in Diestel.

Proof. Since the average degree is $2d$, we can be sure that the graph has some connected component with at least $2d$ vertices.

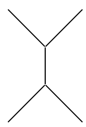
Consider now a longest path $P = (v_0, \dots, v_k)$ in such a component, where we have indexed the vertices along the path. Suppose now $k < 2d$. Since P is a longest path, all neighbors of v_0 must be among $v_1, \dots, v_k$. A similar observation holds for v_k. The Pigeonhole principle then tells us that there is an edge (v_i, v_{i+1}) such that both (v_0, v_{i+1}) and (v_i, v_k) are edges. This makes

$$v_0, v_{i+1}, v_{i+2}, \dots, v_k, v_i, v_{i-1}, \dots, v_1, v_0$$

into a cycle, and we could then construct a longer path P, as we know that the connected component contains at least $2d$ vertices. Hence, $k \geq 2d$ and we are done. □

From this lemma with $d = c/2$, we can conclude that every graph on n vertices and $cn/2$ edges contains a path with c edges.

NOTE THAT EXTREMAL GRAPHS in the sense that adding any edge creates a longer path, might *not* be cliques. [32]

[32] 

Longest path is 4, but adding any additional edge creates a path of length at least 5.

Trouble with tournaments

Let $D(m, k)$ be the smallest integer such that every tournament[33] on $D(m, k)$ vertices either has a totally ordered m-clique, or k edge-disjoint directed 3-cycles. Find good lower and upper bounds on $D(m, k)$.

[33] directed complete graph

First observation is that we clearly have that $D(m, k)$ is weakly increasing in m and k. Also, note that any induced subgraph H that does not contain a directed 3-cycle, is a totally ordered clique.

Claim: We have that

$$D(m, k) \leq km \text{ for } m \geq 3 \text{ and } k \geq 1. \tag{40}$$

Consider a directed complete graph on km vertices, and partition the vertices into k m-subsets. If such an m-subsets is a totally ordered m-clique, we are done. If it is not totally ordered, it must contain a directed 3-cycle. This proves the inequality.

It makes sense[34] to extend the definition and let $D(m, \infty)$ be the minimal number of vertices needed to guarantee that a tournament on this number of vertices contains a totally ordered m-clique.

[34] Note that $D(\infty, k) = \infty$ as there are arbitrary large tournaments without directed cycles, so it does not help much to introduce this extension.

Claim: We have that

$$D(m, \infty) \leq 2^{m-1}. \tag{41}$$

We do proof by induction. The cases $m = 1$ and $m = 2$ are clear. Now suppose we have a tournament on 2^m vertices. Take any vertex v — either in-degree or out-degree is at least 2^{m-1}. In the first[35]

[35] The second case is treated similarly.

case, let H be the subgraph induced by all vertices with an edge directed to v. As H has at least 2^{m-1} vertices, it has by induction a totally ordered m-clique, and together with v, this makes a totally ordered $(m+1)$-clique.

It follows that for fixed m, the sequence

$$D(m,1) \leq D(m,2) \leq D(m,3) \leq \cdots$$

eventually stabilizes.

Claim: We have that

$$D(m,k) \leq m + 2(k-1). \tag{42}$$

Suppose G has $m+2(k-1)$ vertices, but not k edge-disjoint 3-cycles. Let S be a maximal edge-disjoint set of 3-cycles — there are at most $k-1$ of these. In each such cycle, remove two of the vertices and the adjacent edges. This completely remove all edges of the 3-cycles in S. This creates a subgraph H, on (at least) m vertices. Now, H cannot contain a 3-cycle, as that would violate the maximality of S. But if H does not contain a 3-cycle, it must be totally ordered.

See also https://hal.archives-ouvertes.fr/hal-02277436/document, https://drops.dagstuhl.de/opus/volltexte/2019/10971/pdf/LIPIcs-MFCS-2019-27.pdf for complexity results

Claim: We[36] have that

$$D(m,k+1) \leq D(m,k) + 2. \tag{43}$$

[36] Argument due to L. Heldin

Let G be a tournament on $D(m,k)+2$ vertices. We need to show that G has either an m-clique, or $k+1$ directed 3-cycles. If G has an m-clique, we are done, so suppose it does not. Then G has at least a 3-cycle, C. Remove two of its vertices and adjacent edges. This leaves $D(m,k)$ vertices, which do not have an m-clique, so it must have k edge-disjoint 3-cycles. These are also edge-disjoint from C, so the original G does indeed have either an m-clique or $(k+1)$ edge-disjoint 3-cycles. Note how this inequality is very similar to (42).

FOR LOWER BOUNDS, we need to invent explicit constructions of extremal graphs not containing m-cliques nor too many 3-cycles.

Claim: We[37] have that

$$D(m,k) \geq \begin{cases} m+k-1 & \text{if } m \geq 2k-1 \\ \lfloor m/2 \rfloor & \text{if } m < 2k-1. \end{cases} \tag{44}$$

[37] Argument due to J. Conneryd, K. Ehringer, G. Nilsson, I. Ohlsson-Ågnell

The first case, $m \geq 2k-1$ is constructed as follows. Start with a disjoint union of $(k-1)$ directed 3-cycles, and $m-2k+1$ isolated vertices. We order these components such that the isolated vertices are last, and then direct the remaining edges accordingly. The pigeonhole principle easily shows that any m vertices must contain at least one of the 3-cycles, so it will not be totally ordered. Also, it is fairly easy to verify that there are only $k-1$ edge-disjoint 3-cycles, so this construction gives a lower bound.

For the second case, when $m < 2k-1$, the same construction works where we now instead use $\lfloor (m-1)/2 \rfloor$ 3-cycles, and one isolated vertex.

Claim: We[38] have that

$$D(m,k) \geq \min(2m-3, m+k-2). \tag{45}$$

A graph on $1, 2, \dots, n$ is constructed, so that for $x < y$, the edge (x, y) is directed towards x if $y - x = m - 1$, and towards y otherwise. If $n \leq 2m-3$, one can show[39] that there is no totally ordered m-clique. Furthermore, if $n \leq m+k-2$, there are only $k-1$ edges directed from larger to smaller label, so there cannot be a set of k edge-disjoint directed 3-cycles.

[38] Argument due to A. Becedas

[39] Any m-clique has to vertex labels a, b with same remainder mod m.