The Longest Path in a Random Graph

Miklós Ajtai, János Komlós, Endre Szemerédi

Presented by Zouhaier Ferchiou, Harry Sivasubramaniam

September 21, 2020

Main Result

Main Theorem ($G(n, m)$)

$G(n, (1+\epsilon)\frac{n}{2})$ contains a path of length αn a.a.s, where $\alpha = \alpha(\epsilon) > 0$ is constant.

Main Result

Main Theorem ($G(n, m)$)

$G(n, (1+\epsilon)\frac{n}{2})$ contains a path of length αn a.a.s, where $\alpha = \alpha(\epsilon) > 0$ is constant.

Main Theorem ($G(n, p)$)

For $c > \frac{1}{2}$, $G(n, \frac{c}{n})$ contains a path of length αn a.a.s, where $\alpha = \alpha(c) > 0$ is constant.

Main Result

Main Theorem ($G(n, m)$)

$G(n, (1+\epsilon)\frac{n}{2})$ contains a path of length αn a.a.s, where $\alpha = \alpha(\epsilon) > 0$ is constant.

Main Theorem ($G(n, p)$)

For $c > \frac{1}{2}$, $G(n, \frac{c}{n})$ contains a path of length αn a.a.s, where $\alpha = \alpha(c) > 0$ is constant.

These are equivalent since we can find $\epsilon_1 \leq \epsilon_2$ and a coupling such that a.a.s: $G(n, (1+\epsilon_1)\frac{n}{2}) \subseteq G(n, \frac{c}{n}) \subseteq G(n, (1+\epsilon_2)\frac{n}{2})$. (you have proved this on assignment 1)

Strategy

1. Put away ϵn edges in reserve (they won't be considered until step 3).

Strategy

1. Put away ϵn edges in reserve (they won't be considered until step 3).
2. Show that a linear number of vertices can be covered with disjoint big cycles ($\geq \delta \log(n)$) using non reserve edges. Throw away the rest of the graph.

Strategy

1. Put away ϵn edges in reserve (they won't be considered until step 3).
2. Show that a linear number of vertices can be covered with disjoint big cycles ($\geq \delta \log(n)$) using non reserve edges. Throw away the rest of the graph.
3. Split cycles into arcs, and connect them with the reserve edges.

Strategy

1. Put away ϵn edges in reserve (they won't be considered until step 3).
2. Show that a linear number of vertices can be covered with disjoint big cycles ($\geq \delta \log(n)$) using non reserve edges. Throw away the rest of the graph.
3. Split cycles into arcs, and connect them with the reserve edges.
4. Find a long path in the arc-arc graph, inducing a long path in $G(n, p)$.

Strategy

1. Put away ϵn edges in reserve (they won't be considered until step 3).
2. Show that a linear number of vertices can be covered with disjoint big cycles ($\geq \delta \log(n)$) using non reserve edges. Throw away the rest of the graph.
3. Split cycles into arcs, and connect them with the reserve edges.
4. Find a long path in the arc-arc graph, inducing a long path in $G(n, p)$.

Note: we will switch between considering $G(n, \frac{c}{n})$and $G(n, \alpha n)$ for different lemmas. Since these can be coupled, the results will hold for both a.a.s.

Strategy

Note that step 4 :

- Find a long path in the arc-arc graph

Requires a a similar theorem to our objective for directed graphs, which we will present in detail on Wednesday.

Strategy

Note that step 4 :

- Find a long path in the arc-arc graph

Requires a a similar theorem to our objective for directed graphs, which my mates will go present in detail on Wednesday.

Main Theorem ($D(n, p)$)

For $c > \frac{1}{2}$, $D(n, \frac{c}{n})$ contains a directed path of length $\alpha' n$ a.a.s, where $\alpha' = \alpha'(c) > 0$ is constant.

1 - Reserve edges

When considering $G(n, \frac{c}{n})$,$c > 1$, we can do a 2 step randomization with $\frac{c'}{n}$ and $\frac{\epsilon}{n}$ so the total probability that each edge appears is $\frac{c}{n}$, with $c' > 1$ and $\epsilon > 0$.

2 - Reserve edges

When considering $G(n, \frac{c}{n})$,$c > 1$, we can do a 2 step randomization with $\frac{c'}{n}$ and $\frac{\epsilon}{n}$ so the total probability that each edge appears is $\frac{c}{n}$, with $c' > 1$ and $\epsilon > 0$.
We will prove that we can a.a.s cover order n vertices with big cycles from $G(n, \frac{c'}{n})$ and we can 'connect them' with edges from $G(n, \frac{\epsilon}{n})$.

First we give an upper bound that $G(n, \frac{c}{n})$is a forest. This will be useful to claim that most of the graph is covered by disjoint cycles.

First we give an upper bound that $G(n, \frac{c}{n})$is a forest. This will be useful to claim that most of the graph is covered by disjoint cycles.

Lemma

For $p = \frac{c}{n}, c > 1$:

$$\Pr(G(n, \frac{c}{n})\text{is a forest}) < 2\sqrt{n}t^n.$$

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

Forest bound

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

Forest bound

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

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

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

Forest bound

To bound the $G_k(n)$ 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 $x = ye^{-y}$.

Forest bound

To bound the $G_k(n)$ 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 $x = ye^{-y}$.
Note that we are interested in bounding a sum of the form $\sum_{k=0}^{n} G_k(n)t^k$, so we can do the following :

Forest bound

To bound the $G_k(n)$ 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 $x = ye^{-y}$.
Note that we are interested in bounding a sum of the form $\sum_{k=0}^{n} G_k(n)t^k$, so we can do the following :

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

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

Forest bound

Now we can simply substitute $t = \frac{1-p}{p} = \frac{n}{c} - 1$ and $y = 1$:

Forest bound

Now we can simply substitute $t = \frac{1-p}{p} = \frac{n}{c} - 1$ and $y = 1$ to get :

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

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

Forest bound

Now we can simply substitute $t = \frac{1-p}{p} = \frac{n}{c} - 1$ and $y = 1$ to get :

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

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

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

Forest bound

Now we can simply substitute $t = \frac{1-p}{p} = \frac{n}{c} - 1$ and $y = 1$ to get :

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

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

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

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

(here we used that $N \leq \frac{n^2}{2}$ and $n! \leq e(\frac{n}{e})^n\sqrt{n}$), which is the upper Sterling approximation.)

Forest bound

Now we can simply substitute $t = \frac{1-p}{p} = \frac{n}{c} - 1$ and $y = 1$ to get :

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

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

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

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

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

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

Consider a collection of disjoint cycles, maximal with respect to the total number of vertices covered, say kn.

Consider a collection of disjoint cycles, maximal with respect to the total number of vertices covered, say kn.
$\implies G(n, \frac{c}{n})$contains a forest of size $n(1-k)$.

2 - Disjoint long cycles

Now since every subgraph of size $n' = n(1-k)$ of $G(n, \frac{c}{n})$behaves like $G(n', \frac{(1-k)c}{n'})$, we can upper bound:

$$\Pr(G(n, \frac{c}{n}) \text{ contains a forest of size } n')$$

Now since every subgraph of size $n' = n(1-k)$ of $G(n, \frac{c}{n})$behaves like $G(n', \frac{(1-k)c}{n'})$, we can upper bound:

$$\Pr(G(n, \frac{c}{n}) \text{ contains a forest of size } n')$$

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

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

2 - Disjoint long cycles

Now since every subgraph of size $n' = n(1-k)$ of $G(n, \frac{c}{n})$behaves like $G(n', \frac{(1-k)c}{n'})$, we can upper bound:

$$\Pr(G(n, \frac{c}{n}) \text{ contains a forest of size } n')$$

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

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

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

Hence we have

$$\Pr(G(n, \frac{c}{n}) \text{ contains a forest of size } n') \leq 2\sqrt{1-k}e^{nf(k)}$$

where $f(k) = (1-k)(1-\log(1-k)) + \frac{1}{2}\frac{\log n}{n} + (1-k)\log t_{c(1-k)}$.

Hence we have

$$\Pr(G(n, \frac{c}{n}) \text{ contains a forest of size } n') = 2\sqrt{1-k}e^{nf(k)}$$

where $f(k) = (1-k)(1-\log(1-k)) + \frac{1}{2}\frac{\log n}{n} + (1-k)\log t_{c(1-k)}$.

Recall that : $\log t_{c(1-k)} = \log[c(1-k)e^{\frac{1}{2c(1-k)} - \frac{c(1-k)}{2}}] =$
$\log c + \log(1-k) + \frac{1}{2c(1-k)} - \frac{c(1-k)}{2}$

Hence we have

$$\Pr(G(n, \frac{c}{n}) \text{ contains a forest of size } n') = 2\sqrt{1-k}\,e^{nf(k)}$$

where $f(k) = (1-k)(1-\log(1-k)) + \frac{1}{2}\frac{\log n}{n} + (1-k)\log t_{c(1-k)}$.

Recall that : $\log t_{c(1-k)} = \log[c(1-k)e^{\frac{1}{2c(1-k)} - \frac{c(1-k)}{2}}] =$
$\log c + \log(1-k) + \frac{1}{2c(1-k)} - \frac{c(1-k)}{2}$

Therefore : $f(k) = (1+o(1))[(1-k)(1+\log c) + \frac{1}{2c} - (1-k)^2\frac{c}{2}]$

Hence we have

$$\Pr(G(n,\frac{c}{n}) \text{ contains a forest of size } n') = 2\sqrt{1-k}e^{nf(k)}$$

where $f(k) = (1-k)(1-\log(1-k)) + \frac{1}{2}\frac{\log n}{n} + (1-k)\log t_{c(1-k)}$.

Recall that : $\log t_{c(1-k)} = \log[c(1-k)e^{\frac{1}{2c(1-k)} - \frac{c(1-k)}{2}}] =$
$\log c + \log(1-k) + \frac{1}{2c(1-k)} - \frac{c(1-k)}{2}$

Therefore :
$f(k) = (1+o(1))[(1-k)(1+\log c) + \frac{1}{2c} - (1-k)^2\frac{c}{2}] = (1+o(1))f'(k)$

Hence we have

$$\Pr(G(n, \frac{c}{n}) \text{ contains a forest of size } n') = 2\sqrt{1-k}e^{nf(k)}$$

where $f(k) = (1-k)(1-\log(1-k)) + \frac{1}{2}\frac{\log n}{n} + (1-k)\log t_{c(1-k)}$.

Recall that : $\log t_{c(1-k)} = \log[c(1-k)e^{\frac{1}{2c(1-k)} - \frac{c(1-k)}{2}}] =$
$\log c + \log(1-k) + \frac{1}{2c(1-k)} - \frac{c(1-k)}{2}$

Therefore :
$f(k) = (1+o(1))[(1-k)(1+\log c) + \frac{1}{2c} - (1-k)^2\frac{c}{2}] = (1+o(1))f'(k)$
Note that $f'(0) < 0$ for $c > 1$ and $f'(1) > 0$, with f' monotone and continuous on [0,1], so there must be a root x_0, and picking $k = \frac{x_0}{2}$ ensures that a.a.s kn vertices are covered by disjoint cycles as $f'(k) < 0$.

It remains to prove that not many of these cycles are big.

It remains to prove that not many of these cycles are big.
$\Longrightarrow$ We show that by proving we can't cover δn vertices with small cycles ($< \lambda \log n$) for any $\delta > 0$. (so we lose at most $o(n)$ to small cycles)

Hence we bound :

$$\Pr(\delta n \text{ vertices are covered by cycles of size } \leq \lambda \log n)$$

Hence we bound :

$$\Pr(\delta n \text{ vertices are covered by cycles of size } \leq \lambda \log n)$$

$$\leq \binom{n}{\delta n} 2^{\delta n} (\delta n)! p^{\delta(n)} \frac{1}{(\frac{\delta n}{\lambda \log n})!}$$

since $\delta n!$ determines an ordering of the vertices and $2^{\delta n}$ determines a subset of the vertices denoting the breakpoints of when a new cycle starts. The denominator is due to the permutation of cycle placements, since each cycle is small there must be at least $\frac{\delta n}{\lambda \log n}$ of them.

Now note that for sufficiently large m:

$$2^m m! p^m \frac{1}{(\frac{m}{\gamma \log m})!} < (mpe^{-\frac{1}{\gamma}})^m$$

Now note that for sufficiently large m:

$$2^m m! p^m \frac{1}{(\frac{m}{\gamma \log m})!} < (mpe^{-\frac{1}{\gamma}})^m$$

Hence from above applied to δn we get :

$$\Pr(\delta n \text{ vertices are covered by cycles of size } \leq \lambda \log n)$$

$$\leq \binom{n}{\delta n} 2^{\delta n} (\delta n)! p^{\delta(n)} \frac{1}{(\frac{\delta n}{\lambda \log n})!}$$

2 - Disjoint long cycles

Now note that for sufficiently large m:

$$2^m m! p^m \frac{1}{(\frac{m}{\gamma \log m})!} < (mpe^{-\frac{1}{\gamma}})^m$$

Hence from above applied to δn we get :

$$\Pr(\delta n \text{ vertices are covered by cycles of size } \leq \lambda \log n)$$

$$\leq \binom{n}{\delta n} 2^{\delta n} (\delta n)! p^{\delta(n)} \frac{1}{(\frac{\delta n}{\lambda \log n})!}$$

$$\leq \binom{n}{\delta n} (\delta c e^{-\frac{1}{\gamma}})^{\delta n}$$

Now note that for sufficiently large m:

$$2^m m! p^m \frac{1}{(\frac{m}{\gamma \log m})!} < (mpe^{-\frac{1}{\gamma}})^m$$

Hence from above applied to δn we get :

$$\Pr(\delta n \text{ vertices are covered by cycles of size } \leq \lambda \log n)$$

$$\leq \binom{n}{\delta n} 2^{\delta n} (\delta n)! p^{\delta(n)} \frac{1}{(\frac{\delta n}{\lambda \log n})!}$$

$$\leq \binom{n}{\delta n} (\delta c e^{-\frac{1}{\gamma}})^{\delta n}$$

$$\leq (\frac{e}{\delta} \delta c e^{-\frac{1}{\gamma}})^{\delta n} = o(1)$$

when γ is small enough, and hence a.a.s kn vertices are covered by disjoint long cycles.

Given disjoint long cycles ($> \lambda \log n$) that cover at least kn vertices, do the following :

Given disjoint long cycles ($> \lambda \log n$) that cover at least kn vertices, do the following :

- cut each of them into arcs of length L to obtain $A_1..A_{n'}$. (Throw away $< L$ leftover)

Given disjoint long cycles ($> \lambda \log n$) that cover at least kn vertices, do the following :

- cut each of them into arcs of length L to obtain $A_1..A_{n'}$.
- Arbitrarily assign a source/sink to endpoints of each arc.

3 - Arc connect

Given disjoint long cycles ($> \lambda \log n$) that cover at least kn vertices, do the following :

- cut each of them into arcs of length L to obtain $A_1..A_{n'}$.
- Arbitrarily assign a source/sink to endpoints of each arc.

Fix $\delta_1 < \frac{1}{4}$. We call the first $\delta_1 L$ vertices of A_i, starting from the source, the tail of A_i, and the last $\delta_1 L$ vertices, ending with the sink, the tail of A_i.

3 - Arc connect

Given disjoint long cycles ($> \lambda \log n$) that cover at least kn vertices, do the following :

- cut each of them into arcs of length L to obtain $A_1..A_{n'}$.
- Arbitrarily assign a source/sink to endpoints of each arc.

Fix $\delta_1 < \frac{1}{4}$. We call the first $\delta_1 L$ vertices of A_i, starting from the source, the tail of A_i, and the last $\delta_1 L$ vertices, ending with the sink, the tail of A_i.

- Construct a directed graph D on vertices $\{1...n'\}$.
- create an edge in D from A_i to A_j if there is a reserve edge connecting a vertex in the head of A_i to a vertex in the tail of A_j.

3 - Arc connect

For each possible edge of D, it appears with probability $p' = 1 - (1 - \frac{\epsilon}{n})^{(\delta L)^2}$ (since we have $(\delta L)^2$ possible pairs of vertices inducing the edge) independently from the rest.

For each possible edge of D, it appears with probability $p' = 1 - (1 - \frac{\epsilon}{n})^{(\delta L)^2}$ (since we have $(\delta L)^2$ possible pairs of vertices inducing the edge) independently from the rest
$\implies D$ behaves like a directed random graph with edge probability p'.

To use the theorem for the directed case, we still need to show that D is dense enough.

To use the theorem for the directed case, we still need to show that D is dense enough.
First note that $n' \sim \frac{kn}{L}$, and hence we have :

$$p' = 1 - (1 - \frac{\epsilon}{n})^{(\delta L)^2} \geq (1 - o(1))\frac{\epsilon(\delta L)^2}{n} \text{ (union bound)}$$

To use the theorem for the directed case, we still need to show that D is dense enough.
First note that $n' \sim \frac{kn}{L}$, and hence we have :

$$p' = 1 - (1 - \frac{\epsilon}{n})^{(\delta L)^2} \geq (1 - o(1))\frac{\epsilon(\delta L)^2}{n} \text{ (union bound)}$$

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

To use the theorem for the directed case, we still need to show that D is dense enough.
First note that $n' \sim \frac{kn}{L}$, and hence we have :

$$p' = 1 - (1 - \frac{\epsilon}{n})^{(\delta L)^2} \geq (1 - o(1))\frac{\epsilon(\delta L)^2}{n} \text{ (union bound)}$$

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

So by picking L large enough, we can guarantee $\epsilon k \delta^2 L$ is as big of a constant as we wish. (which we can do since the only restriction on L is that it is $o(\log n)$)

4 - Directed Black box

Using the result on dense directed random graphs, we a.a.s guarantee a directed path of length $(1 - \delta)n'$.

Using the result on dense directed random graphs, we a.a.s guarantee a directed path of length $(1-\delta)n'$.
The induced path in the original graph must cover at least $(1-\delta)n'(1-2\delta)L$ vertices.
Hence a.a.s we have a path of length :

$$(1-\delta)n'(1-2\delta)L \sim (1-\delta)(1-2\delta)kn \geq (1-3\delta)kn = \alpha n$$

Where α is a fixed constant.

Any Questions ?

Thank you !

On Wednesday we will investigate the directed case.