CO749 Final Project: Survey of The Longest Path in a Random Graph

Zouhaier Ferchiou and Harry Sivasubramaniam

April 15, 2020

1 Introduction

We survey the results of Ajtai, Komolós and Szemerédi [1] on the longest path in a random graph. An easy consequence of Stirling's approximation is that

$$\binom{n}{k}k!\left(\frac{\alpha}{n}\right)^{k-1} \sim n\alpha^{k-1}e^{-k^2/2n}.$$

Using this and the first moment method, one can show that a.a.s. the longest path in a random graph with $(1-\epsilon)\frac{n}{2}$ edges (and in the directed case with $(1-\epsilon)n$ edges) is of length $O(\log n)$ for $\epsilon > 0$. When $\epsilon = 0$ we get that the longest path in both directed and undirected graphs is of length $O(\sqrt{n \log n})$. The main result of their paper was that for some constant c, a random graph with $(1+\epsilon)\frac{n}{2}$ edges contains a path of length cn and analogously, the random directed graph with $(1+\epsilon)n$ edges contains a directed path of length cn with probability near 1. One may have noticed that we have not made clear which model of random graphs these results hold for. In the original paper, the result is first proven for Erdős-Renyi graphs $\mathcal{G}_{n,p}$ and the directed analogue $\mathcal{D}_{n,p}$. By a standard coupling argument they extend it to the uniform random graph models $\mathcal{G}'_{n,N}$ and $\mathcal{D}'_{n,N}$ where the number of edges N is fixed. In this survey we will focus on the Erdős-Renyi models and omit the details of this coupling. The proof of the main result relies on being able to construct $\mathcal{D}_{n,p}$ in a depth first search (DFS) like manner. DFS is promising as it is likely to discover a linear path as opposed to breadth first search like approaches which would likely construct the whole graph before it exceed depth $\Omega(\log n)$. The above graph process is analyzed by relating it to a Galton-Watson branching process and analyzing that instead. The proof of the undirected case is a consequence of the directed. The result is obtained by constructing a dense directed graph on a a collection of disjoint paths of the original undirected graph, and connecting them by finding a long directed path in the auxilary graph.

2 Preliminaries

In the $\mathcal{G}_{n,p}$ model, the number of vertices n is fixed and each edge is sampled independently with probability p. $\mathcal{D}_{n,p}$ is defined similarly except we now have directions (i.e. edge uv is different from edge vu) so the potential number of edges is $n(n-1)$ while a $\mathcal{G}_{n,p}$ could have at most $\binom{n}{2}$.

2.1 Galton-Watson Branching Process

An intuitive description of this process is to consider the birth and death of particles. At time $t = 0$, a single particle is born. At time $t \geq 1$, each existing particles gives birth to some random number of children and then dies. To formalize this, let Z be a random variable on non-negative integers. Let $X_0 = 1$ and for $t \geq 1$, let X_t denotes the random number of children born at step t where each particle independently has Z children. Clearly $X_0 = 1$ and X_1 has the same distribution as Z. By a simple induction on t, one can show that $\mathbb{E}\, X_t = (\mathbb{E}\, Z)^t$ for $t \geq 1$. Let $\mathcal{E}$ denote the event that $X_t = 0$ for some $t > 0$, i.e., the event the process goes extinct, and let $Q = \Pr[\mathcal{E}]$.

Theorem 1. *If* $\mathbb{E}\, Z > 1$ *and* $\Pr[Z = 0] > 0$ *then* $0 < Q < 1$.

Proof. Let ρ_j be the probability the process becomes extinct after at most j steps and note that $\rho_j = \sum_{i=0}^{\infty} \Pr[Z = i]\rho_{j-1}^i$. Consider the probability generating function $G(x) = \mathbb{E}\, x^Z = \sum_{i=0}^{\infty} \Pr[Z = i]x^i$. It is easy to see that

- $G(1) = 1$ and $G(0) > 0$ as $G(0) = \Pr[Z = 0] > 0$
- G is convex on $(0, 1)$ since $G''(x) > 0$ for all $x \geq 0$ as $\Pr[Z = 0] > 0$
- $G'(1) = \sum_{i=1}^{\infty} ix^{i-1} \Pr[Z = i]\Big|_{x=1} = \mathbb{E}\, Z > 1$.

From the above properties, it is clear the function $G(x) - x$ has a unique positive root in $(0, 1)$. We will show that Q is this root. Observe that since $\rho_n \to Q$ and $\rho_n = G(\rho_{n-1})$, by continuity of G we get $Q = G(Q)$ as desired. $\square$

3 Directed Case

In what follows, fix the edge sampling probability $p = \frac{\alpha}{n}$ where $\alpha > 1$ is a constant. We now present the main theorem for the directed case.

Theorem 2. *There are positive numbers* $c = c(\alpha)$, $K = K(\alpha)$ *and* $\theta = \theta(\alpha) < 1$ *such that*

$$\Pr(D_{n,p} \text{ contains a directed path of length } cn) \geq 1 - K\theta^n.$$

3.1 Directed Graph Process

To prove the above, we will consider the following graph process which generates $\mathcal{D}_{n,p}$ as well as the desired long path. Fix an ordering $v_1, ..., v_n$ of the vertices, let $m = (1 - \delta)n$ for some $\delta = \delta(\alpha) > 0$ chosen later. We can think of the remaining δn vertices as the surplus.

1. Set v_1 as root vertex and initialize m element set $M = \{v_2, ..., v_{m+1}\}$
2. Determine the random number **Ch** of children of v_1 in M according to
$$\lambda_k = \Pr(\mathbf{Ch} = k) = \binom{m}{k} p^k (1 - p)^{m-k}$$
3. Select k points at random from M for the children of v_1 and then replace them in M with the first k unused vertices
4. Recursively repeat on the first child until branch dies, then repeat on the sibling of the node that went extinct and so on (Essentially a depth first search!)
5. If tree dies out, pick unused vertex with smallest index and restart the process until no vertices remain unused

The reader may have noticed the above process has yet to construct $\mathcal{D}_{n,p}$. Rather, it produces a disjoint collection of directed trees. To guarantee a $\mathcal{D}_{n,p}$ we appropriately randomize the remaining edges (i.e., edges between children, from one tree to another, from child to ancestor, etc). Note that adding in these additional edges can only increase the length of the longest path.

Remark 1. *The reason we sample from M is so that the distribution of the children does not change from generation to generation in the corresponding branching process. Without such restriction, the probability that v_1 has k children is $\binom{n-1}{k}p^k(1-p)^{n-1-k}$ while the probability its child has l children is $\binom{n-1-k}{l}p^l(1-p)^{n-1-k-l}$.*

Let $\lambda = mp = (1-\delta)\alpha$ where we may later choose δ small so that $\lambda > 1$. Consider branching process with branching distribution

$$\lambda_k = \Pr(\mathbf{Ch} = k) = \binom{m}{k}p^k(1-p)^{m-k} \sim \frac{(mp)^k}{k!}e^{-mp} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

At time $t = 0$ a single particle is born and gives birth to **Ch** many children and then dies. At each $t \geq 1$ each existing particle independently gives birth to **Ch** children and dies. At each level assign a random order to the branches. Such ordering allows us to talk about the leftmost and rightmost child of a point. Thus the notion of a "leftmost infinite path" (**LIP**) is well defined (The **LIP** will be used later to find our long path). We may now analyze the original graph process by instead studying a collection of branching processes. We can couple the two processes so that when the branching process goes extinct, the graph process also dies. If there are remaining unexplored vertices then we can start another branching process until all the vertices in the original graph have been accounted for. The branching process is an infinite process while the graph only has finitely many vertices. However, we can show that with high probability a **LIP** exists and that we will find a long subpath of the **LIP** well before we run out of points in the original graph.

3.2 Analysis of Branching Process

Let T denote the total population until extinction conditioned on the branching process dying. Let $T_1, T_2, ...$ be a sequence of independent random variables distributed according to T and define $S_k = T_1 + .. + T_k$.

Claim 1. $\Pr(S_{\epsilon n} > \frac{\delta}{2}n)$ *is exponentially small, where $\epsilon = \epsilon(\alpha, \delta)$.*

Claim 1 essentially says that the probability of using up more than half the surplus δn points in the original graph is very low when we run ϵn branching process. Let L_k denote the number of points to the left of **LIP** (including those on the path) up to level k. The next claim tells us we find the desired long path well before we use up more than half the surplus points.

Claim 2. *There exists $c = c(\alpha) > 0$ such that* $\Pr(L_{cn} > \frac{\delta}{2}n)$ *is exponentially small.*

The below claim basically says that the extinction probability is not too dependent on m which is a function of n. This is just a technical detail to show that $Q^{\epsilon n}$ is small for large n.

Claim 3. *For all m, the extinction probability $Q = Q(\lambda, m) < Q_0 = Q_0(\lambda) < 1$.*

With the above claims we are now ready to prove the main theorem. The proofs of the claims can be found in the next section.

Proof of Theorem 2. By Claim 3 and independence of T_i, the first ϵn branching processes go extinct with probability at most $Q_0^{\epsilon n}$. Combining this with claim 1 and 2 by a union bound, we get that with the desired exponential rate on the probability the following holds. There is an index $j \le \epsilon n$ such that $T_1, ..., T_{j-1}$ are finite and T_j in infinite and at most $\frac{\delta}{2}n$ vertices were used in our corresponding graph process up till this point. Also, at most $\frac{\delta}{2}n$ vertices of the original graph were used by taking the points corresponding to the left of **LIP** (inclusive) at level cn. The subpath of **LIP** up till depth cn corresponds to our desired long path in the original graph. Since we had a surplus of $n - m = \delta n$ vertices in the original graph process, where m is the size of the set M we sample from. We have not used up all the surplus so the subpath is in fact a valid path in our original graph. □

Remark 2. *The $1 - K\theta^n$ bound, can be obtained by carefully going through the claims and keeping track of the constants.*

3.3 Proof of Claims

Proof of Claim 3. Let $f(x) = \sum_{k\ge 0} \lambda_k x^k$ be the probability generating function for **Ch**. Since $\mathbb{E}[\mathbf{Ch}] = \lambda > 1$, by the same arguments used to prove theorem 1, we know that Q is the unique root of the equation $Q = f(Q)$ for $0 \le Q < 1$. By the binomial theorem,

$$f(x) = (1 - p + px)^m = (1 - \frac{\lambda(1-x)}{m})^m.$$

$f(x)$ is monotone increasing in m since for $C = \lambda(1-x)$ by AM-GM inequality

$$1 \cdot \left(1 - \frac{C}{m}\right) ... \left(1 - \frac{C}{m}\right) \le \left(\frac{1 + m(1 - \frac{C}{m})}{m+1}\right)^{m+1} = \left(1 - \frac{C}{m+1}\right)^{m+1}.$$

Since $f(x) \le e^{-\lambda(1-x)}$ by the above properties, $Q < Q_0 < 1$ for all m where Q_0 is the unique solution on $(0,1)$ to $Q_0 = e^{-\lambda(1-Q_0)}$. Equivalently, $Q_0 = \frac{x}{\lambda}$ where x is unique value in $(0,1)$ for which $xe^{-x} = \lambda e^{-\lambda}$. This is a well know fact and follows from some simple calculus. Let $h(x) = xe^{-x} - \lambda e^{-\lambda}$ and note that $h'(x) = (1-x)e^{-x} > 0$ on $[0,1)$. Since $h(0) < 0$ and $h(1^-) = e^{-1}(1 - \lambda e^{-(\lambda-1)}) > 0$ and h is increasing on this interval, it follows that there is a unique solution on $(0,1)$. □

Proof of Claim 1. We want to show that for the sum of independent T_i distributed according to T, $\Pr(T_1 + ... + T_{\epsilon n} > \frac{\delta}{2}n)$ has an exponentially small tail. If the moment generating function function $\mathbb{E}\, e^{t_0 T} \le K$, where $K > 0$ depends only on λ and not m, then we may use the Chernoff method to derive such tail. We will first derive the tail and then bound the MGF.

Let $t > 0$ and apply Markov's inequality to get that

$$\Pr(S_{\epsilon n} > \frac{\delta}{2}n) = \Pr[\exp(tS_{\epsilon n}) > \exp(t\delta n/2)] \le \frac{\mathbb{E}[\exp(tS_{\epsilon n})]}{\exp(t\delta n/2)}.$$

By independence of T_i,

$$\mathbb{E}[\exp(tS_{\epsilon n})] = \prod_{i=1}^{\epsilon n} \mathbb{E}\exp(tT) \le K^{\epsilon n},$$

by taking $t = t_0$ (t_0 will be chosen later). Putting this together gives us $\Pr[S_{\epsilon n} > \frac{\delta}{2}n] \le K^{\epsilon n} e^{-t_0 \delta n/2}$ and by picking ϵ small (in terms of α, δ), we can get the desired concentration bound.

It remains to show that $\mathbb{E}[e^{t_0 T}] \leq K$ for some $t_0 = t_0(\lambda) > 0$ and $K = K(\lambda) > 0$. Let Q and f be defined as it was in the proof of claim 3. Let $\overline{T}$ be the total population and $\overline{t}_l = \Pr[\overline{T} = l]$, then we have $\overline{t}_0 = 0$ and $\overline{t}_\infty = 1 - Q$, and for $l \geq 1$

$$\overline{t}_l = \sum_{k=0} \lambda_k \sum_{l_1+...+l_k=l-1:l_i\geq 1} \overline{t}_{l_1} \cdots \overline{t}_{l_k}.$$

Multiplying by x^l and summing for $l = 1, 2, ...$ we get that $\overline{g}(x) = \sum_{l=1} \overline{t}_l x^l$ satisfies $\overline{g}(x) = xf(\overline{g}(x))$. By the Lagrange Implicit Function theorem (a well known result in combinatorial enumeration), $\overline{g}(x)$ is defined uniquely for $x < 1$ and so $\overline{g}(x) < Q$. By continuity $\overline{g}(x)$ can be defined up to $y_0 = \frac{x_0}{f(x_0)} \sim \frac{e^{\lambda-1}}{\lambda}$ where $x_0 = \frac{1-p}{(m-1)p}$ is the maximum of $\frac{x}{f(x)}$ under the assumption $\overline{g} < x_0$ Let T be the total population conditioned on the process going extinct. Thus

$$t_l = \Pr[T = l] = \Pr[\overline{T} = l | \overline{T} < \infty] = \frac{\overline{t}_l}{Q}$$

and $\mathbb{E}T = g'(1) = \frac{1}{1-f'(Q)} \sim \frac{1}{1-\lambda Q}$ where $g(x) = \sum_{l=0} t_l x^l = \overline{g}(x)/Q$.

Since $y_0 \sim \frac{e^{\lambda-1}}{\lambda}$, we pick t_0 to be the value that makes $e^{t_0} = y_0$ and so $\mathbb{E}e^{t_0 T} = g(e^{t_0}) = g(y_0) = \frac{x_0}{Q} < \frac{1}{Q}$. Note that we are done as Q is an increasing function in m so $\frac{1}{Q}$ is bounded uniformly in m as needed. $\square$

The proof of the following claim is quite similar to that of claim 1 so we just give a sketch proof.

Proof (sketch) of Claim 2. Let T_L be the number of points in the union of all trees on the left of **LIP** which start from the root and do not contain any edge from **LIP**. Clearly L_k is the sum of k independent copies of T_L. As seen above, it suffices to get a bound on the MGF of T_L that does not rely on m as we can then use Chernoff's method to get the desired concentration bound. Let $h(x)$ be the generating function of T_L and $K < \infty$ depending only on λ such that $h(a) < K$. Let the distribution of T_L be $(h_1, h_2, ...)$, then $h(x) = \sum_{i=0} h_i x^i$ and by a similar argument to before we get that for $l \geq 1$

$$h_l = \sum_{k=0} \lambda'_k \sum_{l_1+...+l_k=l-1:l_j\geq 1} \overline{t}_{l_1} \cdots \overline{t}_{l_k}$$

where λ'_k is the probability that the node which is the root of an infinite path has k children strictly on the left of **LIP**. Then $\lambda'_k = Q^k \sum_{j=k+1} \lambda_j$ and so

$$\overline{f}(x) = \sum_{k=0} \lambda'_k x^k = \frac{1 - f(xQ)}{1 - xQ}.$$

By multiplying both sides by x^l and summing we get that $h(x) = x\overline{f}(g(x))$. Now we want to show that for some $a > 1$,

$$h(a) = a\overline{f}(g(a)) = a\frac{1 - f(g(a)Q)}{1 - g(a)Q}$$

is bounded uniformly in m. It suffices to show $1 - g(a)Q$ is bounded away from 0, uniformly in m for some $a > 1$. By the same arguments from claim 2, this holds as

$$g(y_0)Q = x_0 = \frac{1-p}{\lambda - p} < \frac{1}{\lambda} < 1, \quad y_0 \geq \frac{e^{\lambda-1}}{\lambda} > 1$$

and so it is possible to bound the MGF. $\square$

Remark 3. *With a bit of calculus it can be shown that the optimal choice of* $\delta = 1 - \frac{\log \alpha}{\alpha+1}$ *gives us that for* $c < \frac{(\alpha-1)-\log\alpha}{\alpha+1}$*,* $\mathcal{D}_{n,\frac{\alpha}{n}}$ *has a cn path with probability at least* $1 - K\theta(n)$*. More on this can be found in the original paper.*

4 Undirected Case

The proof will follow a simple outline : First hide ϵn edges from $\mathcal{G}_{n,p}$, and find a huge collection of long disjoint paths in the rest of the graph. We then connect these paths using the hidden edges by invoking the directed case theorem on an auxiliary graph.

4.1 Existence of long disjoint cycles

Lemma 1. *In* $\mathcal{G}_{n,p}$*, with* $p = \frac{c}{n}, c > 1$ *we have for sufficiently large* n *:*

$$\Pr(\mathcal{G}_{n,p} \textit{ is a forest}) < 2\sqrt{n}t^n.$$

Where $t = t(c) = ce^{\frac{1}{2c} - \frac{c}{2}} < 1$.

Proof. Let $G_k(n)$ be the number of graphs with n labelled vertices consisting of k disjoint trees.

$$\Pr(G(n, \frac{c}{n}) \text{ is a forest}) \leq \sum_{k=0}^{n} G_k(n)p^{n-k}(1-p)^{N-n+k}$$

$$= p^n(1-p)^{N-n}\sum_{k=0}^{n} G_k(n)(\frac{p}{1-p})^k$$

To bound the $G_k(n)t^k$ sum, we will make use of the generating function for disjoint trees on n vertices, given by :

$$\sum_{n=0}^{\infty} G_k(n)\frac{x^n}{n!} = \frac{(y - \frac{y^2}{2})^k}{k!}$$

where $x = ye^{-y}$.

Multiplying every term by t^k then adding we obtain:

$$\sum_{n=0}^{\infty}(\sum_{k=0}^{n} G_k(n)t^k)\frac{x^n}{n!}$$

$$= \sum_{k=0}^{\infty}(\sum_{n=0}^{\infty} G_k(n)\frac{x^n}{n!})t^k$$

$$= \sum_{k=0}^{\infty}(\frac{(y - \frac{y^2}{2})^k}{k!})t^k$$

$$= e^{t(y - \frac{y^2}{2})}$$

Therefore we can bound every single term by the total sum, giving :

$$\sum_{k=0}^{n} G_k(n)t^k = \frac{n!}{x^n}(\sum_{k=0}^{n} G_k(n)t^k)\frac{x^n}{n!}$$

$$\leq \frac{n!}{x^n}[\sum_{n=0}^{\infty}(\sum_{k=0}^{n} G_k(n)t^k)\frac{x^n}{n!}] = \frac{n!}{x^n}e^{t(y-\frac{y^2}{2})}$$

$$= n!(ye^{-y})^{-n}e^{t(y-\frac{y^2}{2})}$$

Where the inequality is valid for every $t < 1$ and $y < 1$. By a continuity argument we can push the statement for $y = 1$, and then take $t = \frac{p}{1-p}$ to obtain :

$$\Pr(\mathcal{G}_{n,p} \text{ is a forest}) = p^n(1-p)^{N-n}\sum_{k=0}^{n} G_k(n)(\frac{p}{1-p})^k$$

$$\leq (\frac{c}{n})^n(1-\frac{c}{n})^{N-n}n!e^ne^{\frac{t}{2}}$$

$$\leq c^ne^{\frac{-c(N-n)}{n}}n!(\frac{e}{n})^ne^{\frac{n}{2c}-\frac{1}{2}}$$

$$\leq c^ne^{\frac{-cn}{2}}e\sqrt{n}e^{\frac{n}{2c}}e^{-\frac{1}{2}}$$

$$\leq 2\sqrt{n}\alpha^n$$

where $\alpha = ce^{\frac{1}{2c}-\frac{c}{2}} < 1$. □

Lemma 2. *Given $c > 1$, there exists $\delta = \delta(c)$, $\gamma = \gamma(c)$, such that a.a.s $\mathcal{G}_{n,p}$ contains δn vertices covered by disjoint cycles of lengths $> \gamma \log n$.*

(In fact, we may strengthen the theorem by proving that the statement holds not only a.a.s but fails with exponentially small probability.)

Proof. Consider a collection of disjoint cycles maximal with respect to the total vertices covered, say kn. This implies that $\mathcal{G}_{n,p}$ contains an induced forest of size $(1-k)n$, found by simply deleting all the vertices covered. Now since every subgraph of size $n' = n(1-k)$ of $\mathcal{G}_{n,p}$ behaves like $\mathcal{G}_{n',p} = \mathcal{G}(n', \frac{(1-k)c}{n'})$, we can upper bound:

$$\Pr(\mathcal{G}_{n,p} \text{ contains a forest of size } n')$$

$$\leq \binom{n}{n'}\Pr(\mathcal{G}(n', \frac{(1-k)c}{n'}) \text{ is a forest})$$

$$\leq \binom{n}{(1-k)n}2\sqrt{n(1-k)}\alpha^{n(1-k)}$$

$$\leq 2\sqrt{1-k}(\frac{e}{(1-k)})^{(1-k)n}\sqrt{n}\alpha^{n(1-k)}$$

Where $\alpha = \alpha_{(1-k)c} = (1-k)ce^{\frac{1}{2c(1-k)}-\frac{(1-k)c}{2}}$

$$\sim e^nf(k)$$

Where $f(k) = \frac{1}{2c} + -k\log k + (1-k)\log c - (1-k)^2\frac{c}{2}$, obtained by simply grouping the exponents.

Note that $f(k)$ is monotone increasing, with $f(1) = \frac{1}{2c} > 0$, and $f(0^+)$ approaching $-\infty$, and hence has a unique root $x_0 > 0$. By choosing $k < x_0$, we are then guaranteed that a.a.s $\mathcal{G}_{n,p}$ does not contain an induced forest of size $(1-k)n$, hence a maximal disjoint union of cycles must cover at least kn vertices.

Now it remains to show that these cycles are likely to be big. We do so by proving that for any $\epsilon > 0$, there exists γ such that a.a.s no ϵn vertices in $\mathcal{G}_{n,p}$ are covered by cycles not exceeding $\gamma \log n$.

Consider a decomposition of ϵn vertices into cycles not exceeding $\gamma \log n$. We can model this structure by an arrangement of the vertices and an index set, where vertices on the same cycle appear in succession as a block following an arbitrary cycle direction, and the index set defines these blocks, i.e. where every cycle ends. This gives an upper bound of $(\epsilon n)!2^{\epsilon n}$ on the number of ordered cycle decompositions on ϵn vertices. Since we are interested in unordered decompositions and constrainted by lengths smaller than $\gamma \log n$, we know that we need at least $\frac{\epsilon n}{\gamma \log n}$, hence the total number of potential decompositions we are interested in cannot exceed

$$2^{\epsilon n}(\epsilon n)! \, \frac{1}{(\frac{\epsilon n}{\gamma \log n})!}$$

due to the possibility of rearranging at least $\frac{\epsilon n}{\gamma \log n}$ blocks in the permutations. Therefore :

$$\Pr(\epsilon n \text{ vertices are covered by small cycles}) \leq \binom{n}{\epsilon n} 2^{\epsilon n}(\epsilon n)! \, p^{\epsilon n} \, \frac{1}{(\frac{\epsilon n}{\gamma \log n})!}$$

$$\leq (\frac{en}{\epsilon n})^n (\epsilon n p e^{-\frac{1}{\gamma}})^{\epsilon n}$$

$$= ((\frac{e}{\epsilon})^{\epsilon} \alpha e^{-\frac{1}{\gamma}})^{\epsilon n} = o(1)$$

if γ is chosen small enough to guarantee $(\frac{e}{\epsilon})^{\epsilon} \alpha e^{-\frac{1}{\gamma}} < 1$.

Therefore choosing $\delta = k - \epsilon$, where $k < x_0$ and $\epsilon > 0$ are arbitrary, we are guaranteed that at least δn vertices are covered by disjoint cycles each of length $\geq \gamma \log n$. $\square$

4.2 Finding a long path

Consider a 3-step randomization process to generate $\mathcal{G}_{n,\frac{c}{n}}$, where every edge will appear first with probability $\frac{c'}{n}$ with $1 < c' < c$ (call these Type 1 edges) and then with probability q had they not already appeared (call these Type 2), where q is chosen so that the probability of every edge appearing as a Type 2 is $\frac{\epsilon}{n} = \frac{c-c'}{n}$. We will use Type 1 edges to find the big cycle cover and Type 2 edges to connect the auxiliary graph.

Fix parameter $\theta < \frac{1}{4}$. Since $c' > 1$, we have proven that a.a.s $\mathcal{G}_{n,\frac{c'}{n}}$ contains at least δn vertices covered by disjoint cycles each of length $\geq \gamma \log n$. Consider such a collection of cycles, and divide each into arcs of length L, to obtain $A_1 .... A_{n'}$ where $n' \sim \frac{\delta n}{L}$ (L is a constant chosen so $\theta^2 \delta L \epsilon > 1$) and throw away potential leftover. Further divide each arc into 3 parts : head, middle and tail, respectively of size θL, $(1-2\theta)L$, θL. Now consider the directed graph D on $\{1, ..., n'\}$, where we have a directed edge from i to j if there is a Type 2 edge from a vertex in the head of A_i to a vertex in tail of A_j.

Notice that D behave like $\mathcal{D}_{n',p'}$, since every directed edge independently has probability $1-(1-\frac{\epsilon}{n})^{\theta l^2}$, as there are θl^2 potential pairings of head-tail vertices creating it, each appearing independently with probability $\frac{\epsilon}{n}$.

To invoke the result on the directed case for D, we need to show that $p' \geq \frac{1}{n'}$, which can be done as follows :

$$p' = 1 - (1 - \frac{\epsilon}{n})^{\theta l^2}$$

$$\geq (1 - o(1)) \frac{\epsilon \theta^2 L^2}{n}$$

$$\sim (1 - o(1))\frac{\epsilon\theta^2 L\delta}{n'}$$
$$\sim (1 - o(1))\frac{\beta}{n'}$$

Where $\beta = \epsilon\theta^2 L\delta > 1$ by our choice of L. Using the result on directed $\mathcal{D}_{n,p}$, we can guarantee that a.a.s there exists a directed path P in D covering at least xn' vertices. Expanding this path in the original graph by following its trajectory along the arcs from tail to head, we are guaranteed a.a.s a path of length at least :

$$xn'(1 - 2\theta)L \sim x(1 - 2\theta)\delta n = \alpha n$$

which concludes the 2nd main result of the paper.

5 Conclusion

The technique used to derive the undirected case from the directed is worth learning as it is useful and elegant. However, it is possible to derive the main result for the undirected case by directly applying a branching process argument. In fact, for the undirected case, there is a much simpler proof due to Krivelevich and Sudakov [2] which does not use branching process. Their proof also use the DFS approach but is much shorter and is pretty self contained as it only relies on basic concentration inequalities.

References

[1] Miklós Ajtai, János Komolós and Endre Szemerédi. The Longest Path in a Random Graph, 1981; Combinatorica. DOI: 10.1007/BF02579172

[2] Michael Krivelevich and Benny Sudakov. The phase transition in random graphs: A simple proof, 2013; Random Structures & Algorithms. DOI: 10.1002/rsa.20470

[3] Noga Alon and Joel H. Spencer. The Probabilistic Method: 4th Edition, 2016; Wiley.