Coupling of Scale-Free and Classical Random Graphs

Béla Bollobás and Oliver Riordan

April 18, 2007

Introduction

Consider a graph where we delete some nodes and look at the size of the largest component remaining.
Just how robust are scale free graphs

- to random failures

Introduction

Consider a graph where we delete some nodes and look at the size of the largest component remaining.
Just how robust are scale free graphs

- to random failures
- to deliberate attacks

Introduction

Consider a graph where we delete some nodes and look at the size of the largest component remaining.
Just how robust are scale free graphs

- to random failures
- to deliberate attacks
- should the adversary have infinite time?

If we take the adversary out of the picture we can phrase the question as.

Introduction

If we take the adversary out of the picture we can phrase the question as.
How many vertices do we need to remove so that the graph breaks up into small pieces?

The Model

Consider the Preferential Attachment Model.
Assume m is fixed, $G[k] \sim G_m^{(k)}$. We create $G[k+1]$ by adding a new vertex $k+1$ with m edges $(k+1, t_1), \ldots, (k+1, t_m)$ where

$$\Pr[t_i = j] = \frac{d_{G[k],i}(j)}{2mk + 2i - 1}$$

Main Result

Theorem 1

There exist $c > 0, m_0$ s.t. for $m \geq m_0$, **whp** $G_m^{(n)}$ the following holds

- (a) Every induced subgraph of size $10n\frac{\log(m)}{m}$ contains a component of size at least $2n\frac{\log m}{m}$
- (b) The graph contains an independent set of size $cn\frac{\log(m)}{m}$

Proof Technique

Proof by coupling

We create two graphs (G_1, G_2) s.t. $G_1 \sim G_m^{(n)}$, $G_2 \sim G(n,p)$.

But G_1 and G_2 are not independent but heavily correlated.

What is coupling?

Let $X \sim D_X$, $Y \sim D_Y$. A coupling is a joint distribution $D_{X,Y}$ s.t. the marginals are correct, i.e. $D_{X,\cdot} = D_X$ and $D_{\cdot,Y} = D_Y$.

What is coupling?

Let $X \sim D_X$, $Y \sim D_Y$. A coupling is a joint distribution $D_{X,Y}$ s.t. the marginals are correct, i.e. $D_{X,\cdot} = D_X$ and $D_{\cdot,Y} = D_Y$. Suppose X, Y are uniform $[0, 1]$ and we let $Y = 1 - X$ then both marginals are still uniform but the joint distribution is far from independent.

Chernoff Bounds

Let $X \sim Bin(n,p)$ then

$$\Pr[X > (\delta+1)np] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{np}$$

$$\Pr[X < (1-\delta)np] < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{np}$$

The Coupling

Theorem 2 Fix $\eta < 1/2$, there exist constants $A, c > 0$ s.t. for fixed m we can construct $G_1 \sim G_m^{(n)}$, $G_2 \sim G(n, \eta m/n)$ s.t. **whp** $e(G_2 \setminus G_1) \leq Ae^{-cm}n$

The Coupling
Sidenote

We can create $G(n, p)$ one node at a time

- $G[1]$ contains a single node

We can create $G(n, p)$ one node at a time

- $G[1]$ contains a single node
- Given $G[k]$

We can create $G(n, p)$ one node at a time

- $G[1]$ contains a single node
- Given $G[k]$
  - generate $Y \sim Bin(k, p)$

The Coupling
Sidenote

We can create $G(n, p)$ one node at a time

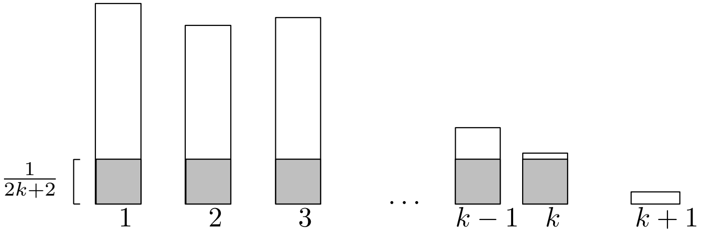
- $G[1]$ contains a single node
- Given $G[k]$
  - generate $Y \sim Bin(k, p)$
  - let S be a random subset of $[k]$ of size Y

We can create $G(n,p)$ one node at a time

- $G[1]$ contains a single node
- Given $G[k]$
  - generate $Y \sim Bin(k,p)$
  - let S be a random subset of $[k]$ of size Y
  - create edges between k and S

We grow G_1, G_2 one node at a time. Start with $G_1[k_0] \sim G_m^{(n)}$, $G_2[k_0] \sim G(k_0, \eta m/n)$, where k_0 grows slowly with n.
Let $t_1, \ldots, t_m$ be the vertices picked when growing $G_1[k]$. For $j \neq k+1$ we have

$$\Pr(t_i = j) = \frac{d_{G[k],i}(j)}{2mk + 2i - 1} \geq \frac{m}{2km + 2m} = \frac{1}{2k+2}$$



Construct $s_1, \ldots, s_m \in [k] \cup \varnothing$, s.t. s_i's are independent $\Pr(s_i = j) = \frac{1}{2k+2}$ for $j \neq \varnothing$ and $s_i = j$ implies $t_i = j$.

Let $X = |\{i|s_i \neq \varnothing\}|$, then $X \sim Bin(m, \frac{k}{2k+2})$.
Let $Y \sim Bin(k, \eta m/n)$, then

$$E[X] = \frac{mk}{2k+2} \geq (1+\epsilon)\eta\frac{mk}{n} = (1+\epsilon)E[Y]$$

So $\Pr[X < Y] \leq Ae^{-c'm}$ for some constants $A, c' > 0$.

All the s_i's are distinct w.p $1 - O(\frac{m^2}{k})$, given this and X the set $S_1 = \{s_i | s_i \neq \varnothing\}$ is a random subset of $[k]$ of size X.

All the s_i's are distinct w.p $1 - O(\frac{m^2}{k})$, given this and X the set $S_1 = \{s_i | s_i \neq \varnothing\}$ is a random subset of $[k]$ of size X.
If $Y \leq X$ we pick a random subset S_2 of size Y from S_1.

All the s_i's are distinct w.p $1 - O(\frac{m^2}{k})$, given this and X the set $S_1 = \{s_i | s_i \neq \varnothing\}$ is a random subset of $[k]$ of size X.
If $Y \leq X$ we pick a random subset S_2 of size Y from S_1.
Otherwise we pick S_2 a random subset of size Y from $[k]$, this happens w.p. $Ae^{-c'm} + o(1)$.

The Coupling

Proof

Set $D_{k+1} = \#$of edges added to G_2 and not G_1.
$\Pr[D_{k+1} > 0] = Ae^{-c'm} + o(1)$, and $D_{k+1} \leq Y$ so
$E[D_{k+1}] \leq Ame^{-c'm}$

The number of edges in $G_2 \setminus G_2$ is at most

$$\binom{k_0}{2} + \sum_{k=k_0+1}^{n} D_k$$

Assuming $k_0 = n^{1/4}$ we see that the sum has expected value at most $Ane^{-c'm}$ and is concentrated around its mean.

The number of edges in $G_2 \setminus G_2$ is at most

$$\binom{k_0}{2} + \sum_{k=k_0+1}^{n} D_k$$

Assuming $k_0 = n^{1/4}$ we see that the sum has expected value at most $Ane^{-c'm}$ and is concentrated around its mean. QED

Proof of Theorem 1(a)

Let G_1, G_2 be as described in Theorem 2. Suppose G_1 has an induced subgraph on a set V, with $|V| = 10n^{\frac{\log(m)}{m}}$ where every component has size less than $2n^{\frac{\log(m)}{m}}$. Then we can partition V into two set V_1, V_2 s.t. $|V_1|, |V_2| \geq 4n^{\frac{\log(m)}{m}}$ and G_1 has no V_1-V_2 edges.

Proof of Theorem 1(a)

Let $x = Ae^{-c'm}n$, if the coupling works then G_2 has at most x, V_1-V_2 edges.
The number of V_1-V_2 edges in G_2 is a binomial with mean

$$\mu = |V_1||V_2|\eta m/n \geq 24\eta(\log(m))^2 n/m$$

Pick m large enough s.t. $x < \mu/100$

Proof of Theorem 1(a)

By Chernoff bounds the probability that we have fewer than x edges crossing V_1, V_2 is at most $e^{-11\mu/12}$. Now

$$\Pr[G_2 \text{ has at most } x \quad V_1 - V_2 \text{ edges}] \leq \binom{n}{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

Proof of Theorem 1(a)

By Chernoff bounds the probability that we have fewer than x edges crossing V_1, V_2 is at most $e^{-11\mu/12}$. Now

$$\Pr[G_2 \text{ has at most } x \quad V_1 - V_2 \text{ edges}] \leq \binom{n}{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\leq \left(\frac{en}{|V|}\right)^{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

Proof of Theorem 1(a)

By Chernoff bounds the probability that we have fewer than x edges crossing V_1, V_2 is at most $e^{-11\mu/12}$. Now

$$\Pr[G_2 \text{ has at most } x \quad V_1 - V_2 \text{ edges}] \le \binom{n}{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\le \left(\frac{en}{|V|}\right)^{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\le \left(\frac{2em}{10\log(m)}\right)^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

Proof of Theorem 1(a)

By Chernoff bounds the probability that we have fewer than x edges crossing V_1, V_2 is at most $e^{-11\mu/12}$. Now

$$\Pr[G_2 \text{ has at most } x \quad V_1 - V_2 \text{ edges}] \leq \binom{n}{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\leq \left(\frac{en}{|V|}\right)^{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\leq \left(\frac{2em}{10\log(m)}\right)^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\leq e^{(10-22\eta)(\log(m))^2 n/m}$$

Proof of Theorem 1(a)

By Chernoff bounds the probability that we have fewer than x edges crossing V_1, V_2 is at most $e^{-11\mu/12}$. Now

$$\Pr[G_2 \text{ has at most } x \quad V_1 - V_2 \text{ edges}] \leq \binom{n}{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\leq \left(\frac{en}{|V|}\right)^{|V|} 2^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\leq \left(\frac{2em}{10\log(m)}\right)^{|V|} e^{-22\eta(\log(m))^2 n/m}$$

$$\leq e^{(10-22\eta)(\log(m))^2 n/m}$$

QED

Overview of the proof

The idea of the proof was this

Overview of the proof

The idea of the proof was this

- We create G_1, G_2 simultaneously s.t.

Overview of the proof

The idea of the proof was this

- We create G_1, G_2 simultaneously s.t.
- a bad event in G_1 implies w.h.p another bad event in G_2

Overview of the proof

The idea of the proof was this

- We create G_1, G_2 simultaneously s.t.
- a bad event in G_1 implies w.h.p another bad event in G_2
- which is easy to show is rare in G_2.

Overview of the proof

The idea of the proof was this

- We create G_1, G_2 simultaneously s.t.
- a bad event in G_1 implies w.h.p another bad event in G_2
- which is easy to show is rare in G_2.
- But we are working in the joint probability space so this implies that the bad event is rare in G_1.

Another Coupling

Theorem 3
Let $\epsilon > 0$ be given, there is a constant C s.t. for fixed m we can couple G_1, G_2 s.t. $G_1 \sim G_m^{(n)}$, $G_2 \sim G(n, Cm/n)$ s.t. **whp** G_2 contains $G_1 \setminus V$ for a set of vertices V s.t. $|\{i \in V | i \geq \epsilon n\}| \leq \epsilon n/m$.

We start with $G_1[k_0] \sim G_m^{(k_0)}$ and $G_2[k_0] \sim G(k_0, Cm/n)$ independent, and here $k_0 = \epsilon n$.
A vertex is *bad at time k* if it has degree at leas Am in $G_1[k]$.

We need the following result from Bollobás et al.: Vertex i has degree $d+m$ in $G_m^{(n)}$ with probability

$$o(n^{-1}) + (1+o(1))\binom{d+m-1}{m-1}(\frac{i}{n})^{m/2}(1-\frac{i}{n})^d$$

i.e. for $i \geq \epsilon n$ it is exponentially small in d for large d
If we choose A large enough the expected number of bad vertices is at most $e^{-Bm}n$ so **whp** there are at most $\epsilon n/m$ bad vertices.

Let $t_1, \ldots, t_m$ be defined as in the previous proof. Let V_k be the set of vertices j, s.t. $k_0 \leq j \leq k$ that are good. Now for $j \in V_k$

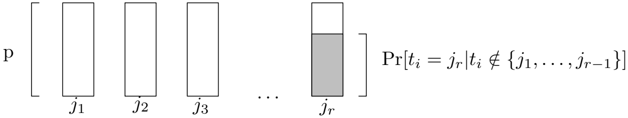
$$\Pr[t_i = j] \leq (Am + m)/(2mk) \leq A/k \leq A\epsilon^{-1}/n$$

and

$$\Pr[t_i \notin V_k] \geq k_0 m/(2mn) = \epsilon/2$$

so $\Pr[t_i = j_r | t_i \notin \{j_1, \ldots, j_{r-1}\}] \leq 2A\epsilon^{-2}/n = p$

Let T_i be a random subset of V_k s.t. every vertex is picked independently with probability p. Then we can couple T_i and t_i s.t. $t_i \in T_i$ if t_i is good.



Let S_1 be the set of good vertices adjacent to $k+1$ in G_1. Let $S_2 = T_1 \cup T_2 \cup \dots T_m$. Then S_2 is a random subset of V_k where every vertex is picked w.p. $1-(1-p)^m \leq mp \leq Cm/n$ for some large C.
This completes the proof since we've constructed S_1 s.t. $S_1 \subseteq S_2$.

For a constant $a > 1$ we know that $G_{n,\frac{a}{n}}$ has an independent set of size $\approx \frac{\log(a)}{a} n$ when $a \to \infty$.
Now couple G_1 and G_2 as in Theorem 3, with $\epsilon = \frac{1}{2}$ and assume $C > 2$. Look at the subgraph of G_2 of vertices that come in after $\frac{n}{2}$.
It has distribution $G_{\frac{n}{2},\frac{Cm}{n}} \sim G_{\frac{n}{2},\frac{Cm/2}{n/2}}$ so it has an independent set of size $\approx n\frac{\log(m)}{Cm}$.
Removing the bad vertices from this set we see that this set is also independent in G_1. So G_1 has an independent set of size $n\frac{\log(m)}{Cm} - \frac{n}{2m}$. By picking m large enough we get the result.

Thank You

The End

Thank You

Questions?