Limited choice and randomness in evolution of networks

Limit Theorems in Probability
IMI-IISC, January 2013

Shankar Bhamidi

Department of Statistics and Operations Research
University of North Carolina

January, 2013

Plan of the talk

Lecture content

- Power of choice in computer science
- *Topic 1:* Bounded size rules
- Critical scaling window and emergence of the giant (joint work with Amarjit Budhiraja and Xuan Wang)
- *Topic 2:* Twitter event networks
- Superstar model (joint work with J.Michael Steele and Tauhid Zaman)

Power of two choices

Application setting

- Consider n bins (servers) into which we are going to sequentially place n balls (jobs).
- Centralized scheme (asking bins current load) computationally expensive and time consuming
- Simplest scheme, each stage choose bin at random and place ball
- Each ball has $\sim Poi(1)$ # of balls at end

$$\text{Max load} \sim \Theta(\log n / \log(\log n))$$

- **Limited choice** Choose 2 bins u.a.r.
- Put ball in bin with minimal # of balls at that stage

$$\text{Max load} \sim \Theta(\log\log n)$$

Network models

Motivation

- Last few years have seen an explosion in empirical data on real world networks.
- Has motivated an interdisciplinary study in understanding the emergence of properties of these network models.
- Formulation of many mathematical models of network formation.

Limited choice

- Incorporate effect of limited choice in network formation
- Simple variants of standard models give much better fit but hard to mathematically analyze

Erdos-Renyi random graph

Setting

- n vertices
- Edge probability t/n
- Phase transition at $t = 1$

# of edges $\sim n/2$

- $t < 1, \mathcal{C}_1(t) \sim \log n$
- $t > 1, \mathcal{C}_1 \sim f(t)n$
-

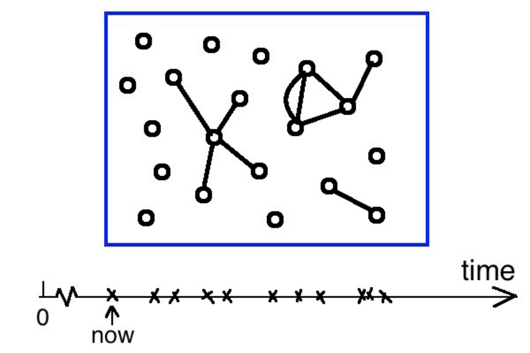
$$t = 1 + \frac{1}{n^{1/3}}$$

Beautiful math theory

Bounded size rules

The Erdős-Rényi random graph of $\mathcal{G}_n^{ER}$

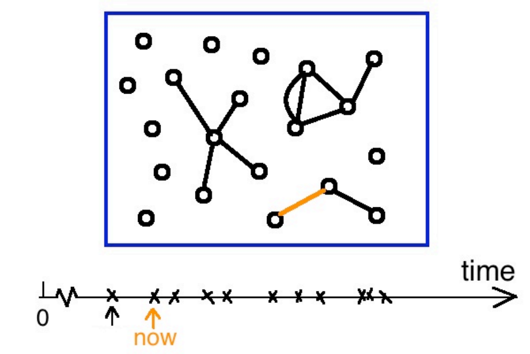
- $\mathcal{G}_n(0) = \mathbf{0}_n$ the graph with n vertices but no edges
- Each step, choose one edge e uniformly among all $\binom{n}{2}$ possible edges, and add it to the graph.
- $\mathcal{G}_n(t)$: add edges at rate $n/2$.



The Erdős-Rényi random graph process

The Erdős-Rényi random graph of $\mathcal{G}_n^{ER}$

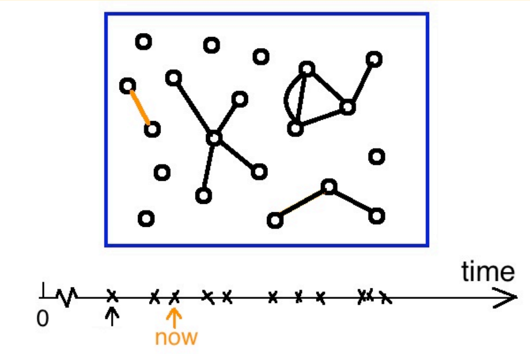
- $\mathcal{G}_n(0) = \mathbf{0}_n$ the graph with n vertices but no edges
- Each step, choose one edge e uniformly among all $\binom{n}{2}$ possible edges, and add it to the graph.
- $\mathcal{G}_n(t)$: add edges at rate $n/2$.



The Erdős-Rényi random graph process

The Erdős-Rényi random graph of $\mathcal{G}_n^{ER}$

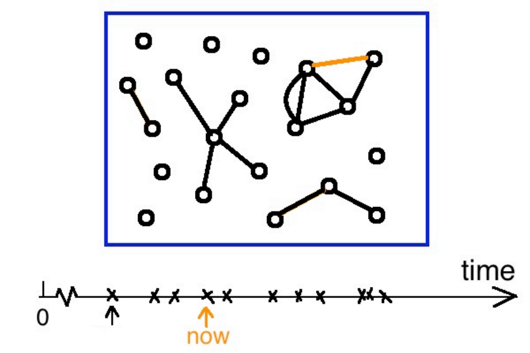
- $\mathcal{G}_n(0) = \mathbf{0}_n$ the graph with n vertices but no edges
- Each step, choose one edge e uniformly among all $\binom{n}{2}$ possible edges, and add it to the graph.
- $\mathcal{G}_n(t)$: add edges at rate $n/2$.



The Erdős-Rényi random graph process

The Erdős-Rényi random graph of $\mathcal{G}_n^{ER}$

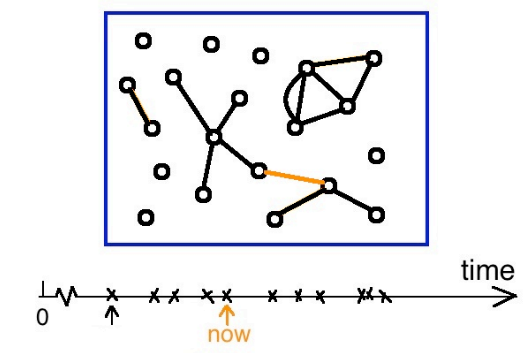
- $\mathcal{G}_n(0) = \mathbf{0}_n$ the graph with n vertices but no edges
- Each step, choose one edge e uniformly among all $\binom{n}{2}$ possible edges, and add it to the graph.
- $\mathcal{G}_n(t)$: add edges at rate $n/2$.



The Erdős-Rényi random graph process

The Erdős-Rényi random graph of $\mathcal{G}_n^{ER}$

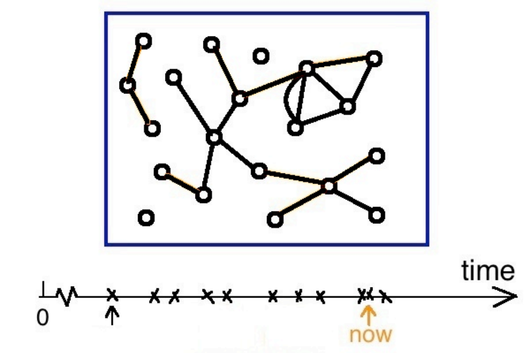
- $\mathcal{G}_n(0) = \mathbf{0}_n$ the graph with n vertices but no edges
- Each step, choose one edge e uniformly among all $\binom{n}{2}$ possible edges, and add it to the graph.
- $\mathcal{G}_n(t)$: add edges at rate $n/2$.



The Erdős-Rényi random graph process

The Erdős-Rényi random graph of $\mathcal{G}_n^{ER}$

- $\mathcal{G}_n(0) = \mathbf{0}_n$ the graph with n vertices but no edges
- Each step, choose one edge e uniformly among all $\binom{n}{2}$ possible edges, and add it to the graph.
- $\mathcal{G}_n(t)$: add edges at rate $n/2$.



The Erdős-Rényi random graph process

The phase transition of $\mathcal{G}_n^{ER}(t)$

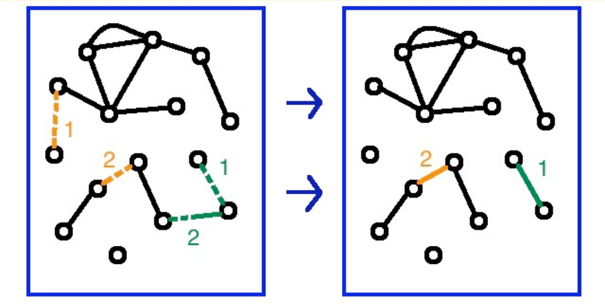
- The giant component: the component contains $\Theta(n)$ vertices.
- Let $\mathcal{C}_n^{(k)}(t)$ be the size of the k^{th} largest component
- $t_c = t_c^{ER} = 1$ is the critical time.
- (super-critical) when $t > 1$, $\mathcal{C}_n^{(1)} = \Theta(n)$, $\mathcal{C}_n^{(2)} = O(\log n)$.
- (sub-critical) when $t < 1$, $\mathcal{C}_n^{(1)} = O(\log n)$, $\mathcal{C}_n^{(2)} = O(\log n)$.
- (critical) when $t = 1$, $\mathcal{C}_n^{(1)} \sim n^{2/3}$, $\mathcal{C}_n^{(2)} \sim n^{2/3}$.

- after initial work by [ER1960], further work by [JKLP1994], finally proved by [Aldous1997].
- Merging dynamics through the scaling window of the components described by a Markov Process called the multiplicative coalescent.
- Formal existence of multiplicative coalescent.

Bounded size rules: Effect of limited choice

[Bohman, Frieze 2001]The Bohman-Frieze random graph

- Motivated by very interesting question of D. Achlioptas. Delay emergence of giant component using simple rules
- Each step, two candidate edges (e_1, e_2) chosen uniformly among all $\binom{n}{2} \times \binom{n}{2}$ possible pairs of edges. If e_1 connects two singletons (component of size 1), add e_1 to the graph; else add e_2.
- Consider continuous time version where between any pair of edges, poisson process with rate $2/n^3$.



The Bohman-Frieze process

[Bohman, Frieze 2001] The delay of phase transition

Consider the continuous time version $\mathcal{G}_n^{BF}(t)$, then there exists $\epsilon > 0$ such that at time $t_c^{ER} + \epsilon$,

$$\mathcal{C}_n^{(1)}(t_c^{ER} + \epsilon) = o(n)$$

[Spencer, Wormald 2004] The critical time

- $t_c^{BF} \approx 1.1763 > t_c^{ER} = 1$.
- (super-critical) when $t > t_c$, $\mathcal{C}_n^{(1)} = \Theta(n)$, $\mathcal{C}_n^{(2)} = O(\log n)$.
- (sub-critical) when $t < t_c$, $\mathcal{C}_n^{(1)} = O(\log n)$, $\mathcal{C}_n^{(2)} = O(\log n)$.

Near Criticality

- Janson and Spencer (2011) analyzed how $s_2(\cdot), s_3(\cdot) \to \infty$ as $t \uparrow t_c$.
- Kang, Perkins and Spencer (2011) analyze the near subcritical $(t_c - \epsilon)$ regime.

General bounded size rules

- Fix $K \geq 1$
- Let $\Omega_K = \{1, 2, \ldots, K, \omega\}$
- $$c(v) = \begin{cases} |C(v)| & \text{if } |C(v)| \leq K \\ \omega & \text{otherwise} \end{cases}$$
- General bounded size rule: subset $F \subset \Omega_K^4$.
- Pick two edges (v_1, v_2) and (v_3, v_4) at random. If $(c(v_1), c(v_2), c(v_3), c(v_4)) \in F$ then choose edge e_1 else e_2

BF model

$K = 1$, $F = \{(1, 1, \alpha, \beta)\}$.

Main questions

- Question: when $t = t_c$, do we have $\mathcal{C}_n^{(1)} \sim n^{2/3}$? How do components merge? scaling window?
- What about the surplus of the largest components in the scaling window?

Notation

- $\mathcal{C}_n^{(i)}(t)$ size of i-th largest component at time t
- Surplus (Complexity) of a component

$$\xi_n^{(i)}(t) = E(\mathcal{C}_n^{(i)}(t)) - (\mathcal{C}_n^{(i)}(t) - 1)$$

- $l_\downarrow^2 = \{(x_i)_{i\geq 1} : x_1 \geq x_2 \geq \cdots \geq 0, \sum_i x_i^2 < \infty\}$
- $l_\downarrow^{2,*} = \left\{(x_i, y_i)_{i\geq 1} : (x_i) \in l_\downarrow^2, y_i \in \mathbb{Z}_+, \sum_i x_i y_i < \infty\right\}$
- $d((x,y),(x',y')) = \sqrt{\sum_i (x_i - x_i')^2} + \sum_i |x_i y_i - x_i' y_i'| + \sum_{i=1}^\infty \frac{|y_i - y_i'|}{2^i}$

The Erdős-Rényi random graph

Theorem (Aldous 1997)

Let $(\mathcal{C}_n^{(1)}(t), \mathcal{C}_n^{(2)}(t), ...)$ *be the component sizes of* $\mathcal{G}_n^{ER}(t)$ *in decreasing order and* $\xi_i(t)$ *the corresponding complexity (surplus). Define rescaled size vector* $\mathbf{C}_n^*(\lambda)$, $-\infty < \lambda < +\infty$ *as*

$$\left(\left(\frac{1}{n^{2/3}} \mathcal{C}_n^{(i)}(t_c + \frac{\lambda}{n^{1/3}}), \xi_n^{(i)}(t_c + \frac{\lambda}{n^{1/3}})\right) : i \geq 1 \right)$$

Then $\mathbf{C}_n(\lambda) \xrightarrow{d} \mathbf{X}(\lambda) = (X(\lambda), \xi(\lambda))$. *Here* $(X(\lambda), -\infty < \lambda < +\infty)$ *is the standard multiplicative coalescent, a continuous time Markov process on the state space* $l^2_\downarrow$.

Distribution for fixed λ

- For fixed $\lambda \in \mathbb{R}$, let
$$W_\lambda(t) = W(t) + \lambda t - \frac{t^2}{2},$$
- $\bar{W}_\lambda(\cdot)$ is the above process reflected at 0.
- $X(\lambda)$ has same distribution as lengths of excursions away from 0 of $\bar{W}(\cdot)$ arranged in decreasing order

The standard multiplicative coalescent $\mathbf{X}(\lambda)$

Dynamics of $\mathbf{X}(\lambda)$

- suppose $\mathbf{X}(\lambda) = (x_1, x_2, x_3, \dots)$, each x_l is viewed as the size of a cluster.
- each pair of clusters of sizes (x_i, x_j) merges at rate $x_i x_j$ into a cluster of size $x_i + x_j$.
- if x_i, x_j is merging, then $(x_1, x_2, x_3, \dots) \rightsquigarrow (x_1', x_2', x_3', \dots)$ where the latter is the re-ordering of $\{x_i + x_j, x_l : l \neq i, j\}$.

Bounded size rules

Theorem (Bhamidi, Budhiraja, Wang, 2012)

Let $(\mathcal{C}_n^{(1)}(t), \mathcal{C}_n^{(2)}(t), \ldots)$ be the component sizes of $\mathcal{G}_n^{BSR}(t)$ in decreasing order and $\xi_i(t)$ the corresponding surplus. Define the rescaled size vector $\mathbf{C}_n(\lambda)$, $-\infty < \lambda < +\infty$ as the vector

$$((\bar{\mathcal{C}}_i(\lambda), \xi_i(\lambda) : i \geq 1) = \left(\frac{\beta^{1/3}}{n^{2/3}}\mathcal{C}_n^{(i)}(t_c + \frac{\beta^{2/3}\alpha\lambda}{n^{1/3}}), \xi_i(t_c + \frac{\beta^{2/3}\alpha\lambda}{n^{1/3}}) : i \geq 1\right)$$

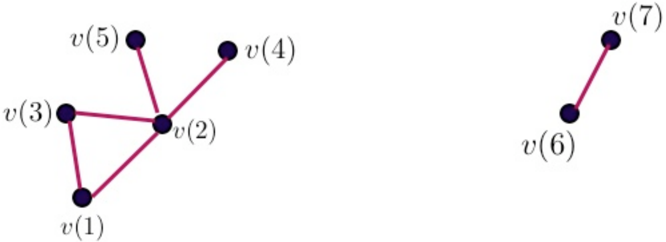
where α, β are constants determined by the BSR process. Then

$$\mathbf{C}_n(\lambda) \xrightarrow{d} \mathbf{X}(\lambda)$$

where $(\mathbf{X}(\lambda), -\infty < \lambda < +\infty)$ is the standard augmented multiplicative coalescent and convergence happens in $l_{\downarrow}^{2,*}$ with metric d.

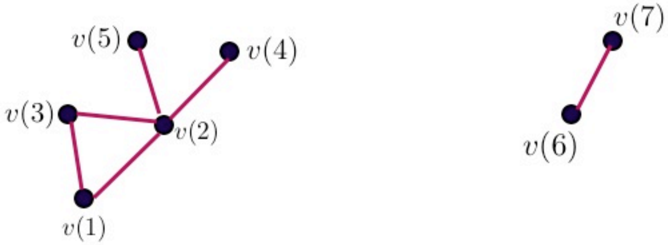
Typical method of proof: Exploration

Typical method of proof: Exploration



Typical method of proof: Exploration

$c(1) = 2$
$c(2) = 2$
$c(3) = 0$
$\cdots$



Typical method of proof

Exploration of the graph

- Explore the components of the graph one by one
- choose a vertex. Let $c(1)$ be the number of children of this vertex
- choose one of the children of this vertex, let $c(2)$ be number of children of this vertex
- continue, when one component completed move onto another component
- Define $Z(0) = 0, \quad Z(i) = Z(i-1) + c(i) - 1$
- $Z(\cdot) = -1$ for the first time when we finish exploring component 1, then hits -2 for first time when exploring component 2 and so on.
- Try to use Martingale functional limit theorem to show $\frac{1}{n^{1/3}} Z(n^{2/3}t) \to_d W^{\lambda}(t)$

Bounded size rules

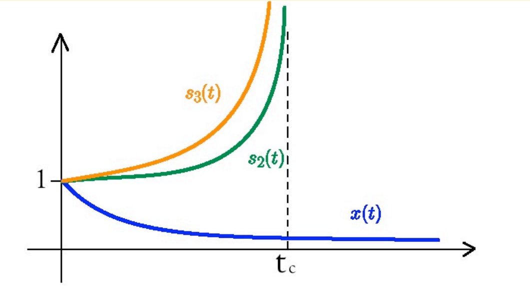
- Hard to think about exploration process especially at criticality
- Turns out: Easier to analyze the entire process

Proof idea: The Bohman-Frieze process

Where does t_c come from ?

Define $X_n(t) = \#$ of singletons, $S_2(t) = \sum_i (\mathcal{C}_n^{(i)}(t))^2$, $S_3(t) = \sum_i (\mathcal{C}_n^{(i)})^3$.
and $\bar{x}_n(t) = X_n(t)/n$, $\bar{s}_2(t) = S_2/n$, $\bar{s}_3(t) = S_3/n$.
Then [Spencer, Wormald 2004] for any fix $t > 0$,

$$\bar{x}_n(t) \xrightarrow{\mathbb{P}} x(t), \qquad \bar{s}_2(t) \xrightarrow{\mathbb{P}} s_2(t), \qquad \bar{s}_3(t) \xrightarrow{\mathbb{P}} s_3(t)$$



Why?

Behavior of $x_n(t)$

- In small time interval $[t, t+\Delta(t))$, $x_n(t) \to x_n(t) - 1/n$ at rate
$$\frac{2}{n^3}\left(\binom{n}{2} - \binom{X_n(t)}{2}\right) X_n(t)(n - X_n(t)) \sim n(1 - x_n^2(t))x_n(t)(1 - x_n(t))$$
- $[t, t+\Delta(t))$, $x_n(t) \to x_n(t) - 2/n$ at rate
$$\frac{2}{n^3}\left[\binom{X_n(t)}{2}\binom{n}{2} + \left(\binom{n}{2} - \binom{X_n(t)}{2}\right)\binom{X_n(t)}{2}\right] \sim \frac{1}{2}(x_n^2(t) + (1 - x_n^2(t)x_n^2(t)))$$
- Suggests that $x_n(t) \to x(t)$ where
$$x'(t) = -x^2(t) - (1 - x^2(t))x(t) \qquad \text{for } t \in [0, \infty,) \qquad x(0) = 1$$
- Similar analysis suggests that for $\bar{s}_2(t), \bar{s}_3(t)$
$$s_2'(t) = x^2(t) + (1 - x^2(t))s_2^2(t) \qquad \text{for } t \in [0, t_c), \qquad s_2(0) = 1$$
$$s_3'(t) = 3x^2(t) + 3(1 - x^2(t))s_2(t)s_3(t) \qquad \text{for } t \in [0, t_c), \qquad s_3(0) = 1.$$

The Bohman-Frieze process

Scaling exponents of s_2 and s_3 (Janson, Spencer 11)

- Functions $x(t), s_2(t), s_3(t)$ are determined by some differential equations
- Differential equations imply $\exists$ constants α, β such that $t \uparrow t_c$

$$s_2(t) \sim \frac{\alpha}{t_c - t}$$

$$s_3(t) \sim \beta (s_2(t))^3 \sim \beta \frac{\alpha^3}{(t_c - t)^3}$$

I: Regularity conditions of the component sizes at "$-\infty$"

- Let $\bar{\mathbf{C}}(\lambda) = n^{-2/3}\mathbf{C}\left(t_c + \beta^{2/3}\alpha\lambda/n^{1/3}\right)$.
- For $\delta \in (1/6, 1/5)$ let $t_n = t_c - n^{-\delta} = t_c + \beta^{2/3}\alpha\frac{\lambda_n}{n^{1/3}}$, then $\lambda_n = -\beta^{2/3}\alpha n^{1/3-\delta}$.
- Need to verify the three conditions

$$\frac{\sum_i \left(\bar{\mathcal{C}}_i(\lambda_n)\right)^3}{\left[\sum_i \left(\bar{\mathcal{C}}_i(\lambda_n)\right)^2\right]^3} \xrightarrow{\mathbb{P}} 1 \qquad \Leftrightarrow \qquad \frac{n^2 S_3(t_n)}{S_2^3(t_n)} \xrightarrow{\mathbb{P}} \beta$$

$$\frac{1}{\sum_i \left(\bar{\mathcal{C}}_i(\lambda_n)\right)^2} + \lambda_n \xrightarrow{\mathbb{P}} 0 \qquad \Leftrightarrow \qquad \frac{n^{4/3}}{S_2(t_n)} - \frac{n^{-\delta+1/3}}{\alpha} \xrightarrow{\mathbb{P}} 0$$

$$\frac{\bar{\mathcal{C}}_1(\lambda_n)}{\sum_i \left(\bar{\mathcal{C}}_i(\lambda_n)\right)^2} \xrightarrow{\mathbb{P}} 0 \qquad \Leftrightarrow \qquad \frac{n^{2/3}\mathcal{C}_n^{(1)}(t_n)}{S_2(t_n)} \xrightarrow{\mathbb{P}} 0$$

II: Dynamics of merging in the critical window

The dynamic of merging

- In any small time interval $[t, t+dt)$, two components i and j merge at rate

$$\frac{2}{n^3}\left[\binom{n}{2} - \binom{X_n(t)}{2}\right]\mathcal{C}_i(t)\mathcal{C}_j(t)$$
$$\sim \frac{1}{n}(1-\bar{x}^2(t))\mathcal{C}_i(t)\mathcal{C}_j(t)$$

Let $\lambda = (t-t_c)n^{1/3}/\alpha\beta^{2/3}$ be rescaled time paramter, rate at which two components merge

$$\gamma_{ij}(\lambda) \sim \frac{(1-x^2(t_c+\beta^{2/3}\alpha\frac{\lambda}{n^{1/3}}))}{n}\frac{\beta^{2/3}\alpha}{n^{1/3}}\mathcal{C}_i\left(t_c+\frac{\beta^{2/3}\alpha\lambda}{n^{1/3}}\right)\mathcal{C}_j\left(t_c+\frac{\beta^{2/3}\alpha\lambda}{n^{1/3}}\right)$$
$$= \alpha\left(1-x^2\left(t_c+\beta^{2/3}\alpha\frac{\lambda}{n^{1/3}}\right)\right)\bar{\mathcal{C}}_i(\lambda)\bar{\mathcal{C}}_j(\lambda)$$
$$= \bar{\mathcal{C}}_i(\lambda)\bar{\mathcal{C}}_j(\lambda) \qquad \text{since } \alpha(1-x^2(t_c)) = 1$$

How to check regularity conditions

Analysis of $\mathcal{C}_n^{(1)}(t)$

Key point: need to get refined bounds on maximal component in barely subcritical regime.

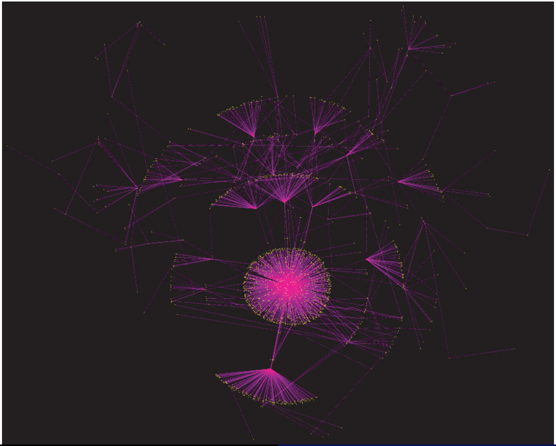
Lemma (Bounds on the largest component)

Let $\delta \in (0, 1/5)$, t_c be the critical time for the BF process, $\mathcal{C}_n^{(1)}(t)$ be the size of the largest component. Then there exists a constant $B = B(\delta)$ such that as $n \to +\infty$,

$$\mathbb{P}\{\mathcal{C}_n^{(1)}(t) \leq \frac{B \log^4 n}{(t_c - t)^2} \text{ for all } t < t_c - n^{-\delta}\} \to 1$$

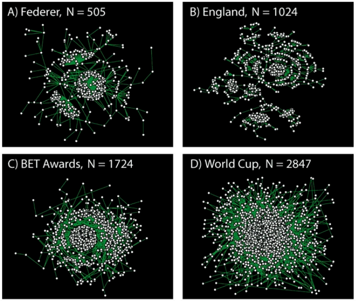
From the Retweet Graph to the Superstar Model

- Joint work with J Michael Steele (Wharton) and Tauhid Zaman (MIT).
- Retweet graph: Given a topic and a time frame — form all the (undirected) *retweet arcs* and look at the graph you get.



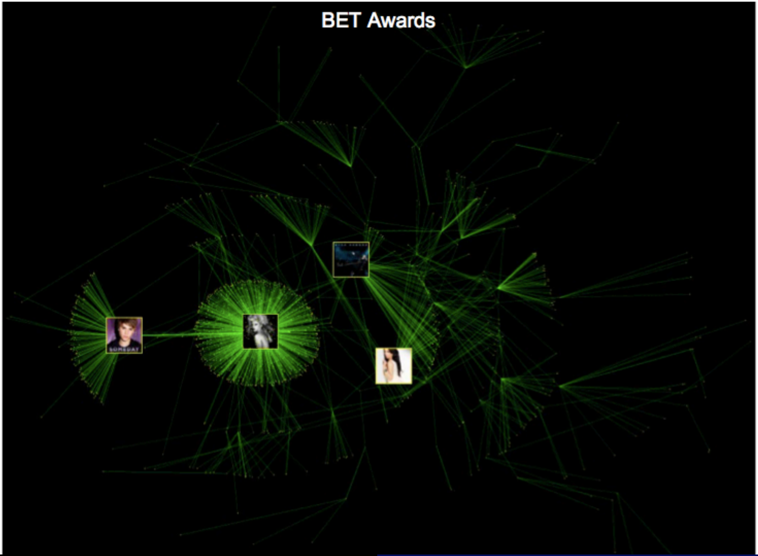
Some Empirical Retweet Graphs

- Retweet graphs were constructed for 13 different public events [1]
  - Sports, breaking news stories, and entertainment events
  - Time range for each topic was between 4-6 hours
- Graphs are very tree-like (few cycles)
- Graphs each have one giant component which we want to study
- We treat the graph as undirected

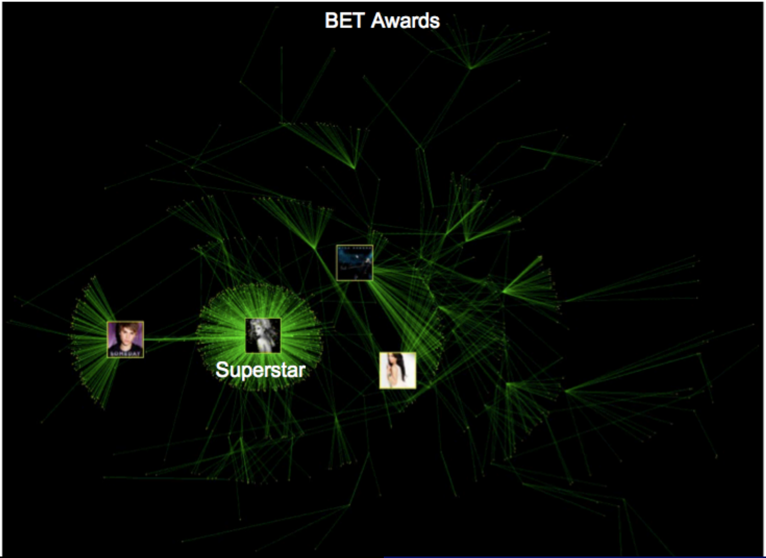


[1] Data courtesy of Microsoft Research, Cambridge, MA

The superstar model

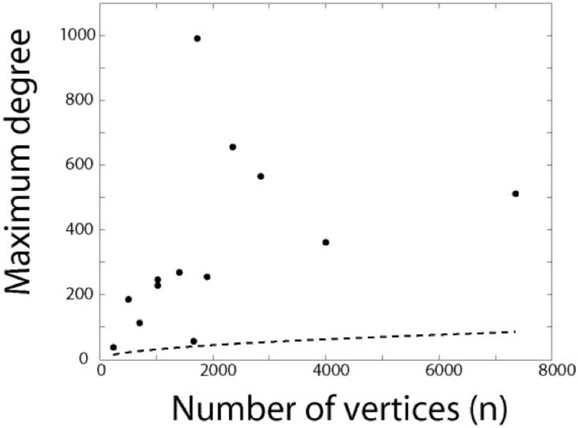


The superstar model

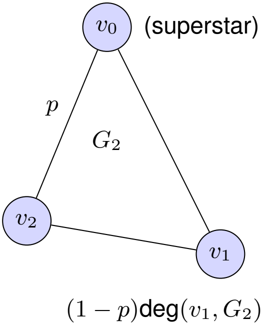


The superstar model

- Max degree in retweet graph is on the order of graph size (i.e. $M_G \sim pn$)
- Preferential attachment predicts sub-linear max degree



The Superstar Model



- Attach to superstar with probability p
- Else with probability $1-p$ attach to one of the non-superstar vertices.
- Non-SS Attachment Rule: probability proportional to its degree (preferential attachment rule)

The only model parameter is p: The superstar parameter

This is a very simple model: But (1) it has empirical benefits and (2) it is tractable — though not particularly easy.

Superstar Degree

Theorem

Let $\deg(v_0, G_n)$ *be the superstar degree. Then we have that*

$$\frac{\deg(v_0, G_n)}{n} \to p \qquad \text{with probability 1 as } n \to \infty$$

- Empirically the Superstar degree is $\Theta(n)$ and the Superstar Model "Bakes this into the Cake"
- But that is ALL that is baked in...
- The value of p determines other features of the graph — the Superstar Model is *testable*.

Non-Superstar Degree

Theorem

Let $\deg_{\max}(G_n)$ *be the maximal non-superstar degree:*

$$\deg_{\max}(G_n) = \max_{1 \le i \le n} \deg(v_i, G_n)$$

and let

$$\gamma = \frac{1-p}{2-p}.$$

Then there exists a non-degenerate, strictly positive random variable Δ^* *such that*

$$n^{-\gamma}\deg_{\max}(G_n)) \to \Delta^* \qquad \text{with probability 1 as } n \to \infty$$

- Maximal non-superstar degree = $\Theta(n^{\gamma})$

Realized Degree Distribution in the Superstar Model

Theorem

Let $f(k, G_n)$ be the realized degree distribution of G_n under the Superstar model,

$$f(k, G_n) = n^{-1} \left|\{1 \le j \le n : \deg(v_j, G_n) = k\}\right|$$

and introduce the superstar model scaling constant

$$f_{SM}(k,p) = \frac{2-p}{1-p}(k-1)! \prod_{i=1}^{k} \left(i + \frac{2-p}{1-p}\right)^{-1}.$$

We then have

$$f(k, G_n) \to f_{SM}(k,p) \qquad \text{with probability 1 as } n \to \infty$$

- The degree distribution scales like $k^{-\beta}$, where $\beta = 3 + p/(1-p)$
- This contrasts with the preferential attachment model which scales like k^{-3}

Height result

Theorem

Let $W(\cdot)$ be the Lambert special function with $W(1/e) \approx 0.2784$. Then with probability one we have

$$\lim_{n\to\infty} \frac{1}{\log n}\mathcal{H}(G_n) = \frac{1-p}{W(1/e)(2-p)}.$$

Superstar Model vs Preferential Attachment

Model	Superstar Model	Preferential Attachment
Superstar Degree	$\Theta(n)$	NA
Maximal non-superstar degree exponent	$\frac{1-p}{2-p}$	$\frac{1}{2}$
Degree distribution power-law exponent	$3+\frac{p}{1-p}$	3

Superstar Model Predictions

- Use actual data to fit the superstar degree and predict the degree distribution
- Consider the observed degree distribution for each empirical retweet graph:

$$f(k, G_n) = n^{-1} \left|\{1 \le j \le n : \deg(v_j, G_n) = k\}\right|$$

- Consider the theoretical asymptotic degree distribution under the Superstar Model

$$f_{SM}(k,p) = \frac{2-p}{1-p}(k-1)! \prod_{i=1}^{k} \left(i + \frac{2-p}{1-p}\right)^{-1}.$$

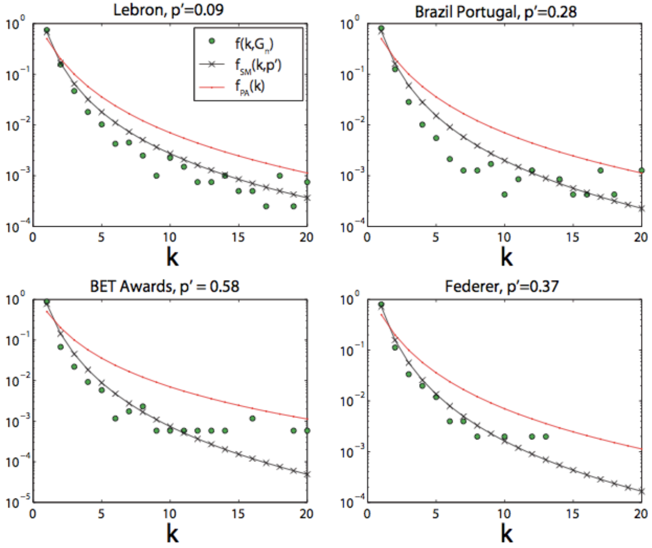
- Bottom Line: We get a nice fit “observed vs predicted”

$$f(k, G_n) \approx f_{SM}(k, \hat{p}) \qquad \text{where} \qquad \hat{p} = \frac{\text{observed superstar degree}}{n}$$

- Comparison: Preferential Attachment always predicts...

$$f_{PA}(k) = \frac{4}{k(k+1)(k+2)}$$

Degree distribution



The Superstar Model and the Realized Degree Distribution: Bottom Line

- The Superstar Model implies a mathematical link between the superstar degree and the degree distribution of the non-superstars.
- When we look at Twitter data for actual events, we see (1) a superstar and (2) a degree distribution of non-superstars that is more compatible with the superstar model than with the preferential attachment model.
- The first property was “baked” into our model, but the second was not. It’s an honest discovery.
- Next: How Can one Analyze the Superstar Model?

Basic Link: Branching Processes

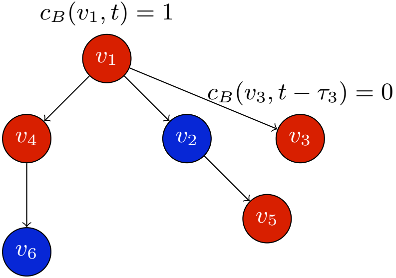
- Proto-Idea: Branching processes have a natural role almost anytime one considers a stochastically evolving tree.
- More Concrete Observation: If the birth rates depend on the number of children, the arithmetic of the Poisson process relates nicely to the arithmetic of preferential attachment.
- Creating the Superstar: Yule processes don't come with a superstar. Still, not terribly hard to move to multi-type branching processes. In a world with multiple types, you have the possibility of doing some surgery that let you build a super star.
- Realistic Expectations: The paper is a dense 29 pages.
- News You Can Use? One can see the benefits of using multi-type branching processes. One can see that the connection between the Yule process and preferential attachment is natural.

Introduction of a Special Branching Process

- Two types of vertices: red and blue
- Each vertex gives birth to vertices according to a non-homogeneous Poisson process that has rate proportional to (1+ number of blue children)

$$c_B(v,t) = \text{number of blue children of } v \text{ at } t \text{ time units after the birth of v}$$

- At birth vertex is painted red with probability p and painted blue with probability $1-p$

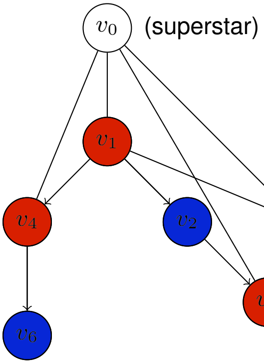


$$\begin{aligned} \mathcal{F}(t) &= \text{Branching process at time } t \\ \tau_n &= \inf\{t : |\mathcal{F}(t)| = n\} \end{aligned}$$

Surgery: From BP Model to Superstar Model

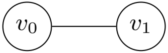
- Add an exogenous superstar vertex v_0 to the vertex set
- For each red vertex remove the edge from parent and create an undirected edge to the superstar vertex v_0
- With the surgery done, all edges are made undirected and all colors are erased

$\mathcal{F}(\tau_6)$



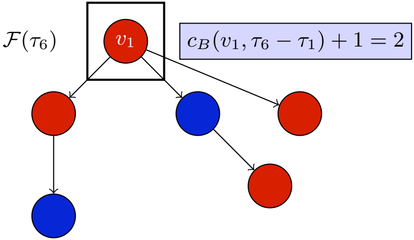
Relating the BP Construction with the Superstar Model

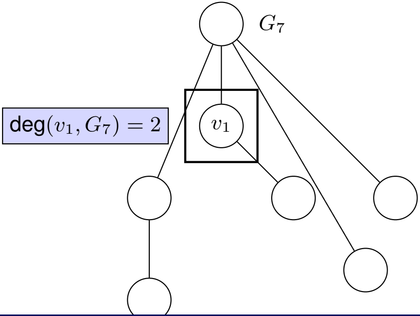
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}
- Base case: $S(\tau_1) = G_2$



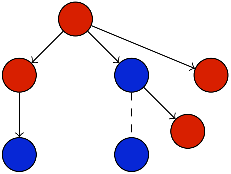
- Need to show that $S(\tau_n)$ and G_{n+1} have same probabilistic evolution
- Superstar: probability of joining superstar = probability of red vertex being born = p
- Same probability for S and G
- Non-superstars: degree of vertex = number of blue children + 1

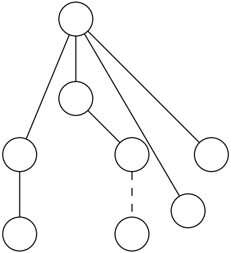
$$\deg(v_k, G_{n+1}) = c_B(v_k, \tau_n - \tau_k) + 1$$





Further Linking of the BP Model and the Superstar Model





$$\mathbb{P}\left(v_n \text{ joins } v_k | G_n\right) = \mathbb{P}\left(v_n \text{ is blue and born to } v_k | \mathcal{F}(\tau_{n-1})\right)$$

$$\begin{aligned}\mathbb{P}\left(v_n \text{ joins } v_k | G_n\right) &= (1-p)\frac{\deg(v_k, G_n)}{\sum_{v_j \in G_n \setminus v_0} \deg(v_j, G_n)} \\ &= (1-p)\frac{\deg(v_k, G_n)}{2(n-1) - \deg(v_0, G_n)}\end{aligned}$$

$$\mathbb{P}\left(v_n \text{ is blue and born to } v_k | \mathcal{F}(\tau_{n-1})\right) = (1-p)\frac{c_B(v_k, \tau_n - \tau_k) + 1}{\sum_{v_k \in \mathcal{F}(\tau_{n-1})} c_B(v_k, \tau_n - \tau_k) + 1}$$

Dynamic random graphs

- Lots of interesting questions
- Understanding what happens for general **unbounded size** rules such as product rule (*explosive percolation*).
- Small variants of standard models turn out to be technically much more challenging, requiring the development of new machinery.
- For the superstar model, a simple tweak gave much better fit to the data (one parameter p).

Thank you for your attention.