Lecture 2: Dynamic network models
Probabilistic and statistical methods for networks
Berlin Bath summer school for young researchers

Shankar Bhamidi

Department of Statistics and Operations Research
University of North Carolina

August, 2017

Motivation

Preferential attachment

- Last few years: enormous interest in formulating models to "explain" real-world networks (e.g. network of webpages, the Internet, social networks, gene regulatory networks etc).

Motivation

Preferential attachment

- Last few years: enormous interest in formulating models to "explain" real-world networks (e.g. network of webpages, the Internet, social networks, gene regulatory networks etc).
- One *central* model: preferential attachment. Amongst the first models for which explicit power-law degree distribution derived.

Motivation

Preferential attachment

- Last few years: enormous interest in formulating models to "explain" real-world networks (e.g. network of webpages, the Internet, social networks, gene regulatory networks etc).
- One *central* model: preferential attachment. Amongst the first models for which explicit power-law degree distribution derived.

Math world

- Main thrust: asymptotic information on the degree distribution.

Motivation

Preferential attachment

- Last few years: enormous interest in formulating models to "explain" real-world networks (e.g. network of webpages, the Internet, social networks, gene regulatory networks etc).
- One *central* model: preferential attachment. Amongst the first models for which explicit power-law degree distribution derived.

Math world

- Main thrust: asymptotic information on the degree distribution.
- Largely based on *recursions and concentration inequalities*.

Motivation

Preferential attachment

- Last few years: enormous interest in formulating models to "explain" real-world networks (e.g. network of webpages, the Internet, social networks, gene regulatory networks etc).
- One *central* model: preferential attachment. Amongst the first models for which explicit power-law degree distribution derived.

Math world

- Main thrust: asymptotic information on the degree distribution.
- Largely based on *recursions and concentration inequalities*.
- *What if we wanted asymptotics for Global characteristics e.g. spectral distribution of adjacency matrix?*

Motivation

Preferential attachment

- Last few years: enormous interest in formulating models to "explain" real-world networks (e.g. network of webpages, the Internet, social networks, gene regulatory networks etc).
- One *central* model: preferential attachment. Amongst the first models for which explicit power-law degree distribution derived.

Math world

- Main thrust: asymptotic information on the degree distribution.
- Largely based on *recursions and concentration inequalities*.
- *What if we wanted asymptotics for Global characteristics e.g. spectral distribution of adjacency matrix?*
- *How to analyze variants such as limited choice or non-local preferential attachment. Analysis? Performance in practice?*

Outline of the talk

1. Preferential attachment model

Outline of the talk

1. Preferential attachment model
2. Continuous time branching processes

Outline of the talk

1. Preferential attachment model
2. Continuous time branching processes
3. Local weak convergence

Outline of the talk

1. Preferential attachment model
2. Continuous time branching processes
3. Local weak convergence
4. Random adjacency matrices and Spectra (Arnab Sen, Steve Evans, SB)

Outline of the talk

1. Preferential attachment model
2. Continuous time branching processes
3. Local weak convergence
4. Random adjacency matrices and Spectra (Arnab Sen, Steve Evans, SB)
5. Variants I: Superstar model (Mike Steel, Tauhid Zaman, SB)

Outline of the talk

1. Preferential attachment model
2. Continuous time branching processes
3. Local weak convergence
4. Random adjacency matrices and Spectra (Arnab Sen, Steve Evans, SB)
5. Variants I: Superstar model (Mike Steel, Tauhid Zaman, SB)
6. Variants II: Preferential attachment with choice (Omer Angel, Robin Pemantle, SB)

Basic model of growing trees

Setting

- At time 2 start with two vertices labelled with $[2] := \{1,2\}$ connected by single directed edge $1 \rightsquigarrow 2$.

Basic model of growing trees

Setting

- At time 2 start with two vertices labelled with $[2] := \{1,2\}$ connected by single directed edge $1 \rightsquigarrow 2$.
- Let $\mathcal{T}_n = (\mathcal{V}_n := [n], \mathcal{E}_n)$ be the tree at time n.

Basic model of growing trees

Setting

- At time 2 start with two vertices labelled with $[2] := \{1,2\}$ connected by single directed edge $1 \rightsquigarrow 2$.
- Let $\mathcal{T}_n = (\mathcal{V}_n := [n], \mathcal{E}_n)$ be the tree at time n.
- *Attractiveness function:* Assume we are given a (possibly **random**) function $f(\cdot, n) : V_n \to \mathbb{R}^+$.

Basic model of growing trees

Setting

- At time 2 start with two vertices labelled with $[2] := \{1,2\}$ connected by single directed edge $1 \rightsquigarrow 2$.
- Let $\mathcal{T}_n = (\mathcal{V}_n := [n], \mathcal{E}_n)$ be the tree at time n.
- *Attractiveness function:* Assume we are given a (possibly **random**) function $f(\cdot, n) : V_n \to \mathbb{R}^+$.

Dynamics

- At time $n+1$, new node labelled $n+1$ enters system.
- Node $n+1$ attaches to node in $\mathcal{T}_n$ with probability proportional to $f(\cdot, n)$.

Basic model of growing trees

Setting

- At time 2 start with two vertices labelled with $[2] := \{1,2\}$ connected by single directed edge $1 \rightsquigarrow 2$.
- Let $\mathcal{T}_n = (\mathcal{V}_n := [n], \mathcal{E}_n)$ be the tree at time n.
- *Attractiveness function:* Assume we are given a (possibly **random**) function $f(\cdot, n) : V_n \to \mathbb{R}^+$.

Dynamics

- At time $n+1$, new node labelled $n+1$ enters system.
- Node $n+1$ attaches to node in $\mathcal{T}_n$ with probability proportional to $f(\cdot, n)$.
- Most examples we consider: $f(v,n) = f(D(v,n))$ where $f : \{0,1,\ldots\} \to (0,\infty)$ is a fixed function.

Basic model of growing trees

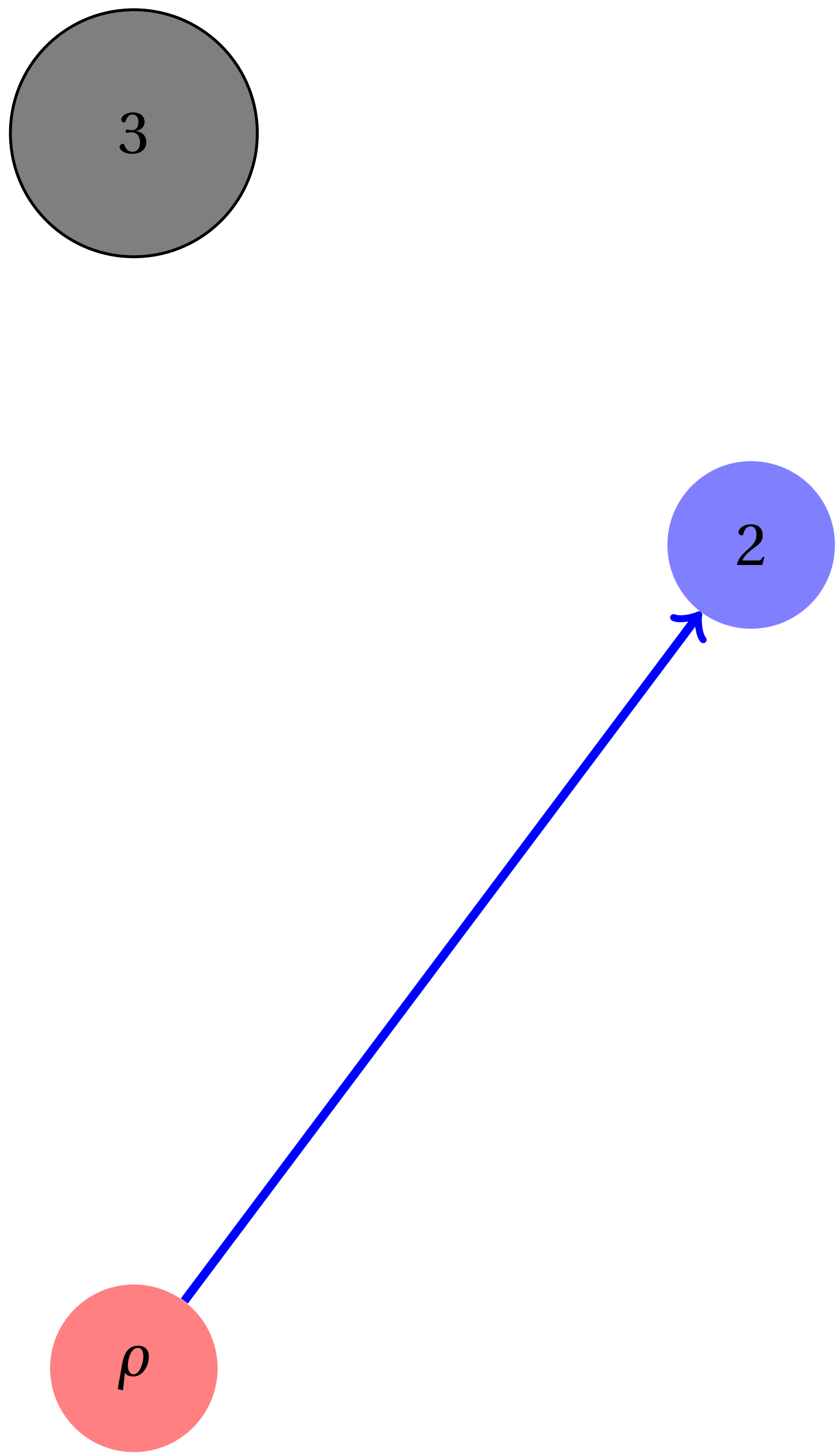
Setting

- At time 2 start with two vertices labelled with $[2] := \{1,2\}$ connected by single directed edge $1 \rightsquigarrow 2$.
- Let $\mathcal{T}_n = (\mathcal{V}_n := [n], \mathcal{E}_n)$ be the tree at time n.
- *Attractiveness function:* Assume we are given a (possibly **random**) function $f(\cdot, n) : V_n \to \mathbb{R}^+$.

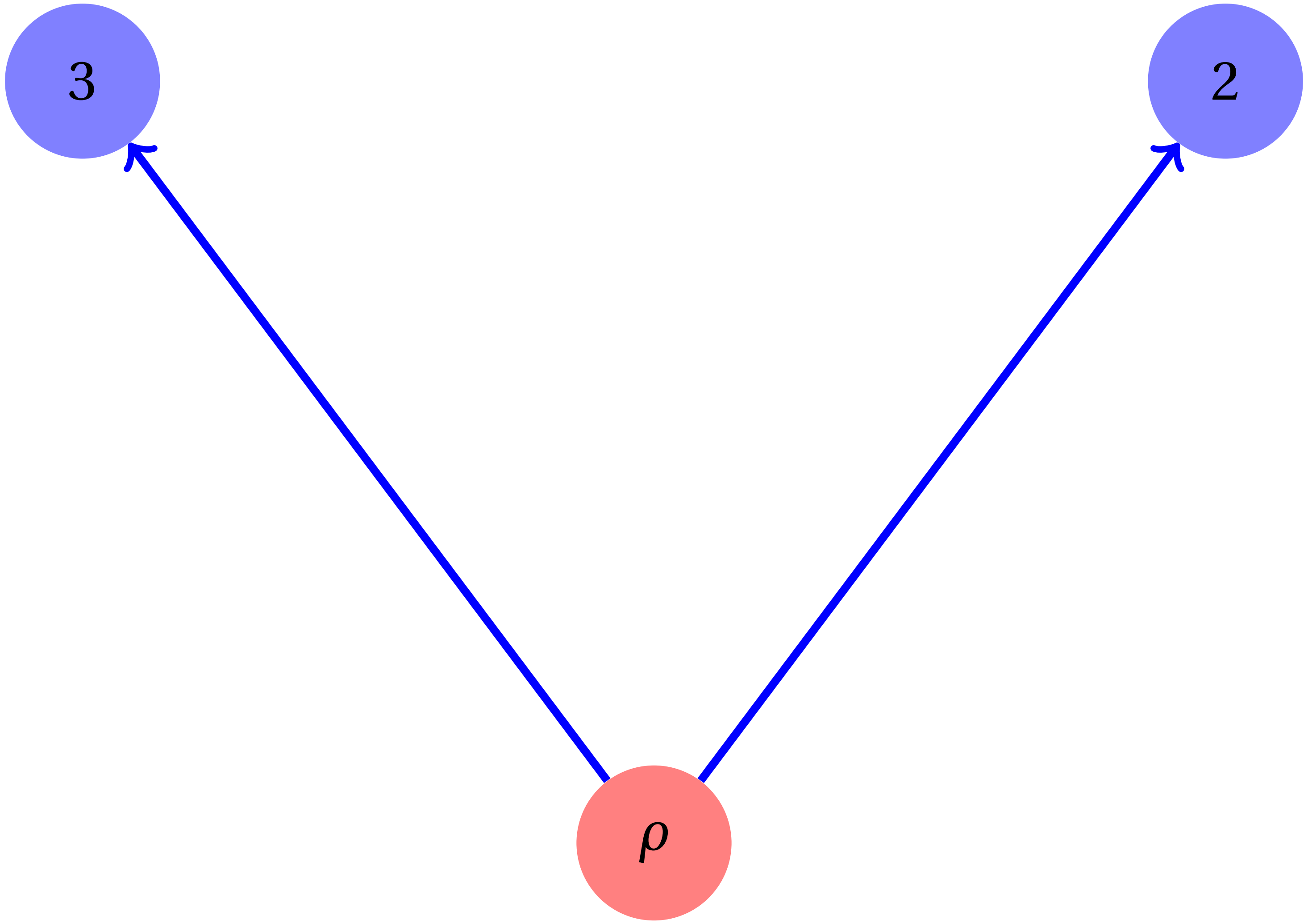
Dynamics

- At time $n+1$, new node labelled $n+1$ enters system.
- Node $n+1$ attaches to node in $\mathcal{T}_n$ with probability proportional to $f(\cdot, n)$.
- Most examples we consider: $f(v,n) = f(D(v,n))$ where $f : \{0,1,\ldots\} \to (0,\infty)$ is a fixed function.
- $D(v,n) =$ *out-degree* of node v at time n.

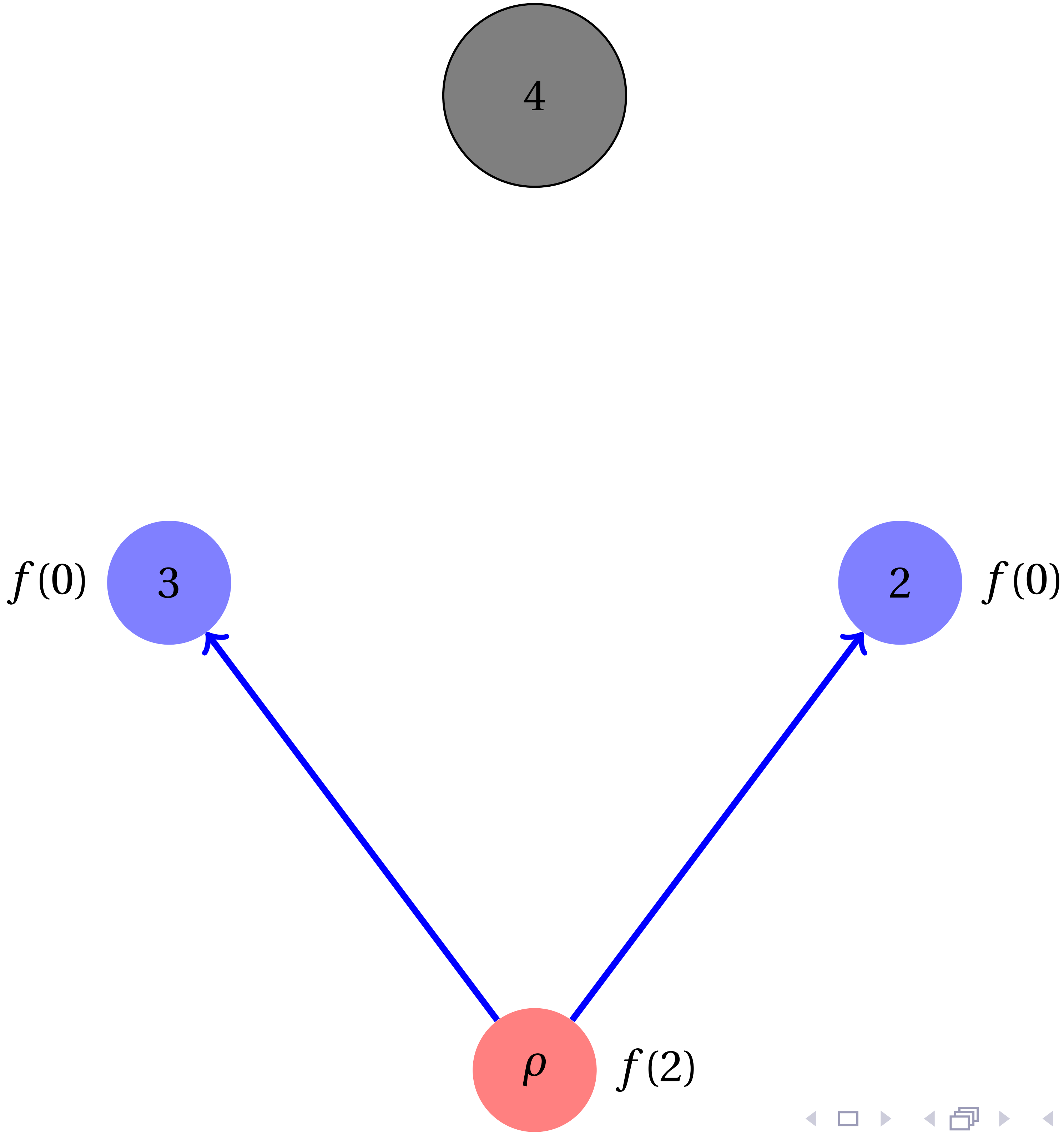
Preferential attachment: Example $\mathscr{T}_3$



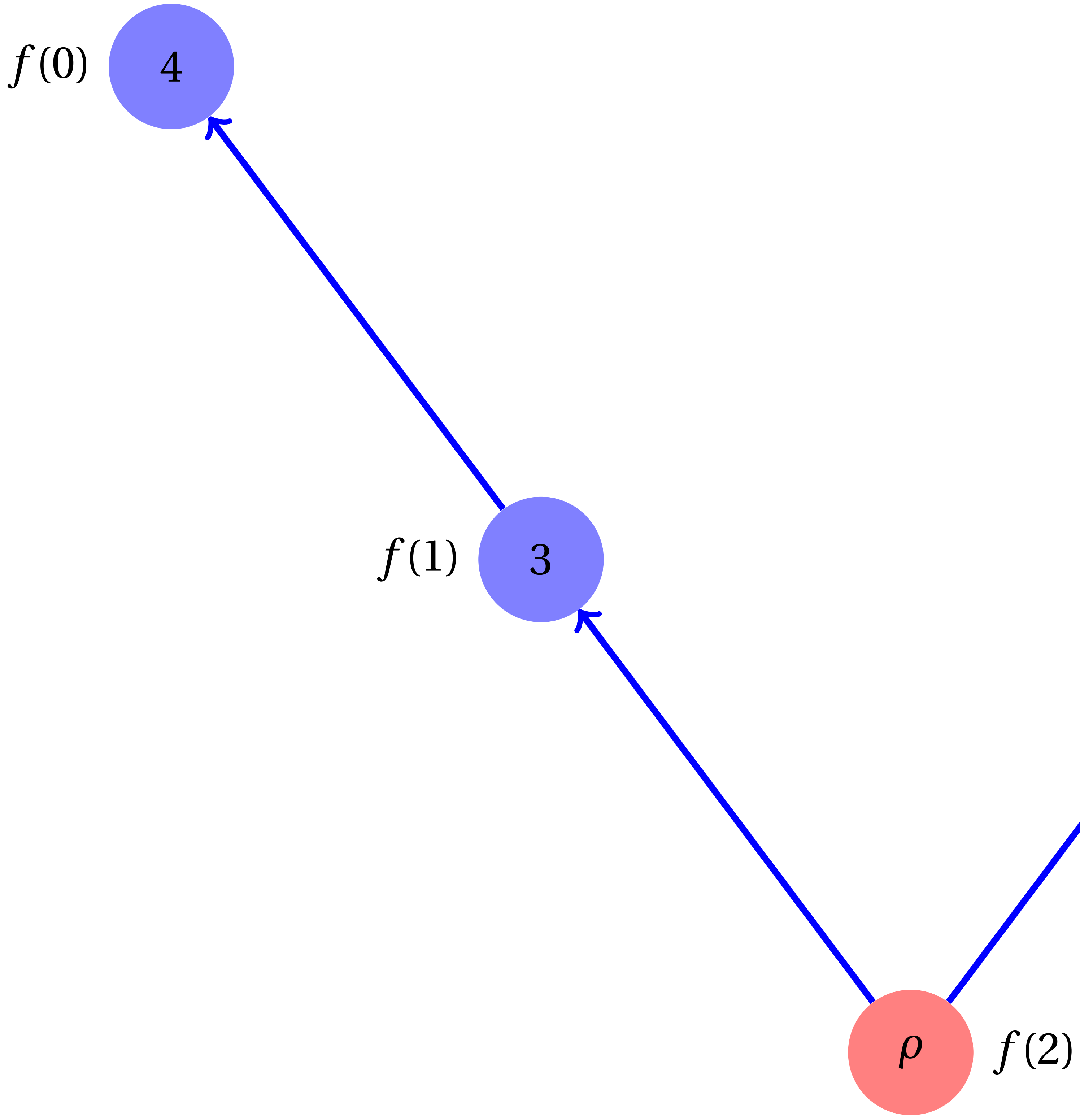
Preferential attachment: Example



Preferential attachment: Example



Preferential attachment: Example $\mathscr{T}_4$



Examples of attractiveness functions

Attachment trees

1. **YSBA** *(Yule-Simon-Barabasi-Albert)* **model:** $f(k) = k + 1$

Examples of attractiveness functions

Attachment trees

1. **YSBA** *(Yule-Simon-Barabasi-Albert)* **model:** $f(k) = k + 1$

Examples of attractiveness functions

Attachment trees

1. **YSBA** *(Yule-Simon-Barabasi-Albert)* **model:** $f(k) = k + 1$
2. **Linear Preferential attachment model:** $f(k) = k + 1 + \alpha, \ \alpha > 0$

Examples of attractiveness functions

Attachment trees

1. **YSBA** *(Yule-Simon-Barabasi-Albert)* **model:** $f(k) = k + 1$
2. **Linear Preferential attachment model:** $f(k) = k + 1 + \alpha, \ \alpha > 0$

Examples of attractiveness functions

Attachment trees

1. **YSBA** *(Yule-Simon-Barabasi-Albert)* **model:** $f(k) = k+1$
2. **Linear Preferential attachment model:** $f(k) = k+1+\alpha,\ \alpha > 0$
3. **Random fitness models** Every new vertex v given a fitness $f_v \sim \nu$ (independent across vertices).
   (a) *Multiplicative fitness:* $f(k) = f_v(k+1)$.
   (b) *Additive fitness:* $f(k) = k+1+f_v$.

Examples of attractiveness functions

Attachment trees

1. **YSBA** *(Yule-Simon-Barabasi-Albert)* **model:** $f(k) = k + 1$
2. **Linear Preferential attachment model:** $f(k) = k + 1 + \alpha$, $\alpha > 0$
3. **Random fitness models** Every new vertex v given a fitness $f_v \sim \nu$ (independent across vertices).
   (a) *Multiplicative fitness:* $f(k) = f_v(k+1)$.
   (b) *Additive fitness:* $f(k) = k + 1 + f_v$.

Examples of attractiveness functions

Attachment trees

1. **YSBA** *(Yule-Simon-Barabasi-Albert)* **model:** $f(k) = k + 1$
2. **Linear Preferential attachment model:** $f(k) = k + 1 + \alpha,\ \alpha > 0$
3. **Random fitness models** Every new vertex v given a fitness $f_v \sim \nu$ (independent across vertices).
   (a) *Multiplicative fitness:* $f(k) = f_v(k+1)$.
   (b) *Additive fitness:* $f(k) = k + 1 + f_v$.
4. **Sublinear Pref Attachment:** $f(k) = (k+1)^{\alpha},\ 0 < \alpha < 1$

Main math idea

Simple idea [Karlin-Athreya]

- Suppose we have vertex set $\{1, 2, \ldots, m\}$ with associated (*strictly positive*) weights $\{d_1, d_2, \ldots d_m\}$.

Main math idea

Simple idea [Karlin-Athreya]

- Suppose we have vertex set $\{1,2,\ldots,m\}$ with associated (*strictly positive*) weights $\{d_1,d_2,\ldots d_m\}$.
- Want to selected vertex i with probability proportional to d_i.

Main math idea

Simple idea [Karlin-Athreya]

- Suppose we have vertex set $\{1, 2, \ldots, m\}$ with associated (*strictly positive*) weights $\{d_1, d_2, \ldots d_m\}$.
- Want to selected vertex i with probability proportional to d_i.
- Elegant algorithm: Let X_i be independent *rate* d_i exponential random variables.

Main math idea

Simple idea [Karlin-Athreya]

- Suppose we have vertex set $\{1, 2, \ldots, m\}$ with associated (*strictly positive*) weights $\{d_1, d_2, \ldots d_m\}$.
- Want to selected vertex i with probability proportional to d_i.
- Elegant algorithm: Let X_i be independent *rate d_i* exponential random variables.
- Let J be index

$$X_J = \min_{1 \le i \le m} X_i.$$

Main math idea

Simple idea [Karlin-Athreya]

- Suppose we have vertex set $\{1, 2, \ldots, m\}$ with associated (*strictly positive*) weights $\{d_1, d_2, \ldots d_m\}$.
- Want to selected vertex i with probability proportional to d_i.
- Elegant algorithm: Let X_i be independent *rate* d_i exponential random variables.
- Let J be index
$$X_J = \min_{1 \le i \le m} X_i.$$
- *Then* $\mathbb{P}(J = i) \propto d_i$.

Outline

Continuous time construction

Point process corresponding to attractiveness function f

- $\mathscr{P}$ is Markov pure birth process with rate description

$$\mathbb{P}(\mathscr{P}(t, t+dt] = 1 | \mathscr{P}(t) = k) = f(k)dt.$$

Continuous time construction

Point process corresponding to attractiveness function f

- $\mathscr{P}$ is Markov pure birth process with rate description

$$\mathbb{P}(\mathscr{P}(t, t+dt] = 1 | \mathscr{P}(t) = k) = f(k)dt.$$

- For example, for $f(k) = k+1$ (usual preferential attachment model) we get the Yule process.

Continuous time construction

Point process corresponding to attractiveness function f

- $\mathscr{P}$ is Markov pure birth process with rate description

$$\mathbb{P}(\mathscr{P}(t, t+dt] = 1 | \mathscr{P}(t) = k) = f(k)dt.$$

- For example, for $f(k) = k+1$ (usual preferential attachment model) we get the Yule process.

Continuous time branching process $\mathscr{F}(t)$

1. Start with a single node at time 0 giving birth to children at times of $\mathscr{P}$.
2. Each node born behaves in the same manner (has it's own independent point process of births).

Continuous time construction

Point process corresponding to attractiveness function f

- $\mathscr{P}$ is Markov pure birth process with rate description

$$\mathbb{P}(\mathscr{P}(t, t+dt] = 1 | \mathscr{P}(t) = k) = f(k) dt.$$

- For example, for $f(k) = k+1$ (usual preferential attachment model) we get the Yule process.

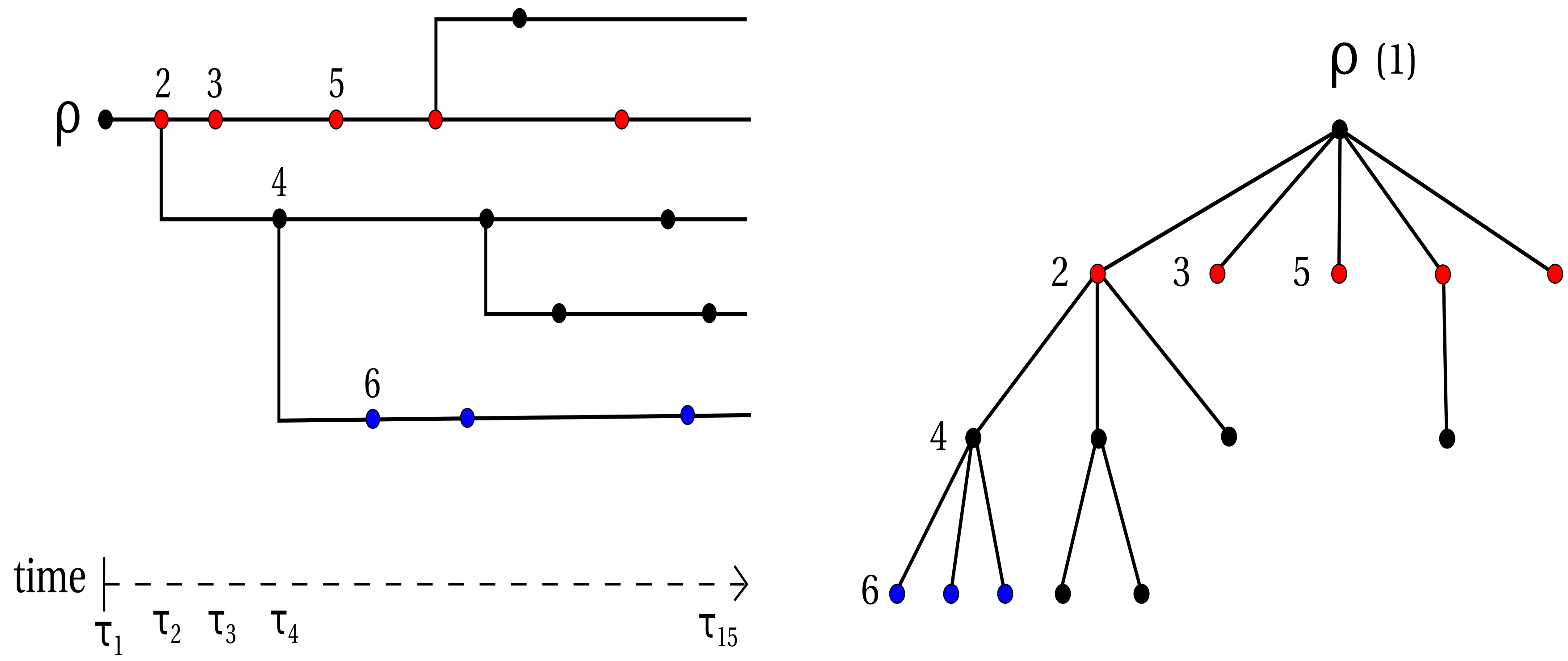
Continuous time branching process $\mathscr{F}(t)$

1. Start with a single node at time 0 giving birth to children at times of $\mathscr{P}$.
2. Each node born behaves in the same manner (has it's own independent point process of births).

Key connection

$\tau_n = \inf\{t : \mathscr{F}(t) = n\}$ *then* $\mathscr{F}(\tau_n) \stackrel{d}{=} \mathcal{T}_n^f$.

Continuous time and discrete time in pictures



Branching process theory

Asymptotics

Conjectured by Euler. Developed by Jagers and Nerman.

- Processes grow exponentially: $|\mathscr{F}(t)| \sim e^{\lambda t}$

Branching process theory

Asymptotics

Conjectured by Euler. Developed by Jagers and Nerman.

- Processes grow exponentially: $|\mathscr{F}(t)| \sim e^{\lambda t}$
- Here λ is a very important characteristic : called the *Malthusian rate of growth*

Branching process theory

Asymptotics

Conjectured by Euler. Developed by Jagers and Nerman.

- Processes grow exponentially: $|\mathscr{F}(t)| \sim e^{\lambda t}$
- Here λ is a very important characteristic : called the *Malthusian rate of growth*
- Given by the formula:

$$\mathbb{E}(\mathscr{P}(T_\lambda)) = 1,$$

  where $T_\lambda \sim \exp(\lambda)$ independent of $\mathscr{P}$.

Exact result

Under technical conditions $(\mathbb{E}(\mathscr{P}(T_\lambda), \log^+ \mathscr{P}(T_\lambda)) < \infty)$:

- $|\mathscr{F}(t)|e^{-\lambda t} \xrightarrow{a.s.} W$.
- In our settings: $W > 0$ a.s.
- Bottom line:

$$\tau_n \sim \frac{1}{\lambda} \log n \pm O_P(1)$$

Case Study: Usual preferential attachment

$f(k) = k + 1$

- Offspring distribution: $\mathscr{P}(\cdot) =$ Yule process

Case Study: Usual preferential attachment

$f(k) = k+1$

- Offspring distribution: $\mathscr{P}(\cdot) =$ Yule process
- Malthusian rate of growth: $\lambda = 2$.

Case Study: Usual preferential attachment

$f(k) = k+1$

- Offspring distribution: $\mathscr{P}(\cdot) =$ Yule process
- Malthusian rate of growth: $\lambda = 2$.

Root degree asymptotics

- Degree of the root $= \mathscr{P}_\rho(\tau_n)$

Case Study: Usual preferential attachment

$f(k) = k + 1$

- Offspring distribution: $\mathscr{P}(\cdot) = \text{Yule process}$
- Malthusian rate of growth: $\lambda = 2$.

Root degree asymptotics

- Degree of the root $= \mathscr{P}_\rho(\tau_n)$
- $\tau_n \sim \frac{1}{2}\log n \pm O_P(1)$. Yule process also grows exponentially: $\mathscr{P}(t) \sim e^t$

Case Study: Usual preferential attachment

$f(k) = k+1$

- Offspring distribution: $\mathscr{P}(\cdot) =$ Yule process
- Malthusian rate of growth: $\lambda = 2$.

Root degree asymptotics

- Degree of the root $= \mathscr{P}_\rho(\tau_n)$
- $\tau_n \sim \frac{1}{2}\log n \pm O_P(1)$. Yule process also grows exponentially: $\mathscr{P}(t) \sim e^t$
- $\deg_n(\rho) = \mathscr{P}(\frac{1}{2}\log n + O_P(1)) \sim O_P(e^{\frac{1}{2}\log n}) = O_P(\sqrt{n})$

Case Study: Usual preferential attachment

$f(k) = k + 1$

- Offspring distribution: $\mathscr{P}(\cdot) =$ Yule process
- Malthusian rate of growth: $\lambda = 2$.

Root degree asymptotics

- Degree of the root $= \mathscr{P}_\rho(\tau_n)$
- $\tau_n \sim \frac{1}{2}\log n \pm O_P(1)$. Yule process also grows exponentially: $\mathscr{P}(t) \sim e^t$
- $\deg_n(\rho) = \mathscr{P}(\frac{1}{2}\log n + O_P(1)) \sim O_P(e^{\frac{1}{2}\log n}) = O_P(\sqrt{n})$
- More refined analysis (Mori/Pekoz+Rollin+Ross) gives

$$\frac{\deg_n(\rho)}{\sqrt{n}} \xrightarrow{a.s.} Z \qquad Z \text{ has explicit recursive construction.}$$

Maximal degree

Basic heuristics for models with "heavy tails"

- Due to exponential growth of the models in the natural "time scale", maximal degree occurs in a finite neighborhood of the root.

Maximal degree

Basic heuristics for models with "heavy tails"

- Due to exponential growth of the models in the natural "time scale", maximal degree occurs in a finite neighborhood of the root.
- For usual preferential attachment model, using explicit distributional properties of Yule process easy to conclude, for any given $\epsilon > 0 \; \exists \; K_\epsilon$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{\max\deg_n}{\sqrt{n}} > K_\epsilon\right) < \epsilon$$

Maximal degree

Basic heuristics for models with "heavy tails"

- Due to exponential growth of the models in the natural "time scale", maximal degree occurs in a finite neighborhood of the root.
- For usual preferential attachment model, using explicit distributional properties of Yule process easy to conclude, for any given $\epsilon > 0 \ \exists \ K_\epsilon$,

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{\max \deg_n}{\sqrt{n}} > K_\epsilon\right) < \epsilon$$

- With a bit more work, possible to deduce distributional convergence for the maximal degree.
- Example of interesting results: Sublinear pref attachment $f(k) = (k+1)^\alpha$

$$\frac{deg_n(\rho)}{(\log n)^{\frac{1}{1-\alpha}}} \xrightarrow{P} \left(\frac{1}{\theta(\alpha)}\right)^{\frac{1}{1-\alpha}}$$

Height asymptotics

$h_n =$ Furthest distance of a vertex from the root ρ in $\mathscr{T}_n$.

Height asymptotics

h_n = Furthest distance of a vertex from the root ρ in $\mathscr{T}_n$.

Kingman's result and Pittel's "proof from the book"

Let $B_k :=$ first time that an individual in the k^{th} generation (namely an individual at graph distance k from the root) is born.

Height asymptotics

h_n = Furthest distance of a vertex from the root ρ in $\mathscr{T}_n$.

Kingman's result and Pittel's "proof from the book"

Let $B_k :=$ first time that an individual in the k^{th} generation (namely an individual at graph distance k from the root) is born.

[Kingman]: There exists a (model dependent) limit constant γ such that:

$$\frac{B_k}{k} \xrightarrow{a.s.} \gamma_{\texttt{model}}$$

Height asymptotics

h_n = Furthest distance of a vertex from the root ρ in $\mathscr{T}_n$.

Kingman's result and Pittel's "proof from the book"

Let $B_k :=$ first time that an individual in the k^{th} generation (namely an individual at graph distance k from the root) is born.

[Kingman]: There exists a (model dependent) limit constant γ such that:

$$\frac{B_k}{k} \xrightarrow{a.s.} \gamma_{\mathtt{model}}$$

For us we have

$$B_{h_n} \leq T_n \leq B_{h_n+1}$$

Height asymptotics

h_n = Furthest distance of a vertex from the root ρ in $\mathscr{T}_n$.

Kingman's result and Pittel's "proof from the book"

Let $B_k :=$ first time that an individual in the k^{th} generation (namely an individual at graph distance k from the root) is born.

[Kingman]: There exists a (model dependent) limit constant γ such that:

$$\frac{B_k}{k} \xrightarrow{a.s.} \gamma_{\texttt{model}}$$

For us we have

$$B_{h_n} \leq T_n \leq B_{h_n+1}$$

Thus

$$\frac{B_{h_n}}{h_n} \leq \frac{\tau_n}{h_n} \leq \frac{B_{h_n+1}}{h_n}$$

Now use the fact [Pittel's argument] that

$$\frac{\tau_n}{\frac{1}{\lambda}\log n} \xrightarrow{a.s.} 1 \Rightarrow \frac{h_n}{\log n} \xrightarrow{P} C_{\texttt{model}}$$

Outline

1. Preferential attachment: Base model
   - Continuous time construction
   - Local weak convergence
2. Twitter event networks and the superstar model
   - Retweet Graph and Superstar Model
   - Main Results
   - Comparison with Preferential Attachment Model
   - Superstar Model: Tools for Analysis
3. Power of choice and random trees
4. Conclusion

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.
- *Infinite path* represents the "*path to the root*" .

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.
- *Infinite path* represents the "*path to the root*" .
- Immediately gives degree distribution asymptotics but gives much much more.

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.
- *Infinite path* represents the "*path to the root*" .
- Immediately gives degree distribution asymptotics but gives much much more.

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.
- *Infinite path* represents the "*path to the root*" .
- Immediately gives degree distribution asymptotics but gives much much more.

Starting point: Age of an individual

$$\mathbb{P}(\text{Age}(V_t) > 10 | \mathscr{F}(t)) = \frac{|\mathscr{F}(t-10)|}{|\mathscr{F}(t)|}$$

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.
- *Infinite path* represents the "*path to the root*" .
- Immediately gives degree distribution asymptotics but gives much much more.

Starting point: Age of an individual

$$\mathbb{P}(\text{Age}(V_t) > 10 | \mathscr{F}(t)) = \frac{|\mathscr{F}(t-10)|}{|\mathscr{F}(t)|} \sim \frac{We^{\lambda(t-10)}}{We^{\lambda t}}$$

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.
- *Infinite path* represents the "*path to the root*" .
- Immediately gives degree distribution asymptotics but gives much much more.

Starting point: Age of an individual

$$\mathbb{P}(\text{Age}(V_t) > 10 | \mathscr{F}(t)) = \frac{|\mathscr{F}(t-10)|}{|\mathscr{F}(t)|} \sim \frac{We^{\lambda(t-10)}}{We^{\lambda t}} = e^{-10\lambda} = \mathbb{P}(T_\lambda > 10)$$

Local asymptotics (Jagers+Nerman/Aldous)

Conceptual point

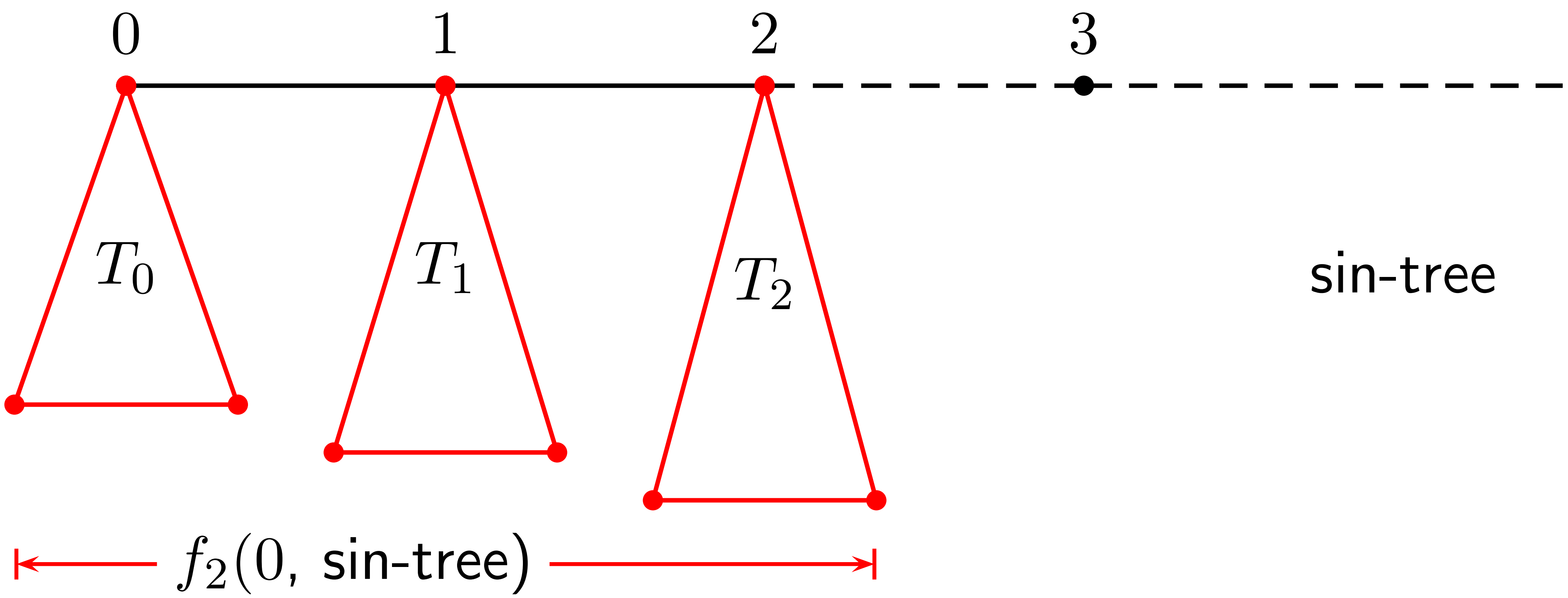
- Construction of infinite (locally finite) rooted trees with a single infinite path. (Trees with one end).
- Local neighborhood of a random node asymptotically looks like neighborhood of the *root* of appropriately constructed `sin`-tree.
- *Infinite path* represents the "*path to the root*" .
- Immediately gives degree distribution asymptotics but gives much much more.

Starting point: Age of an individual

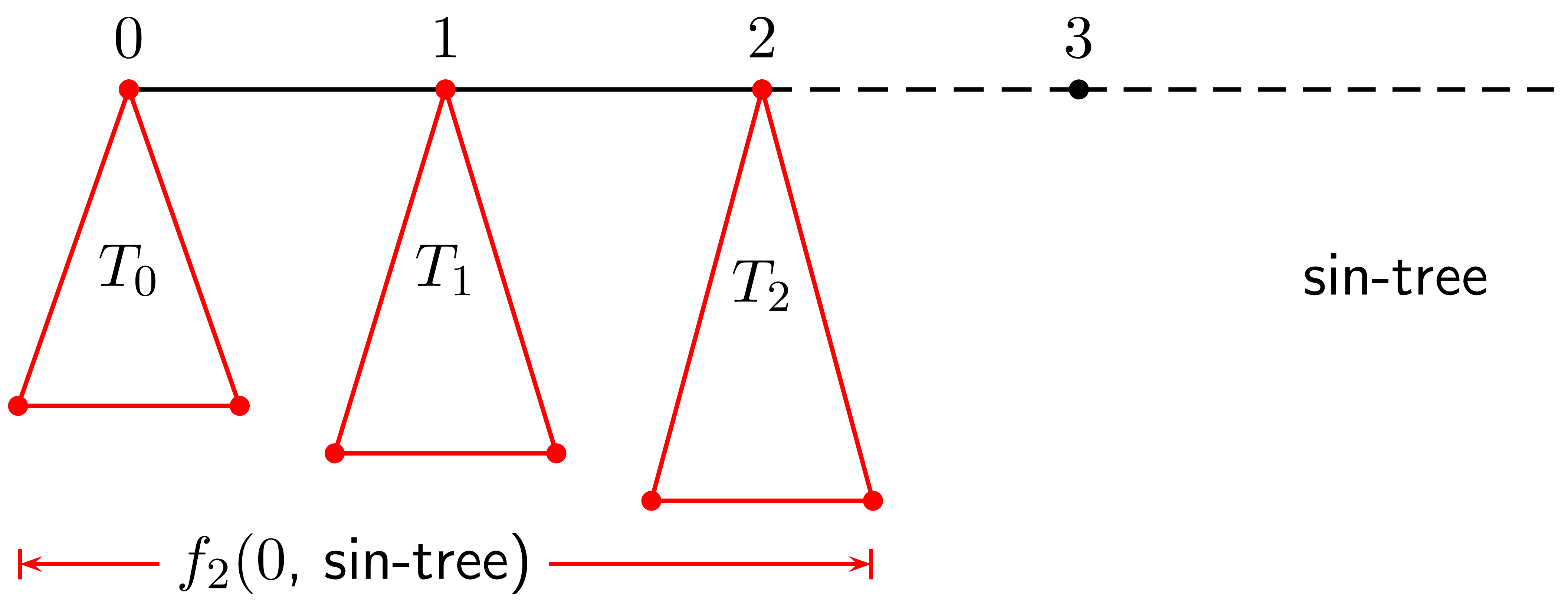
$$\mathbb{P}(\mathrm{Age}(V_t) > 10 | \mathscr{F}(t)) = \frac{|\mathscr{F}(t-10)|}{|\mathscr{F}(t)|} \sim \frac{We^{\lambda(t-10)}}{We^{\lambda t}} = e^{-10\lambda} = \mathbb{P}(T_\lambda > 10)$$

Suggests tree "below" random node looks like $\mathscr{F}(T_\lambda)$ i.e. branching process run for random exponential amount of time.

Sin-tree

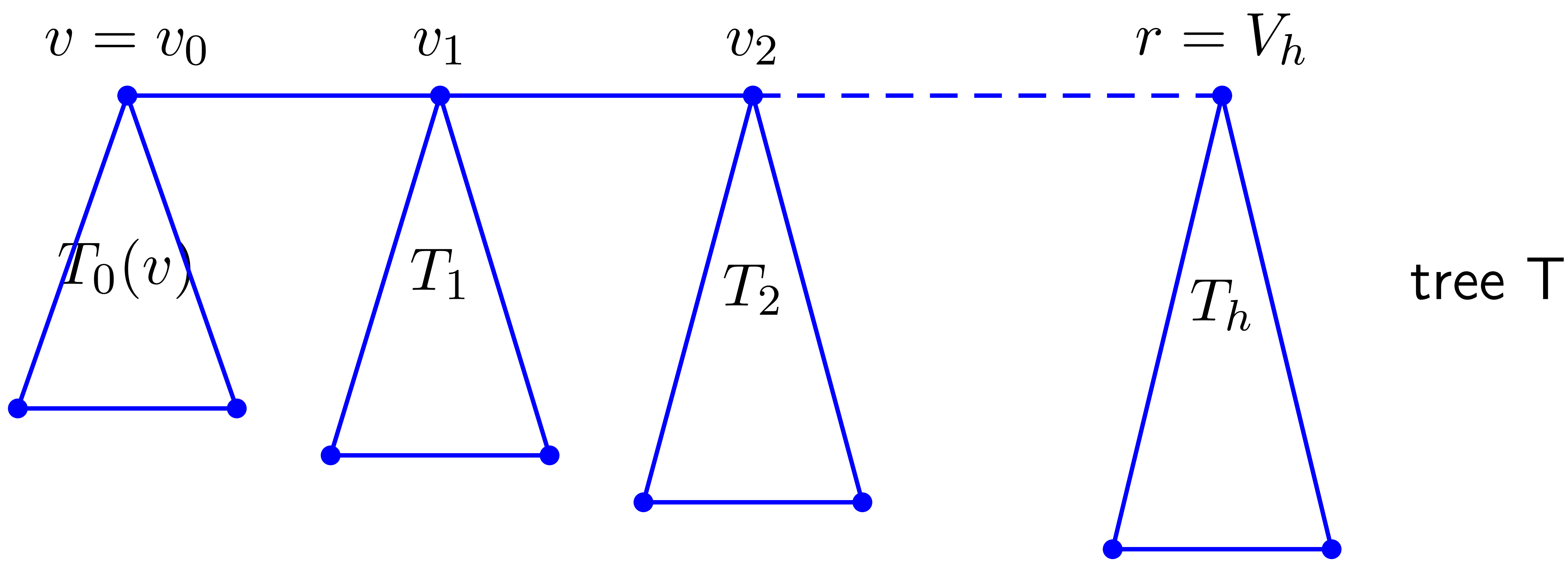


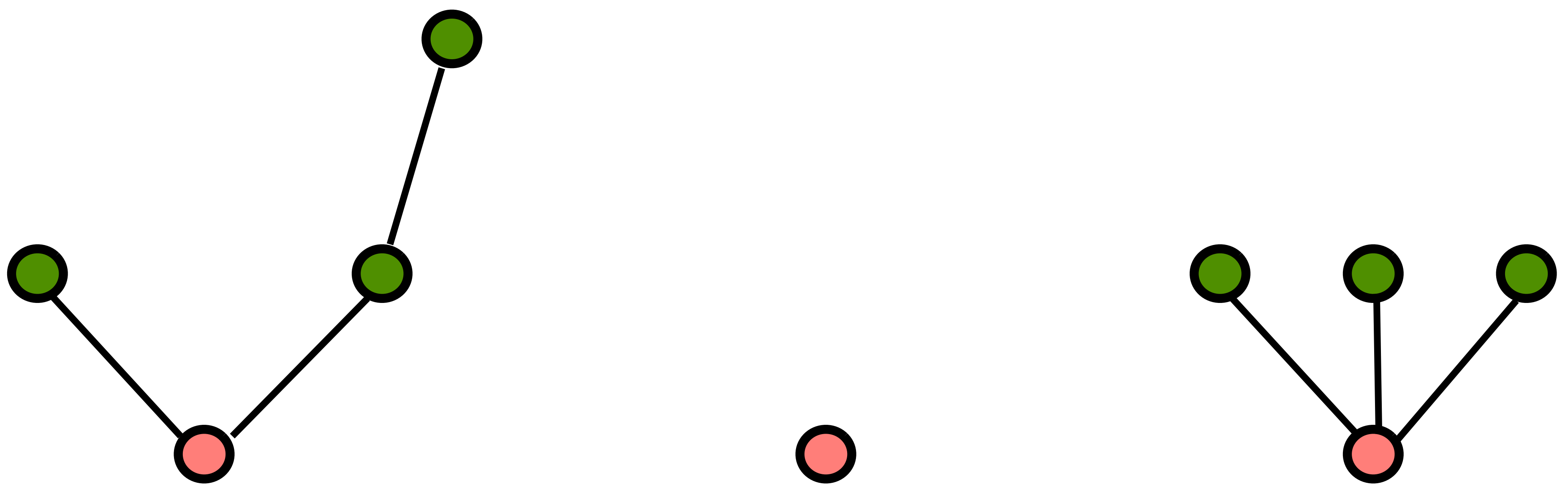
Sin-tree



Can think of random sin-trees

T is a tree with root r. Given a vertex v, there exists a unique path $v_0 = v, v_1, ..., v_h = r$ from v to the root.

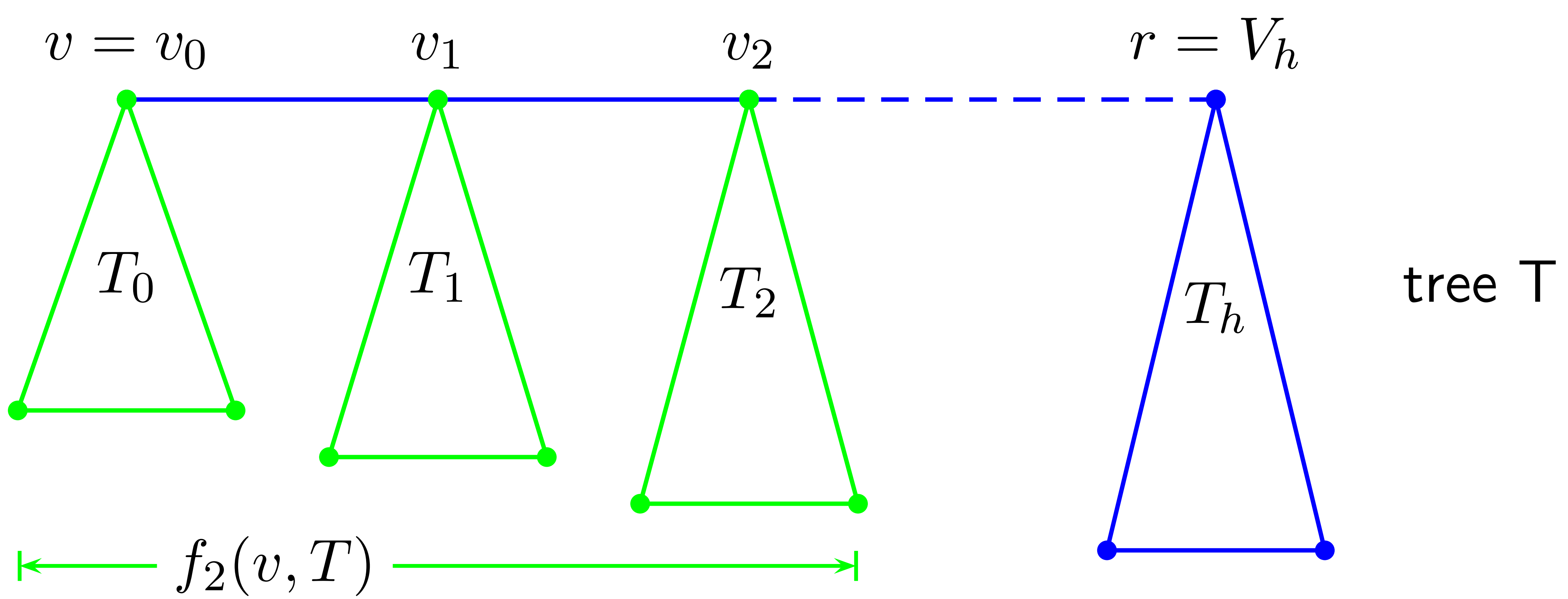




t_0 t_1 t_2

Decompose tree into a sequence of finite rooted subtrees or fringes $(T_0(v), T_1(v), T_2(v), \ldots)$. For each $k \geq 1$,

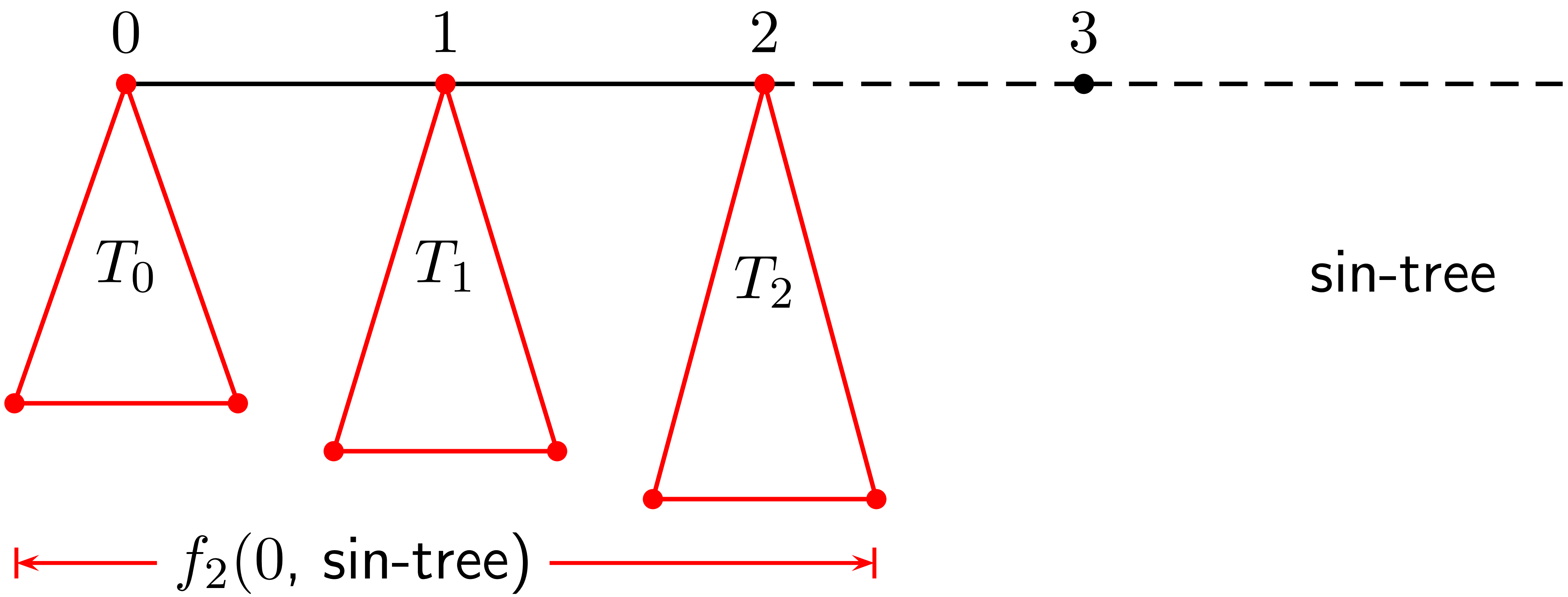
$$\frac{1}{n}\sum_{v\in T} 1(f_k(v,T) = (t_0, t_1, ..., t_k)) \xrightarrow{P} \mathbb{P}_\mu(f_k(0,\mathcal{T}) = (t_0, t_1, ..., t_k)).$$



Convergence in probability fringe sense

Decompose tree into a sequence of finite rooted subtrees or fringes $(T_0(v), T_1(v), T_2(v), \ldots)$. For each $k \geq 1$,

$$\frac{1}{n}\sum_{v\in T} 1(f_k(v,T) = (t_0, t_1, ..., t_k)) \xrightarrow{P} \mathbb{P}_\mu(f_k(0,\mathcal{T}) = (t_0, t_1, ..., t_k)).$$



sin-trees [Jagers+Nerman/Aldous]

Construction

$\mathscr{T}^{\mathtt{sin}}_{\alpha,\mu,\mathscr{P}}$: Random tree with single infinite path.

sin-trees [Jagers+Nerman/Aldous]

Construction

$\mathcal{T}^{\mathtt{sin}}_{\alpha,\mu,\mathscr{P}}$: Random tree with single infinite path.

- $X_0 \sim \exp(\alpha)$ and for $i \geq 1$, $X_i \sim \mu$. $S_n = \sum_0^n X_i$.
- Conditional on the sequence $(S_n)_{n\geq 0}$,
  1. $\mathscr{F}_{X_0}$: continuous time branching process driven by $\mathscr{P}$ observed up to time X_0.
  2. For $n \geq 1$ let $\mathscr{F}_{S_n, S_{n-1}}$: continuous time branching process observed up to time S_n;

sin-trees [Jagers+Nerman/Aldous]

Construction

$\mathcal{T}^{\mathtt{sin}}_{\alpha,\mu,\mathscr{P}}$: Random tree with single infinite path.

- $X_0 \sim \exp(\alpha)$ and for $i \geq 1$, $X_i \sim \mu$. $S_n = \sum_0^n X_i$.
- Conditional on the sequence $(S_n)_{n\geq 0}$,
  1. $\mathscr{F}_{X_0}$: continuous time branching process driven by $\mathscr{P}$ observed up to time X_0.
  2. For $n \geq 1$ let $\mathscr{F}_{S_n, S_{n-1}}$: continuous time branching process observed up to time S_n; *only difference being that the distribution of the points of birth of founding ancestor is $\mathscr{P}$ conditioned to have a birth X_n time units after the birth of the founding ancestor.*
- $\mathtt{sin}$-tree construction: Infinite path is $\mathbb{Z}^+ = 0, 1, 2, \ldots$.

sin-trees [Jagers+Nerman/Aldous]

Construction

$\mathcal{T}^{\mathtt{sin}}_{\alpha,\mu,\mathscr{P}}$: Random tree with single infinite path.

- $X_0 \sim \exp(\alpha)$ and for $i \geq 1$, $X_i \sim \mu$. $S_n = \sum_0^n X_i$.
- Conditional on the sequence $(S_n)_{n\geq 0}$,
  1. $\mathscr{F}_{X_0}$: continuous time branching process driven by $\mathscr{P}$ observed up to time X_0.
  2. For $n \geq 1$ let $\mathscr{F}_{S_n,S_{n-1}}$: continuous time branching process observed up to time S_n; *only difference being that the distribution of the points of birth of founding ancestor is $\mathscr{P}$ conditioned to have a birth X_n time units after the birth of the founding ancestor.*
- $\mathtt{sin}$-tree construction: Infinite path is $\mathbb{Z}^+ = 0, 1, 2, \ldots$.

  0 designated as the root. $\mathscr{F}_{X_0}$ to be rooted at 0 and for $n \geq 1$ consider $\mathscr{F}_{S_n,S_{n-1}}$ to be rooted at n.
- $\mathbf{f}_k(\mathcal{T}^{\mathtt{sin}}_{\alpha,\mu,\mathscr{P}}) = (\mathscr{F}_{X_0}, \mathscr{F}_{S_1,S_0}, \mathscr{F}_{S_2,S_3}, \ldots, \mathscr{F}_{S_k,S_{k-1}})$.

Case study: degree distribution in preferential attachment

- Recall $\lambda = 2$. Offspring distribution: Yule process $\mathscr{P}(\cdot)$.

Case study: degree distribution in preferential attachment

- Recall $\lambda = 2$. Offspring distribution: Yule process $\mathscr{P}(\cdot)$.
- From above discussion: expect degree of a randomly selected vertex to converge to: $1 + \mathscr{P}(T_2) := 1 + C$ *say*.

Case study: degree distribution in preferential attachment

- Recall $\lambda = 2$. Offspring distribution: Yule process $\mathscr{P}(\cdot)$.
- From above discussion: expect degree of a randomly selected vertex to converge to: $1 + \mathscr{P}(T_2) := 1 + C$ *say*.
- Let $\{E_i : i \geq 0\}$ be independent exponential random variables (independent of T_2) with $E_i \sim \exp(i+1)$.

Case study: degree distribution in preferential attachment

- Recall $\lambda = 2$. Offspring distribution: Yule process $\mathscr{P}(\cdot)$.
- From above discussion: expect degree of a randomly selected vertex to converge to: $1 + \mathscr{P}(T_2) := 1 + C$ *say*.
- Let $\{E_i : i \geq 0\}$ be independent exponential random variables (independent of T_2) with $E_i \sim \exp(i+1)$.
- Fix $k \geq 0$. Then

$$\bar{p}_k := \mathbb{P}(C \geq k) = \mathbb{P}(E_0 + E_1 + \cdots E_{k-1} \leq T_2) = \mathbb{E}(\exp(-2(E_0 + E_1 + \cdots + E_{k-1}))).$$

Case study: degree distribution in preferential attachment

- Recall $\lambda = 2$. Offspring distribution: Yule process $\mathscr{P}(\cdot)$.
- From above discussion: expect degree of a randomly selected vertex to converge to: $1 + \mathscr{P}(T_2) := 1 + C$ *say*.
- Let $\{E_i : i \geq 0\}$ be independent exponential random variables (independent of T_2) with $E_i \sim \exp(i+1)$.
- Fix $k \geq 0$. Then

$$\bar{p}_k := \mathbb{P}(C \geq k) = \mathbb{P}(E_0 + E_1 + \cdots E_{k-1} \leq T_2) = \mathbb{E}(\exp(-2(E_0 + E_1 + \cdots + E_{k-1}))).$$

- $$\bar{p}_k := \prod_{i=0}^{k-1} \left(\frac{i+1}{i+1+2} \right)$$

Case study: degree distribution in preferential attachment

- Recall $\lambda = 2$. Offspring distribution: Yule process $\mathscr{P}(\cdot)$.
- From above discussion: expect degree of a randomly selected vertex to converge to: $1 + \mathscr{P}(T_2) := 1 + C$ *say*.
- Let $\{E_i : i \geq 0\}$ be independent exponential random variables (independent of T_2) with $E_i \sim \exp(i+1)$.
- Fix $k \geq 0$. Then
$$\bar{p}_k := \mathbb{P}(C \geq k) = \mathbb{P}(E_0 + E_1 + \cdots E_{k-1} \leq T_2) = \mathbb{E}(\exp(-2(E_0 + E_1 + \cdots + E_{k-1}))).$$
- $$\bar{p}_k := \prod_{i=0}^{k-1} \left(\frac{i+1}{i+1+2} \right)$$
- $$p_k := \mathbb{P}(C = k) = \frac{2}{k+3} \prod_{i=0}^{k-1} \left(\frac{i+1}{i+1+2} \right) = \frac{2}{(k+3)(k+2)(k+1)} \sim \frac{C}{k^3}$$

Case study: degree distribution in preferential attachment

- Recall $\lambda = 2$. Offspring distribution: Yule process $\mathscr{P}(\cdot)$.
- From above discussion: expect degree of a randomly selected vertex to converge to: $1 + \mathscr{P}(T_2) := 1 + C$ *say*.
- Let $\{E_i : i \geq 0\}$ be independent exponential random variables (independent of T_2) with $E_i \sim \exp(i+1)$.
- Fix $k \geq 0$. Then
$$\bar{p}_k := \mathbb{P}(C \geq k) = \mathbb{P}(E_0 + E_1 + \cdots E_{k-1} \leq T_2) = \mathbb{E}(\exp(-2(E_0 + E_1 + \cdots + E_{k-1}))).$$
- $$\bar{p}_k := \prod_{i=0}^{k-1} \left(\frac{i+1}{i+1+2} \right)$$
- $$p_k := \mathbb{P}(C = k) = \frac{2}{k+3} \prod_{i=0}^{k-1} \left(\frac{i+1}{i+1+2} \right) = \frac{2}{(k+3)(k+2)(k+1)} \sim \frac{C}{k^3}$$
- *Power law degree distribution!*

Random matrices

What more can one do with this machinery?

Notation

- A_n adjacency matrix of tree $\mathcal{T}_n$.

Random matrices

What more can one do with this machinery?

Notation

- A_n adjacency matrix of tree $\mathcal{T}_n$.
- $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the n eigen values.

Random matrices

What more can one do with this machinery?

Notation

- A_n adjacency matrix of tree $\mathcal{T}_n$.
- $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the n eigen values.
- $F_n = \frac{1}{n}\sum_1^n \delta_{\lambda_i}$ spectral distribution.

Setting

- For the convergence of spectral distribution can take general families of trees satisfying sin-tree convergence.

Random matrices

What more can one do with this machinery?

Notation

- A_n adjacency matrix of tree $\mathcal{T}_n$.
- $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the n eigen values.
- $F_n = \frac{1}{n}\sum_1^n \delta_{\lambda_i}$ spectral distribution.

Setting

- For the convergence of spectral distribution can take general families of trees satisfying sin-tree convergence.
- For maximal eigen value convergence talking about preferential attachment with $f(v,n) = \text{Deg}(v,n) + a$.

Main result

Theorem (SB, Evans, Sen 08)

(a) Consider a sequence of trees converging in fringe since to a random infinite $\mathtt{sin}$-tree. Then there exists a model dependent probability distribution function F such that

$$d(F_n, F) \xrightarrow{P} 0, \qquad \text{as } n \to \infty.$$

(b) Let $\gamma_\alpha = \alpha + 2$. Then for the linear preferential attachment model

$$\left(\frac{\lambda_1}{n^{1/2\gamma_\alpha}}, \frac{\lambda_2}{n^{1/2\gamma_\alpha}}, \ldots, \frac{\lambda_k}{n^{1/2\gamma_\alpha}} \right) \xrightarrow{d} \nu_k$$

Main result

Theorem (SB, Evans, Sen 08)

(a) Consider a sequence of trees converging in fringe since to a random infinite `sin`-tree. Then there exists a model dependent probability distribution function F such that

$$d(F_n, F) \xrightarrow{P} 0, \qquad \text{as } n \to \infty.$$

(b) Let $\gamma_\alpha = \alpha + 2$. Then for the linear preferential attachment model

$$\left(\frac{\lambda_1}{n^{1/2\gamma_\alpha}}, \frac{\lambda_2}{n^{1/2\gamma_\alpha}}, \ldots, \frac{\lambda_k}{n^{1/2\gamma_\alpha}} \right) \xrightarrow{d} \nu_k$$

Spectral distribution turns out to be a local property of random node, maximal eigen values, local property about the root.

Spectral distribution: Method of proof

Stieltjes transform

$$s(z) = \int_{\mathbb{R}} \frac{1}{x - z} \, dF_n(x)$$

Spectral distribution: Method of proof

Stieltjes transform

$$s(z) = \int_{\mathbb{R}} \frac{1}{x-z} dF_n(x)$$

For eigen value distribution

$$\begin{aligned} s(z) &= \frac{1}{n} \mathrm{Tr}(A - zI)^{-1} \\ &= \frac{1}{n} \sum_{v=1}^{n} R_{vv}(z) \end{aligned}$$

Spectral distribution: Method of proof

Stieltjes transform

$$s(z) = \int_{\mathbb{R}} \frac{1}{x-z} dF_n(x)$$

For eigen value distribution

$$\begin{aligned} s(z) &= \frac{1}{n}\mathrm{Tr}(A - zI)^{-1} \\ &= \frac{1}{n}\sum_{v=1}^{n} R_{vv}(z) \end{aligned}$$

$$R_{vv}(z) = \frac{1}{-z + \sum_1^{N(v)} R_{v_i v_i}(z) + R_{A(v)}^{\mathtt{big}}(z)}$$

Spectral distribution contd

- Fix $Im(z) > 1$. Iterate the above expansion d times. Get a continued fraction upto d terms and some error term.

Spectral distribution contd

- Fix $Im(z) > 1$. Iterate the above expansion d times. Get a continued fraction upto d terms and some error term.
- Not hard to see that for $Im(z) > 1$, this implies that $s_n(z)$ "depends" on the first K terms.

Spectral distribution contd

- Fix $Im(z) > 1$. Iterate the above expansion d times. Get a continued fraction upto d terms and some error term.
- Not hard to see that for $Im(z) > 1$, this implies that $s_n(z)$ "depends" on the first K terms.
- Fringe convergence of the random trees tells you what happens upto distance K for any fixed K.

Spectral distribution contd

- Fix $Im(z) > 1$. Iterate the above expansion d times. Get a continued fraction upto d terms and some error term.
- Not hard to see that for $Im(z) > 1$, this implies that $s_n(z)$ "depends" on the first K terms.
- Fringe convergence of the random trees tells you what happens upto distance K for any fixed K.
- So not hard to show that there exists a fixed Stieltjes transform $s(z)$ such that,

$$s_n(z) \xrightarrow{P} s(z).$$

Properties and questions

Open question

- We have established sufficient conditions for a point $a \in \mathbb{R}$ to be an atom of limiting F
- Implies that for most standard models, limiting F has dense set of atoms

Properties and questions

Open question

- We have established sufficient conditions for a point $a \in \mathbb{R}$ to be an atom of limiting F
- Implies that for most standard models, limiting F has dense set of atoms
- **Open Question:** Does limiting F have absolutely continuous part?
- Connections to areas such as Random Schrodinger operators?

At this point

- If one can embed things in a continuous time all good things happen.
- How far can one push such embeddings? Can these continuous time branching processes arise in the limit even when no such embeddings exist?

Outline

From the Retweet Graph to the Superstar Model

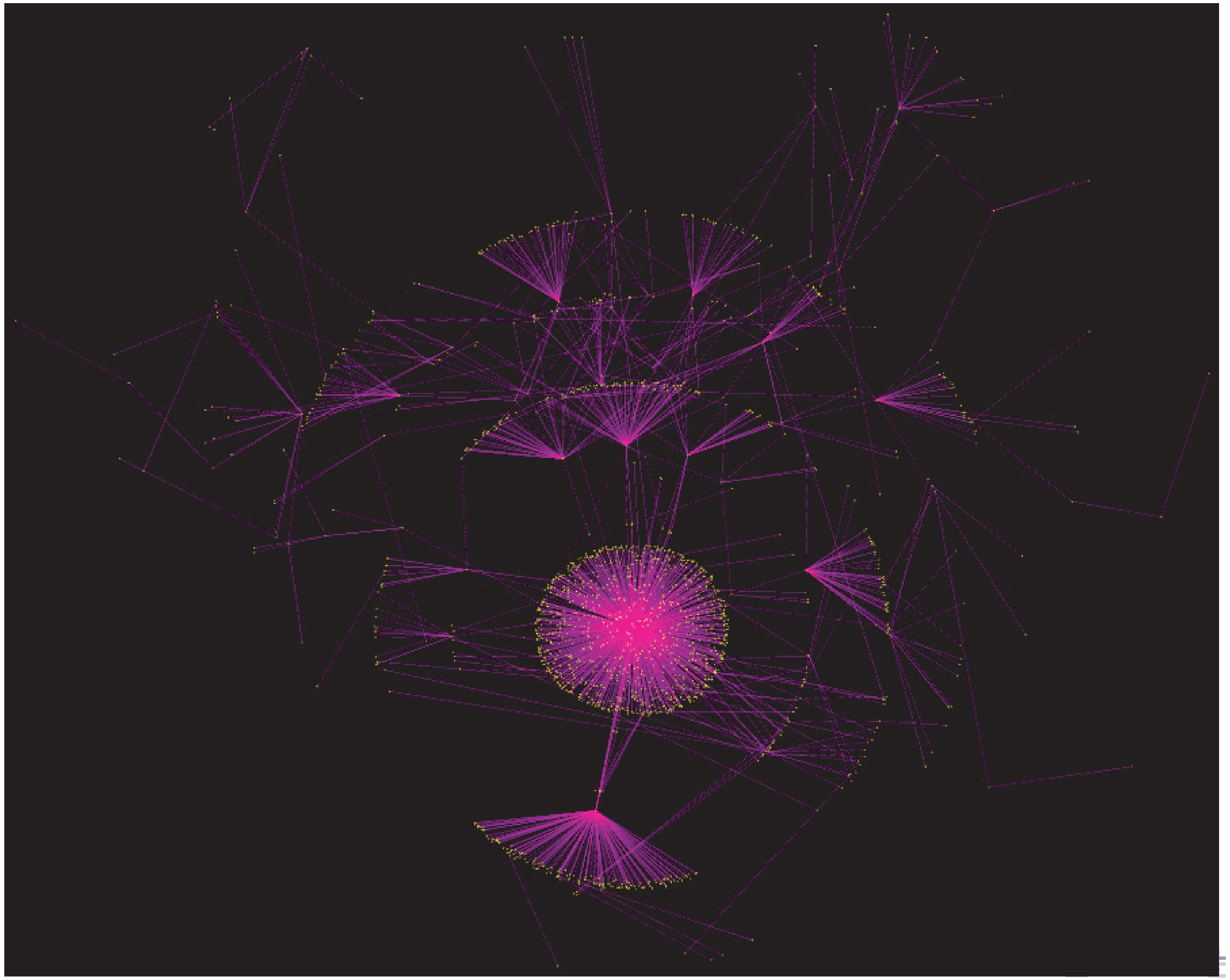
- Joint work with J Michael Steele (Wharton) and Tauhid Zaman (MIT).

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.

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.



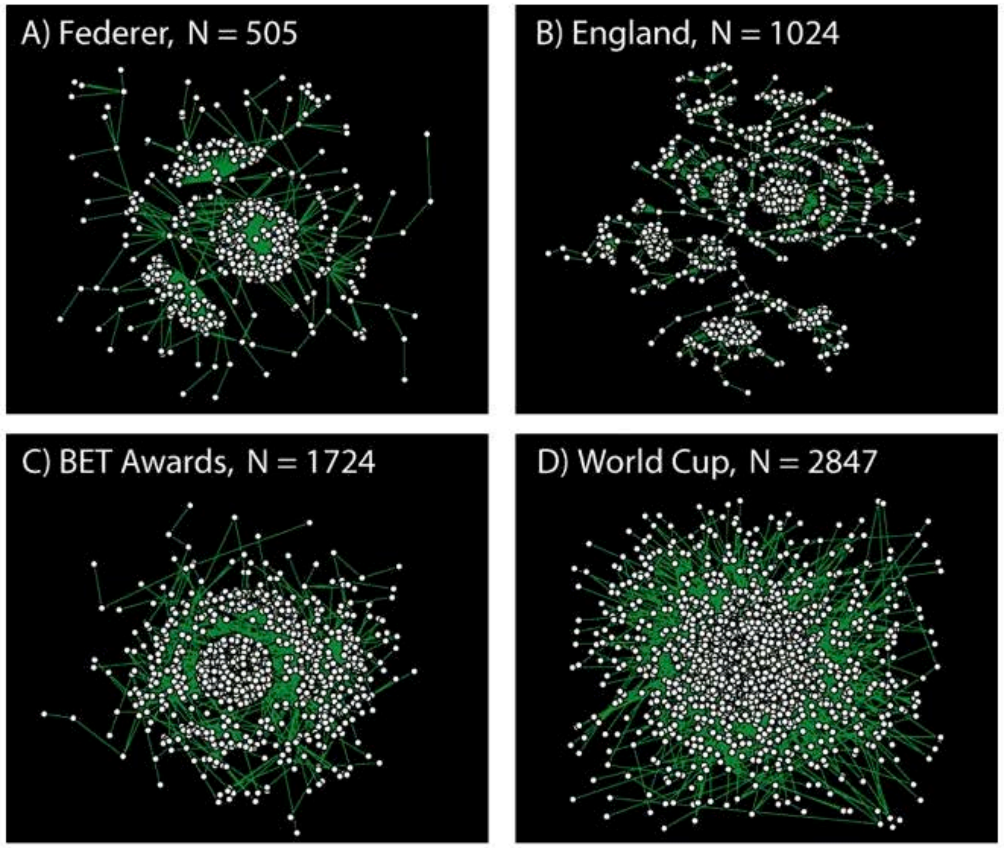
Some Empirical Retweet Graphs

- Retweet graphs were constructed for 13 different public events [1]

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

Some Empirical Retweet Graphs

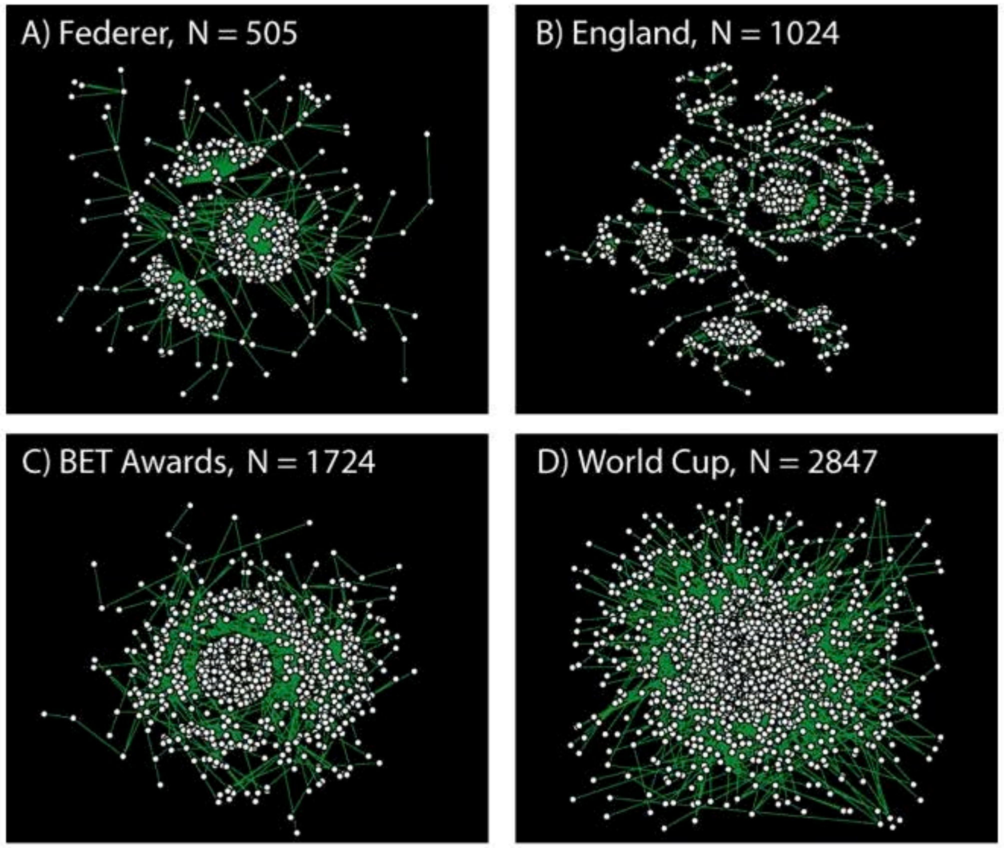
- Retweet graphs were constructed for 13 different public events [1]



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

Some Empirical Retweet Graphs

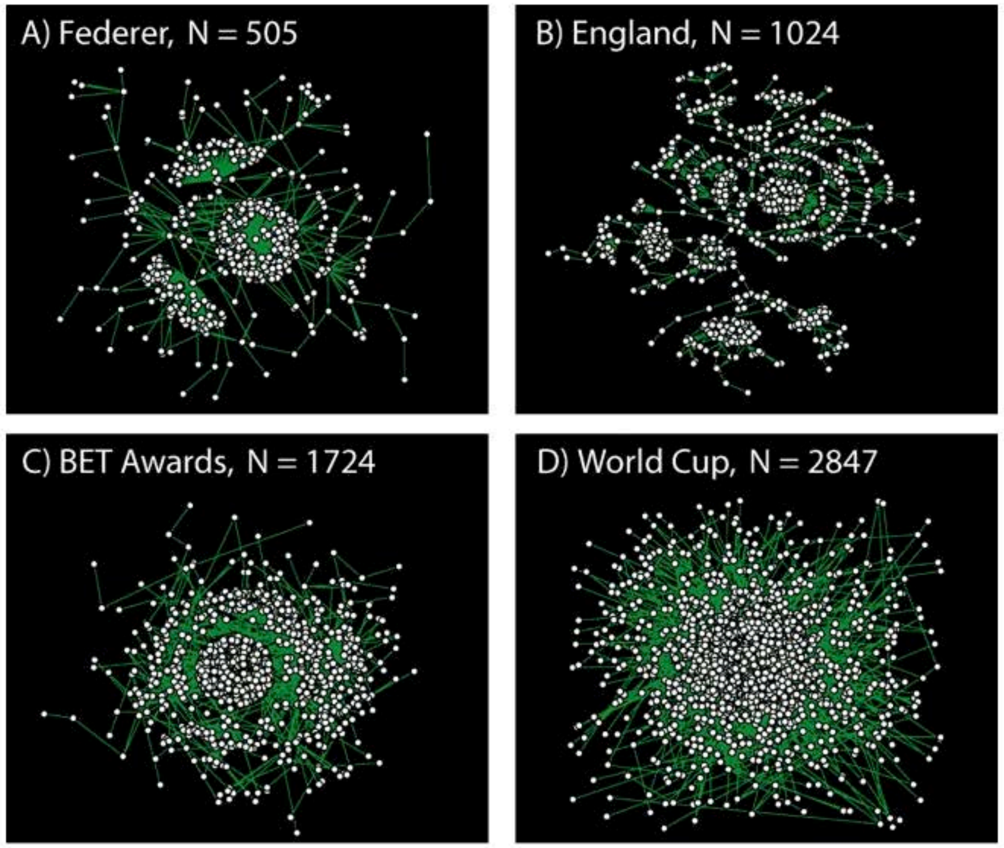
- Retweet graphs were constructed for 13 different public events [1]
  - Sports, breaking news stories, and entertainment events



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

Some Empirical Retweet Graphs

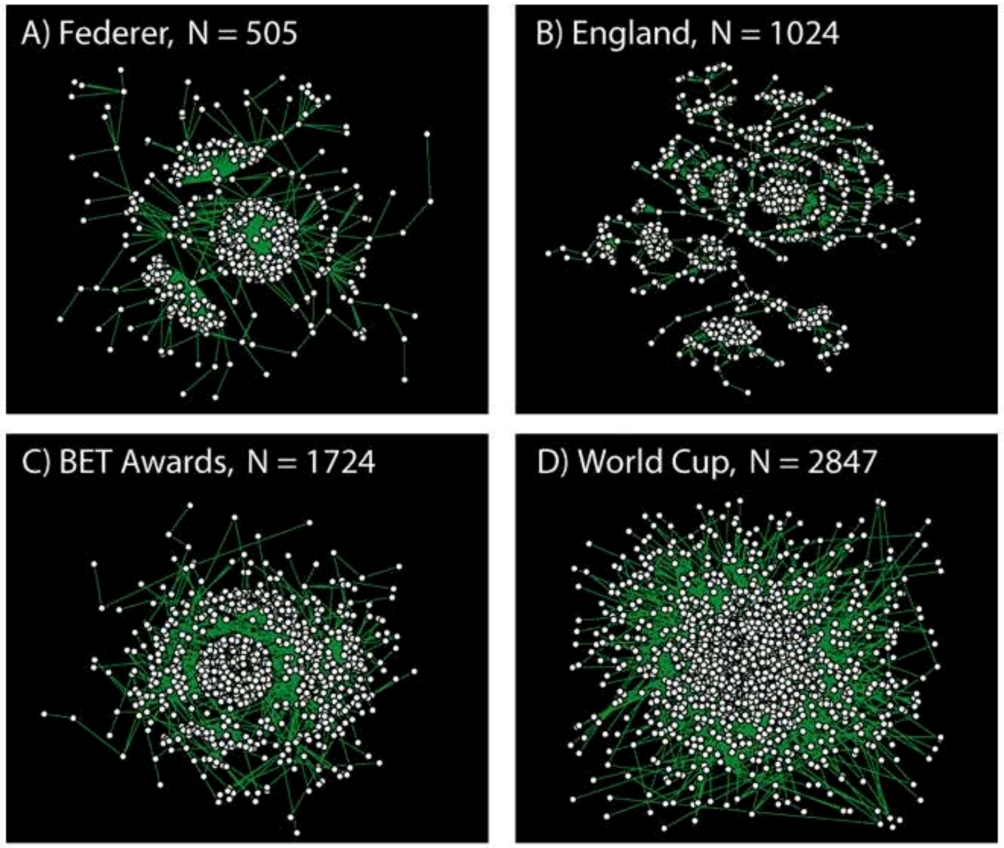
- Retweet graphs were constructed for 13 different public events [1]
  - Sports, breaking news stories, and entertainment events
  - Time range for each topic was between 4-6 hours



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

Some Empirical Retweet Graphs

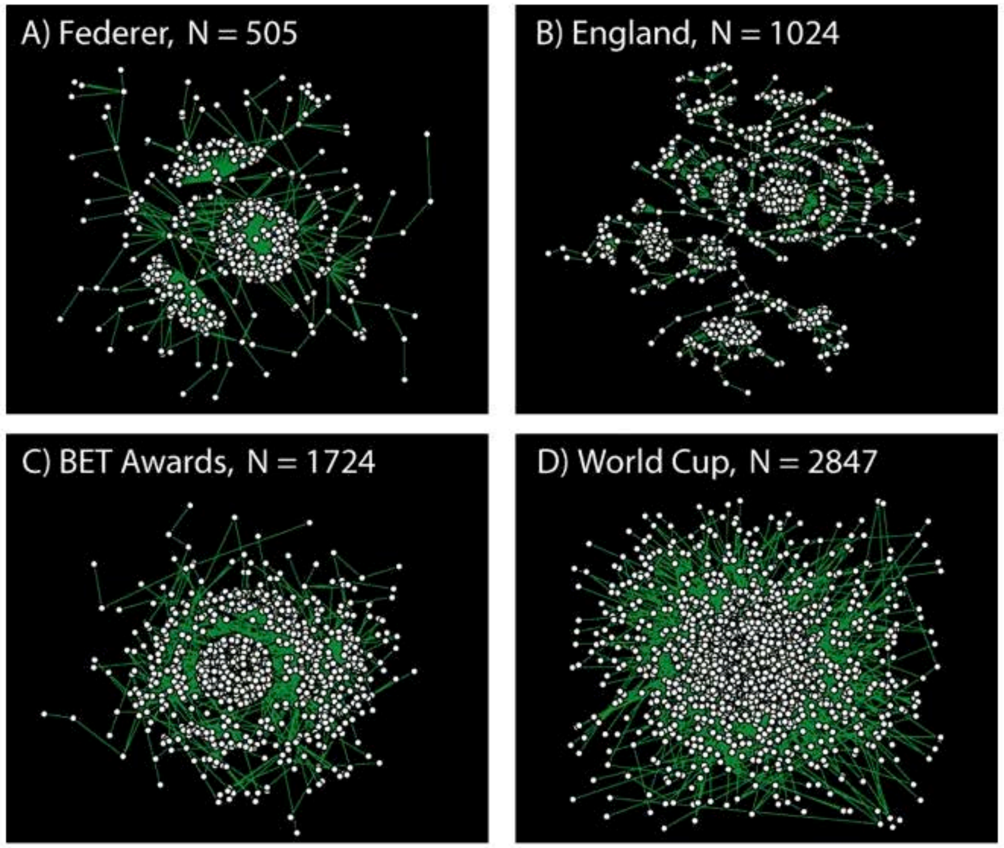
- Retweet graphs were constructed for 13 different public events [1]
  - Sports, breaking news stories, and entertainment events
  - Time range for each topic was between 4-6 hours
- Graphs are very tree-like (few cycles)



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

Some Empirical Retweet Graphs

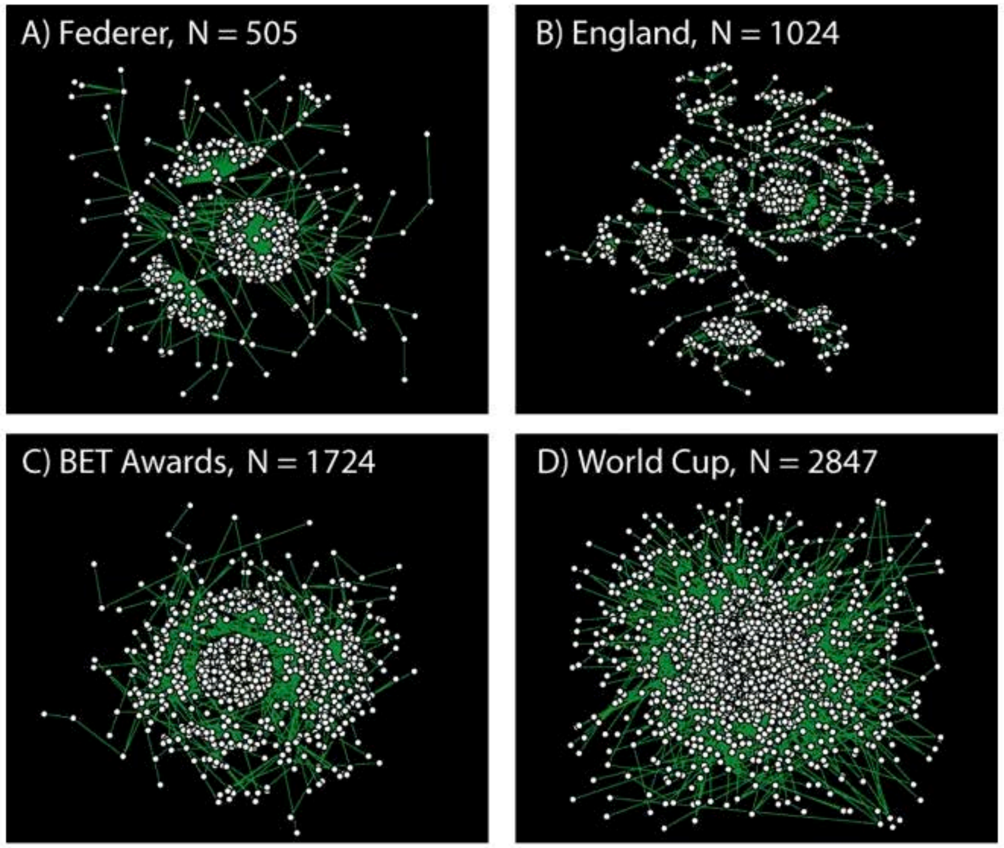
- Retweet graphs were constructed for 13 different public events [1]
  - Sports, breaking news stories, and entertainment events
  - Time range for each topic was between 4-6 hours
- Graphs are very tree-like (few cycles)
- Graphs each have one giant component which we want to study



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

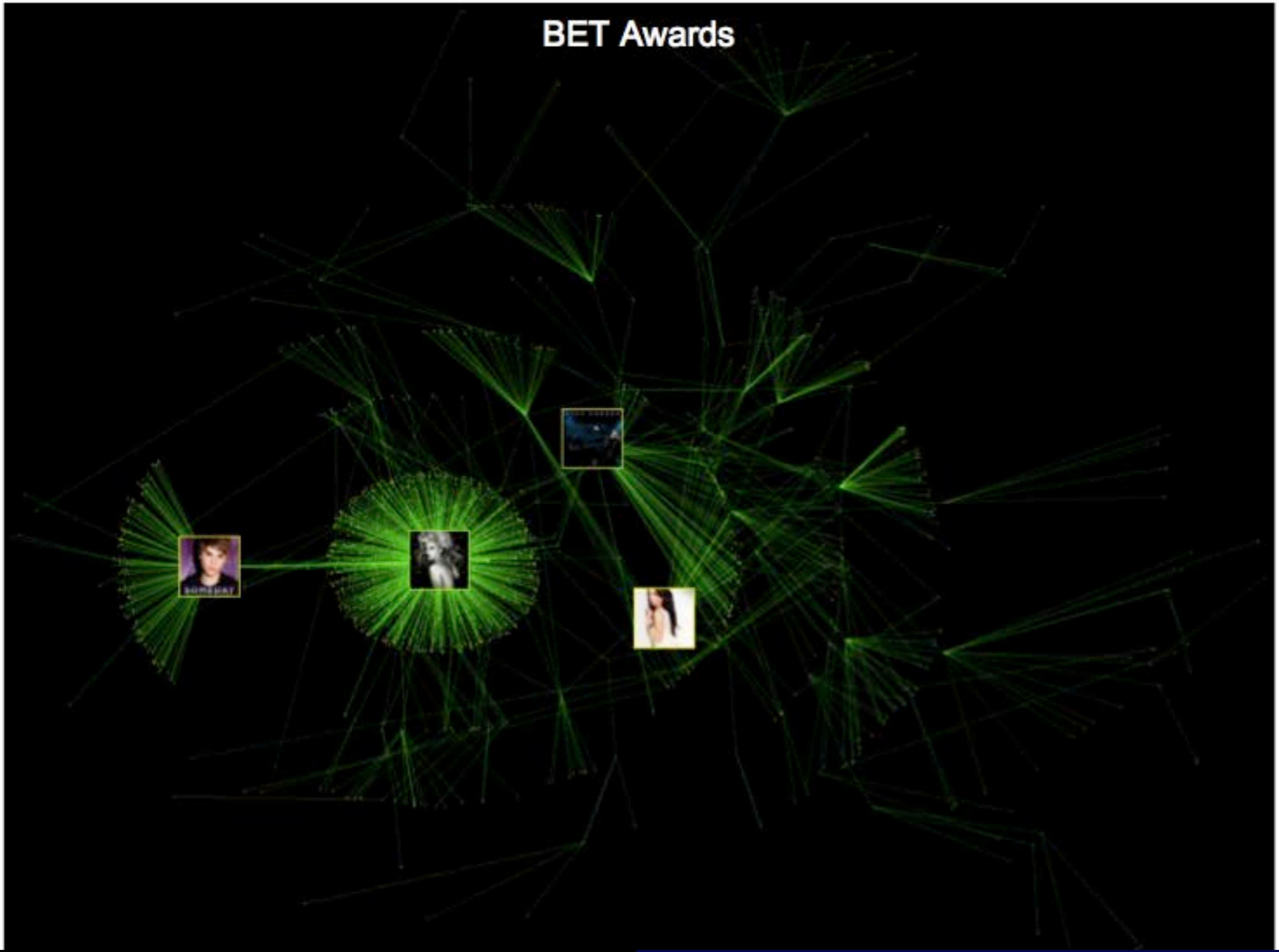
Some Empirical Retweet Graphs

- 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

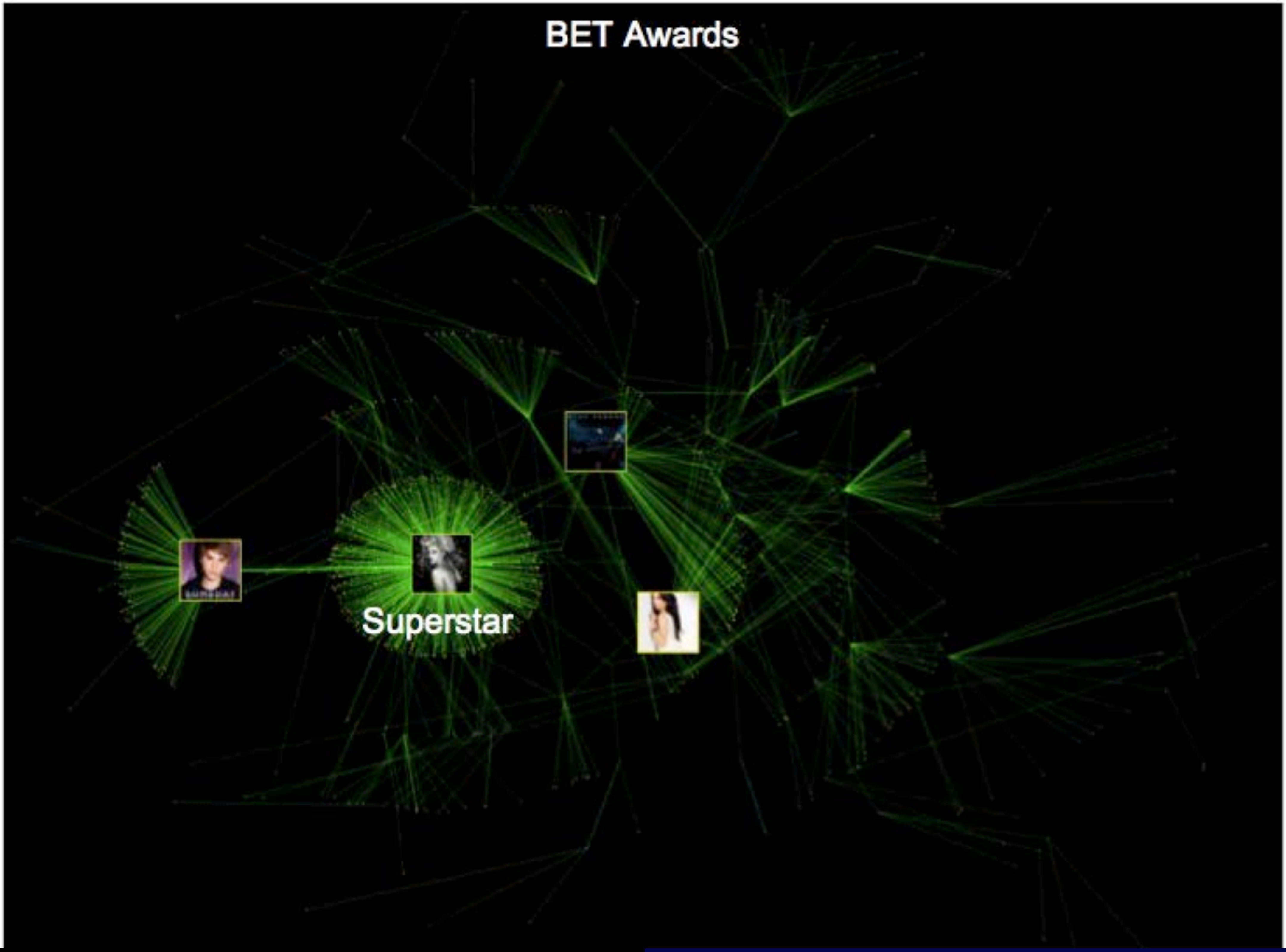


[1] Data courtesy of Microsoft Research, Cambridge, MA

The superstar model

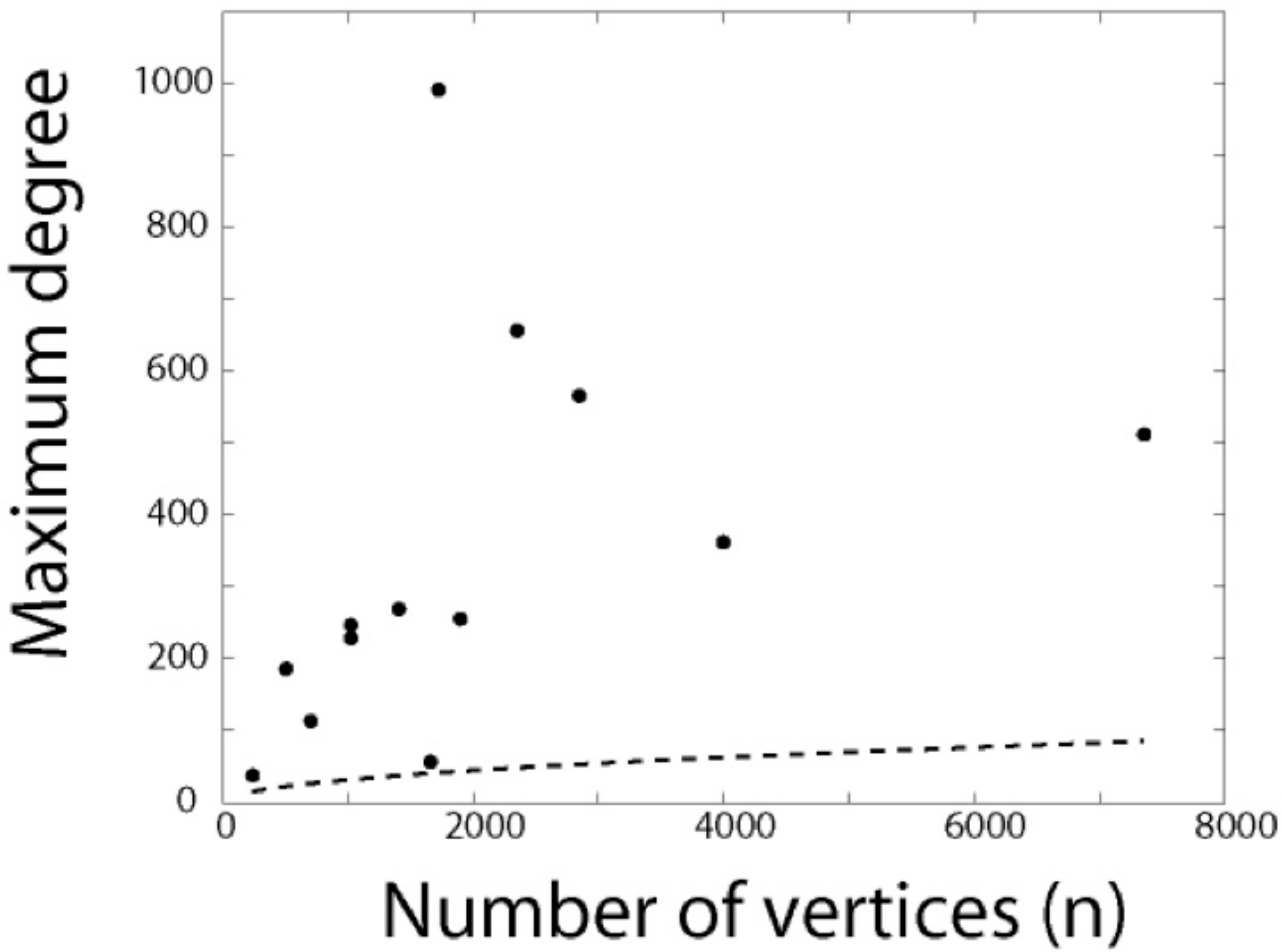


The superstar model

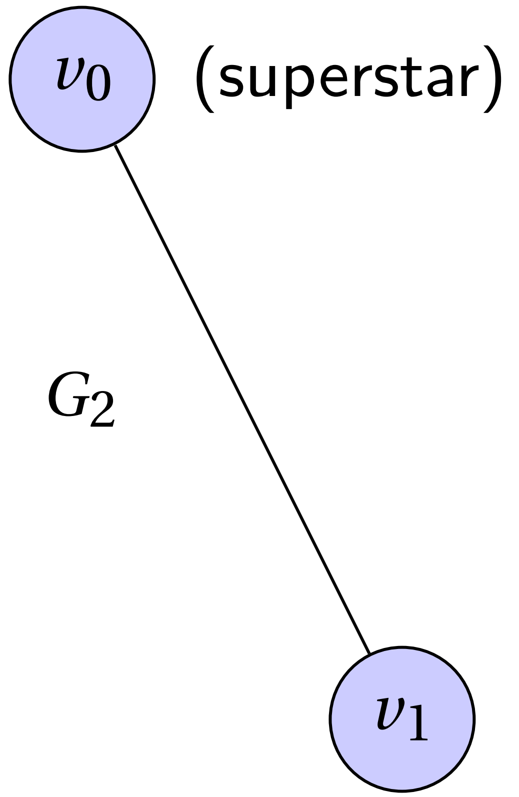


The superstar model

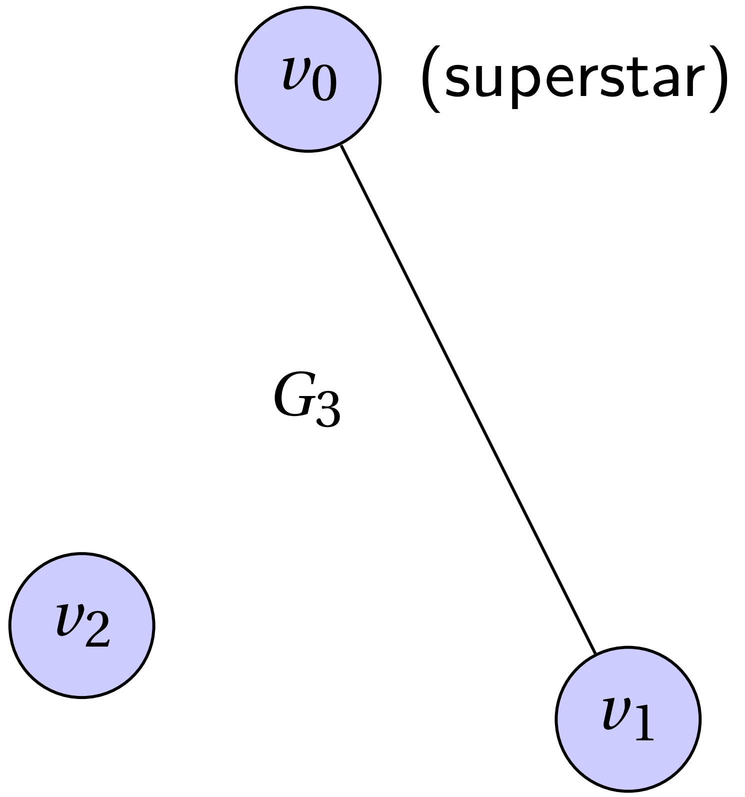
- Max degree in retweet graph is on the order of graph size (i.e. $M_G \sim pn$)
- Preferential attachment predicts sub-linear max degree



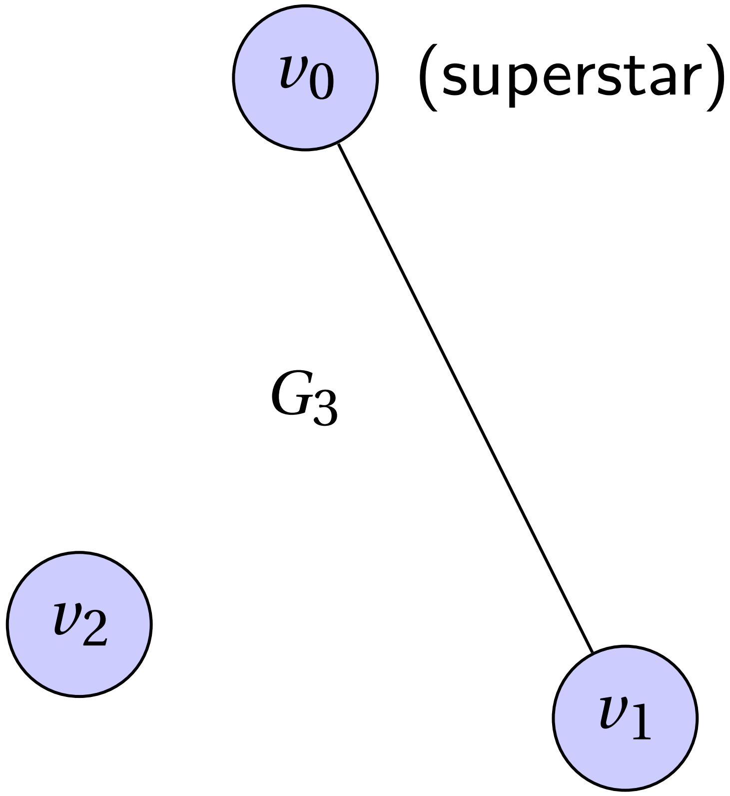
The Superstar Model



The Superstar Model

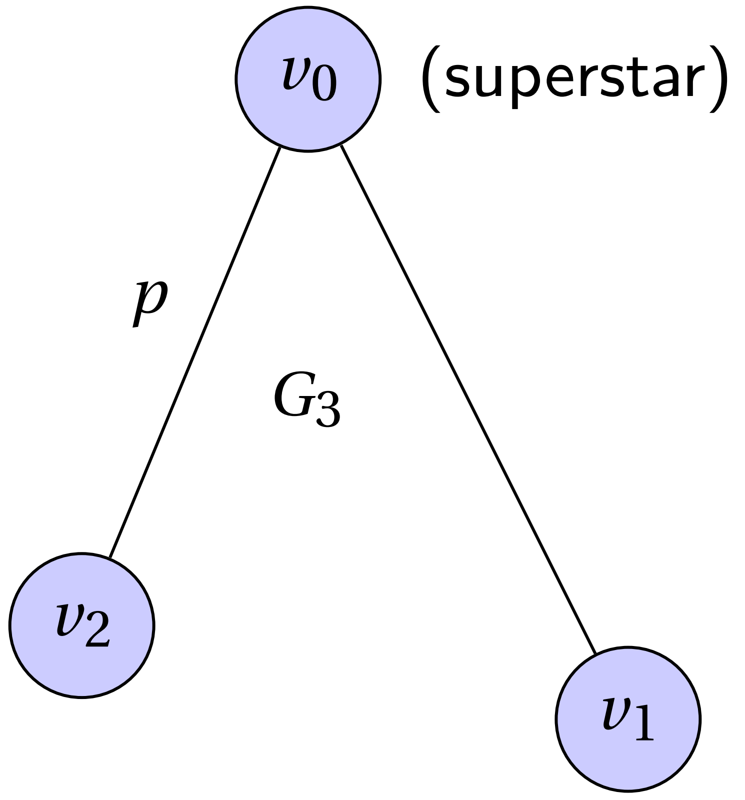


The Superstar Model



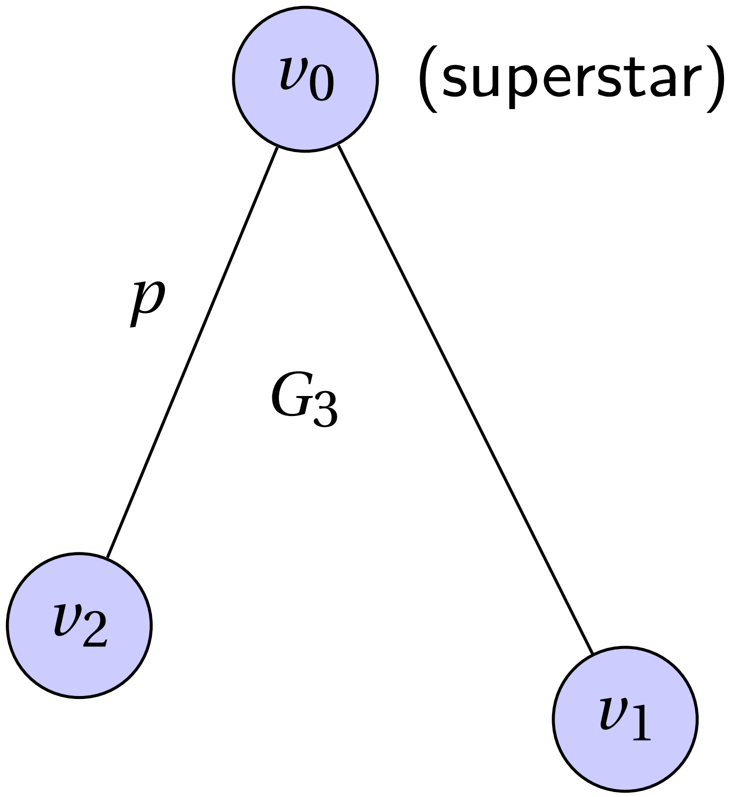
- Attach to superstar with probability p

The Superstar Model



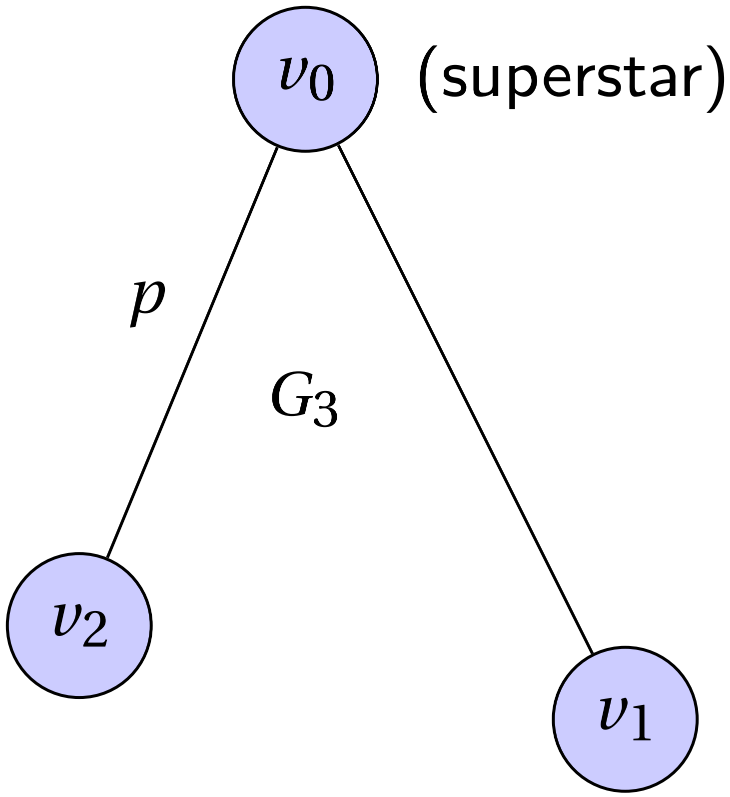
- Attach to superstar with probability p

The Superstar Model



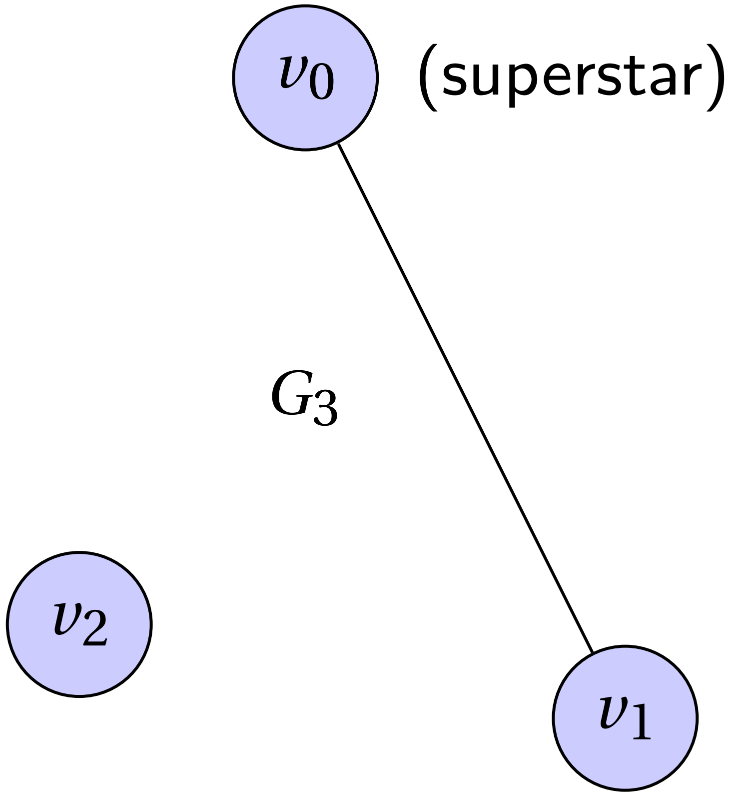
- Attach to superstar with probability p
- Else with probability $1-p$ attach to one of the non-superstar vertices.

The Superstar Model



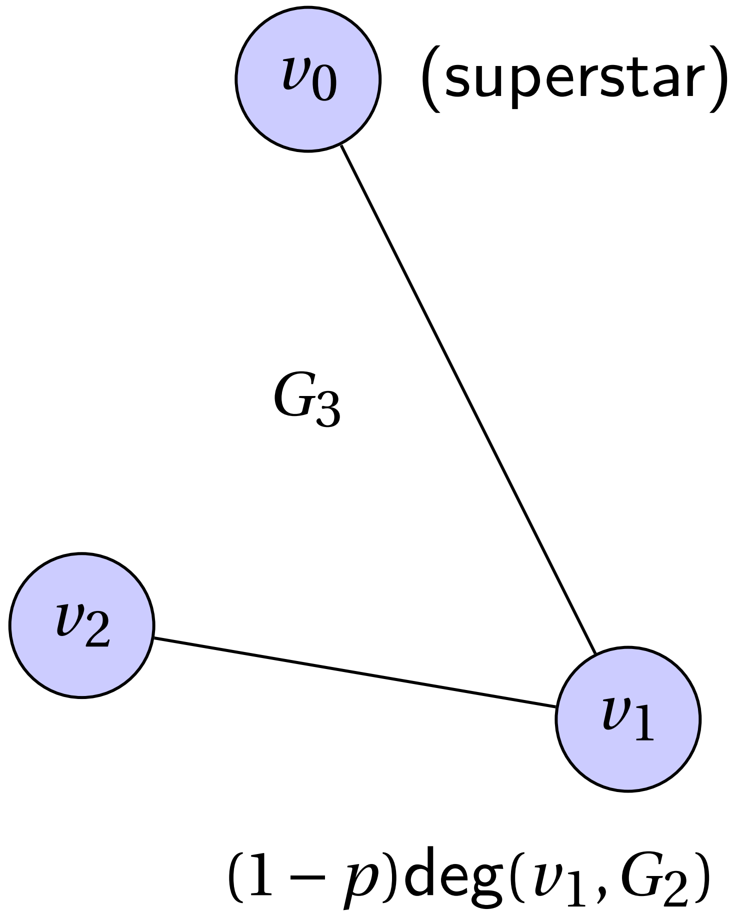
- Attach to superstar with probability p
- Else with probability $1-p$ attach to one of the non-superstar vertices.
- Non-SS Attachment Rule: probability proportional to its degree (preferential attachment rule)

The Superstar Model



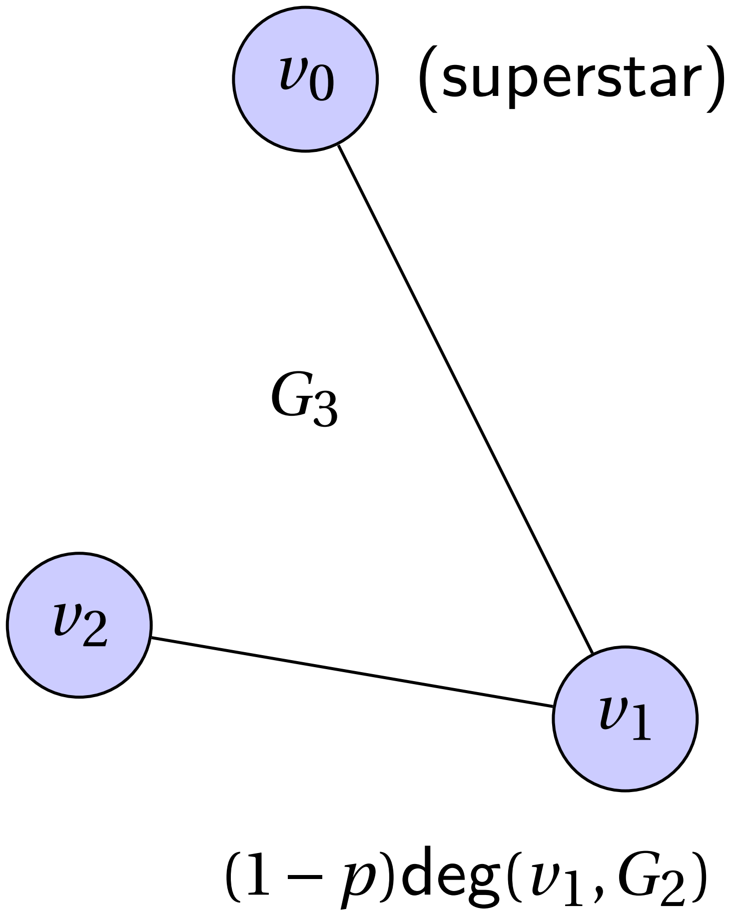
- Attach to superstar with probability p
- Else with probability $1-p$ attach to one of the non-superstar vertices.
- Non-SS Attachment Rule: probability proportional to its degree (preferential attachment rule)

The 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 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.

Outline

Superstar Degree

Theorem

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

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

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"

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...

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*

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

Non-Superstar Degree

Theorem

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

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

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

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

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 $\beta = 3 + p/(1-p)$

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,

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

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

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

Outline

1. Preferential attachment: Base model
   - Continuous time construction
   - Local weak convergence
2. Twitter event networks and the superstar model
   - Retweet Graph and Superstar Model
   - Main Results
   - Comparison with Preferential Attachment Model
   - Superstar Model: Tools for Analysis
3. Power of choice and random trees
4. Conclusion

Superstar Model vs Preferential Attachment

Model	Superstar Model	Preferential Attachment
Superstar Degree		
Maximal non-superstar degree exponent		
Degree distribution power-law exponent		

Superstar Model vs Preferential Attachment

Model	Superstar Model	Preferential Attachment
Superstar Degree	$\Theta(n)$	
Maximal non-superstar degree exponent		
Degree distribution power-law exponent		

Superstar Model vs Preferential Attachment

Model	Superstar Model	Preferential Attachment
Superstar Degree	$\Theta(n)$	NA
Maximal non-superstar degree exponent		
Degree distribution power-law exponent		

Superstar Model vs Preferential Attachment

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

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		

Superstar Model vs Preferential Attachment

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

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

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:

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

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

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

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"

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

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

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

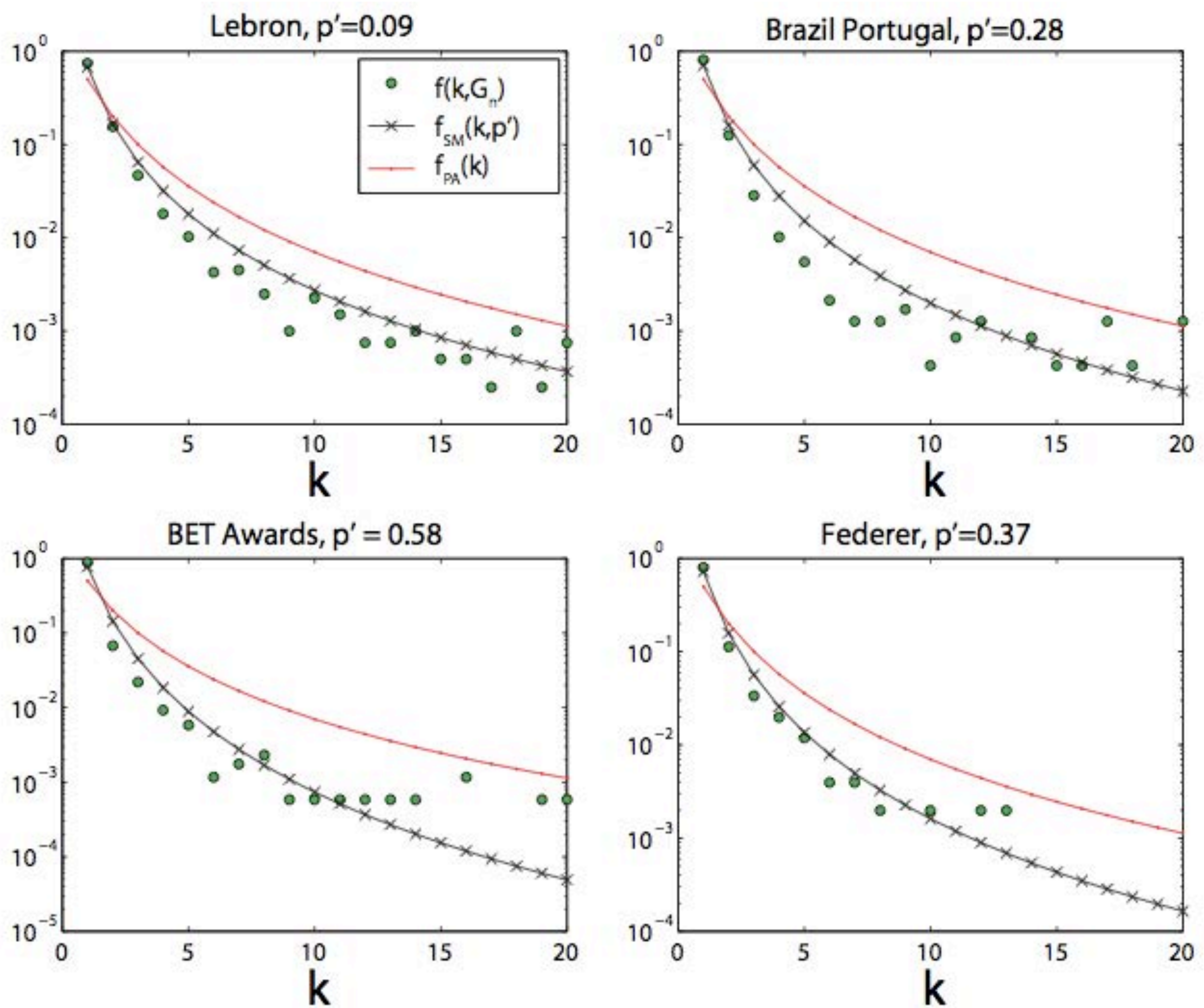
- Bottom Line: We get a nice fit "observed vs predicted"

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

- 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.

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 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.

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?

Outline

Basic Link: Branching Processes

- Proto-Idea: Branching processes have a natural role almost anytime one considers a stochastically evolving tree.

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.

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.

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.

Basic Link: Branching Processes

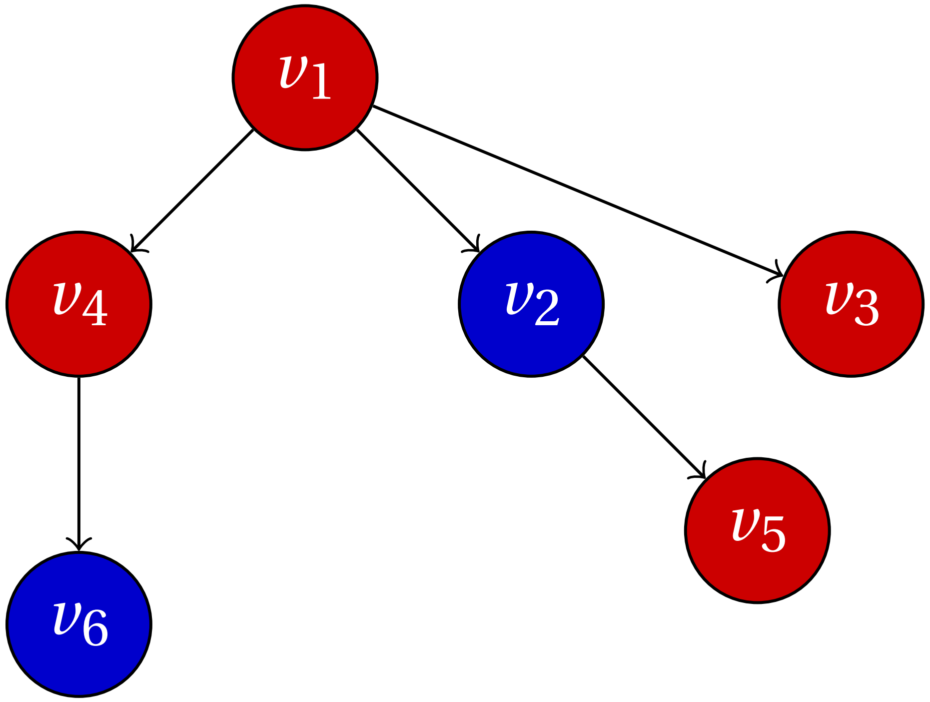
- Proto-Idea: Branching processes have a natural role almost anytime one considers a stochastically evolving tree.
- More Concrete Observation: If the birth rates depend on the number of children, the arithmetic of the Poisson process relates nicely to the arithmetic of preferential attachment.
- Creating 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

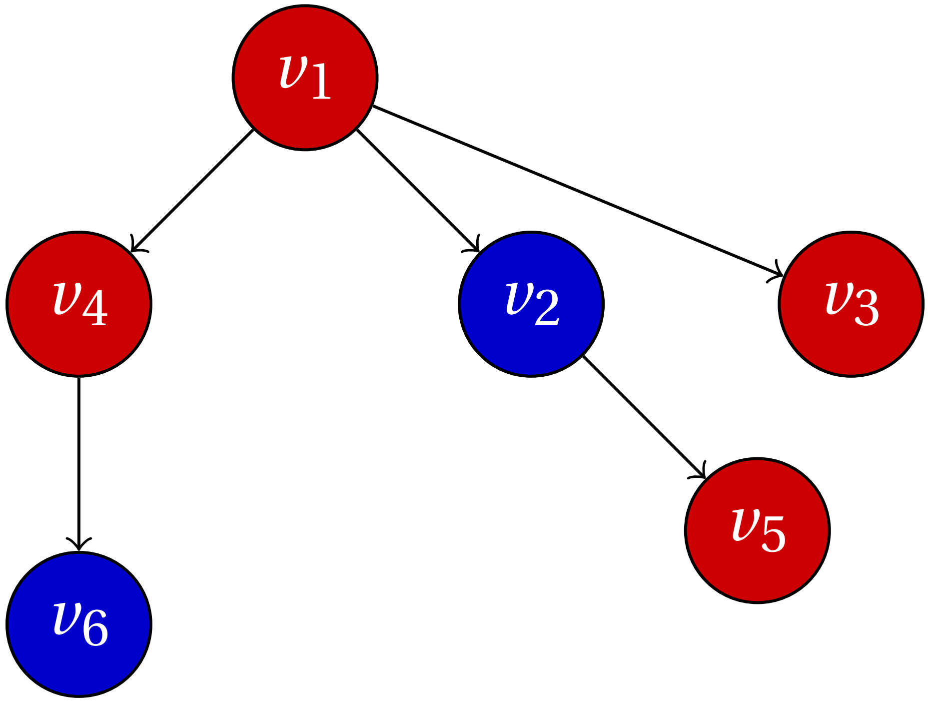
Introduction of a Special Branching Process

- Two types of vertices: red and blue



Introduction of a Special Branching Process

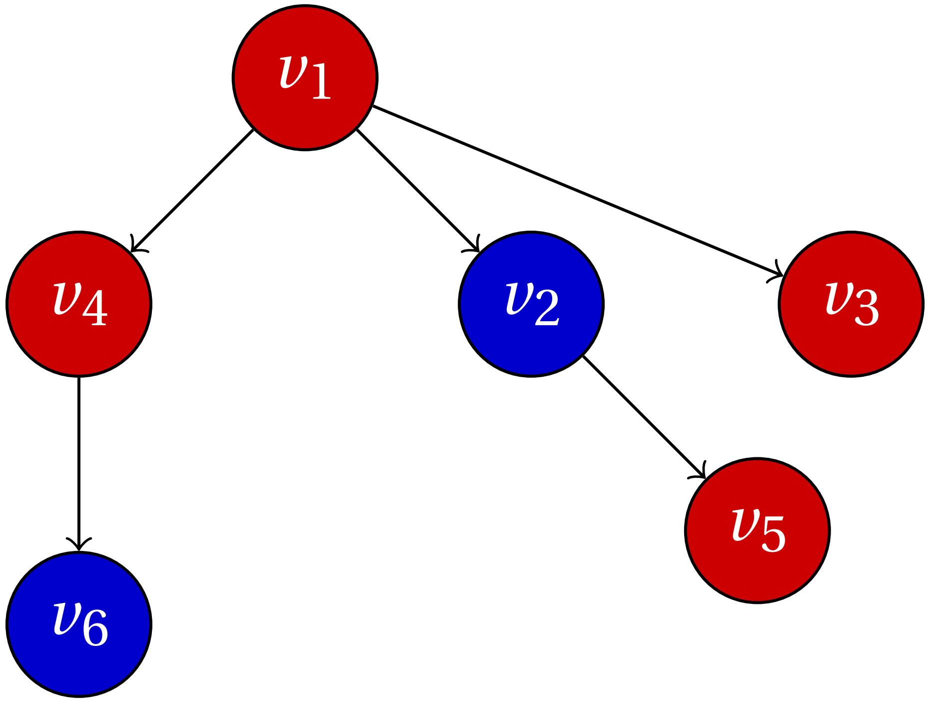
- Two types of vertices: red and blue
- Each vertex gives birth to vertices according to a non-homogeneous Poisson process that has rate proportional to (1+ number of blue children)



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)

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

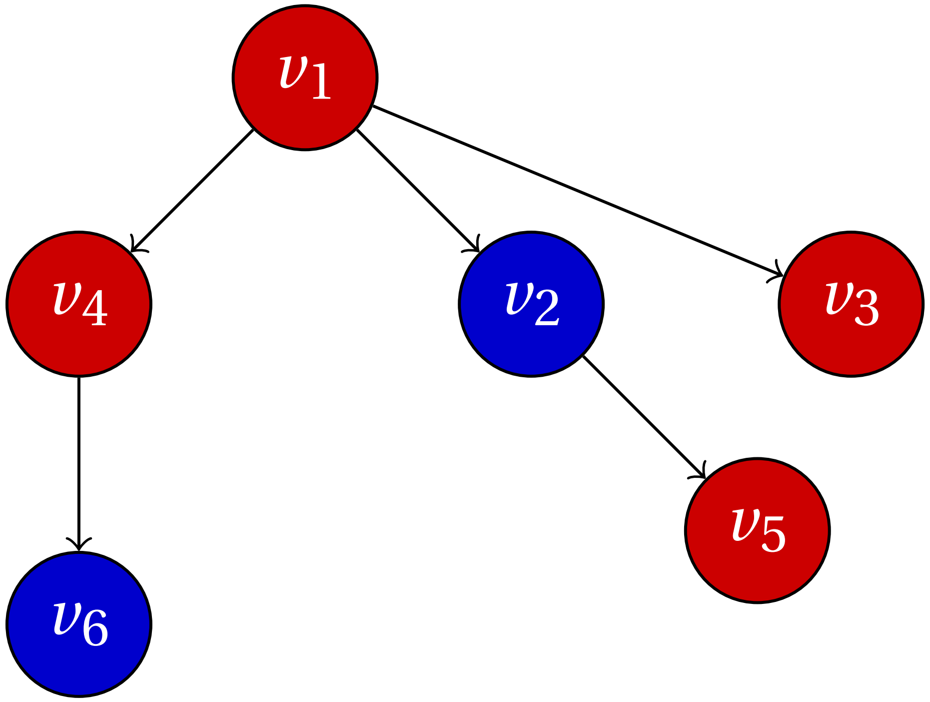


Introduction of a Special Branching Process

- 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 $1-p$

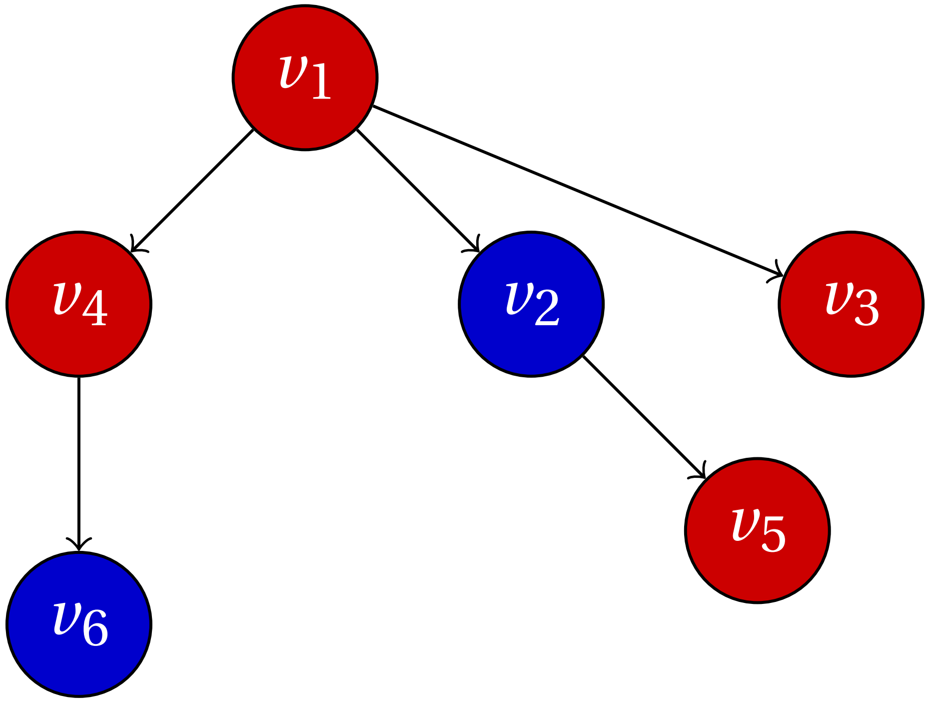


Introduction of a Special Branching Process

- 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 $1-p$



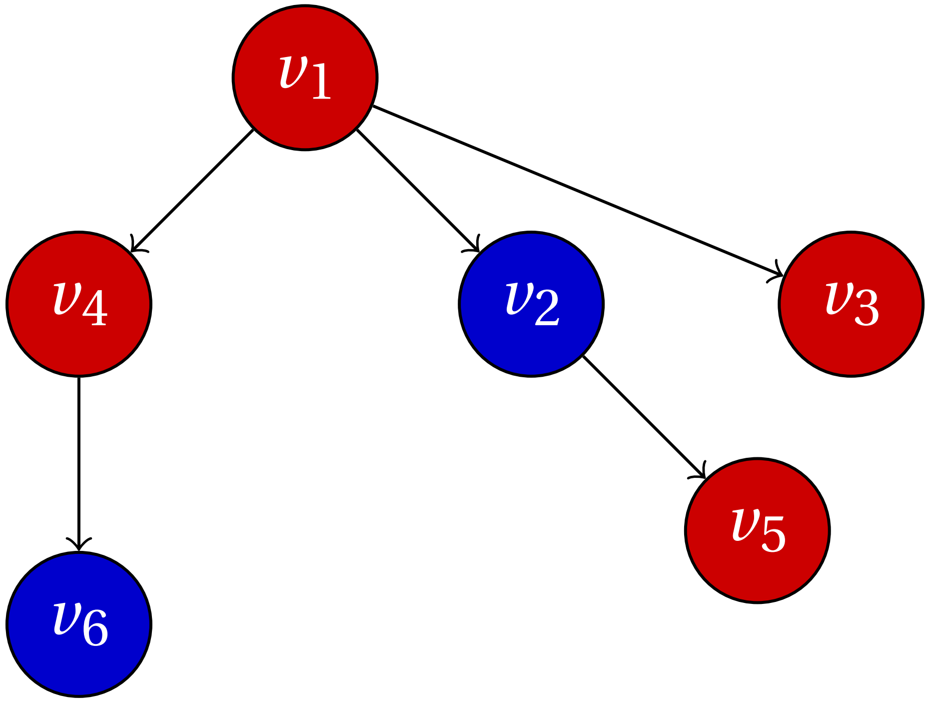
$$\mathscr{F}(t) \quad = \quad \text{Branching process at time } t$$

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 $1-p$



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

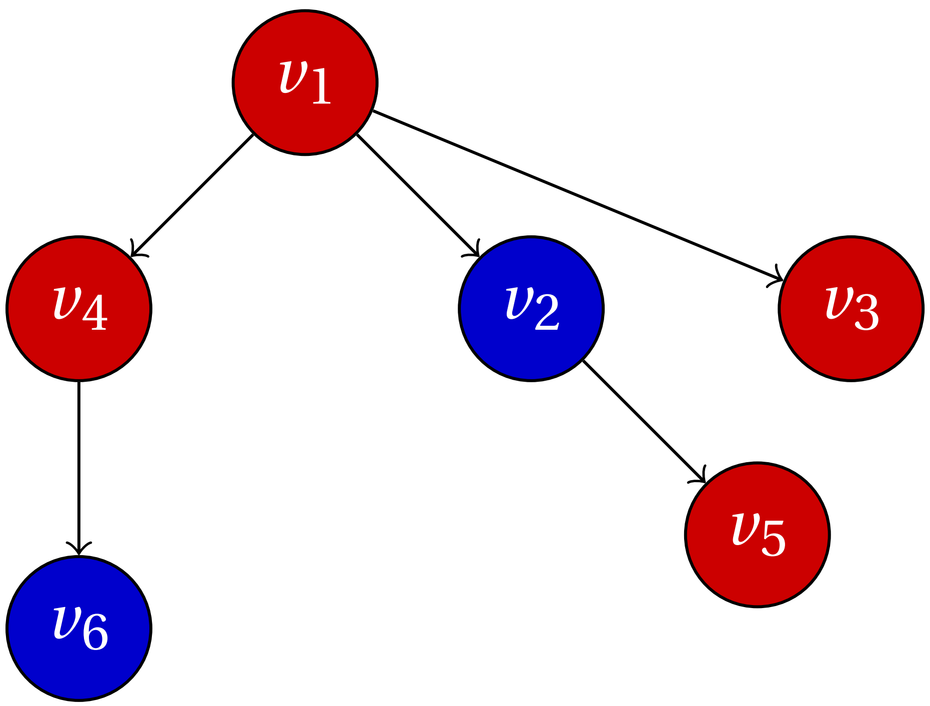
Introduction of a Special Branching Process

- 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 $1-p$

$c_B(v_1, t) = 1$



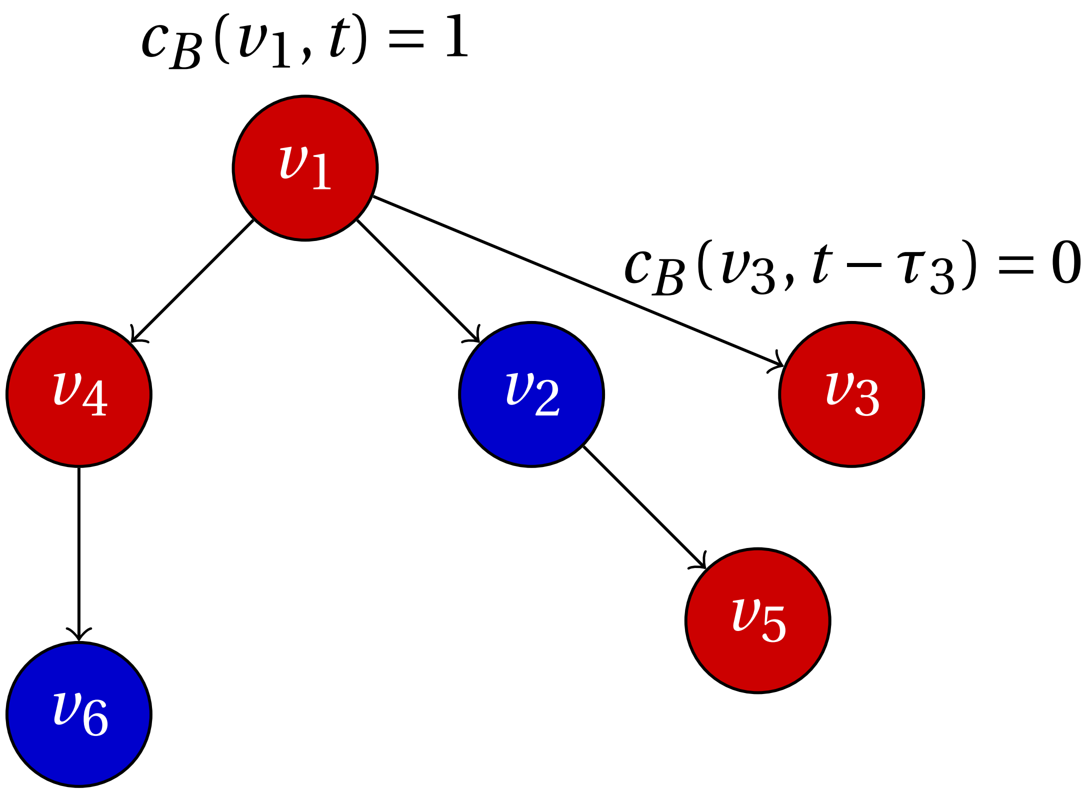
$$\begin{aligned} \mathscr{F}(t) &= \text{Branching process at time } t \\ \tau_n &= \inf\{t : |\mathscr{F}(t)| = n\} \end{aligned}$$

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)

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

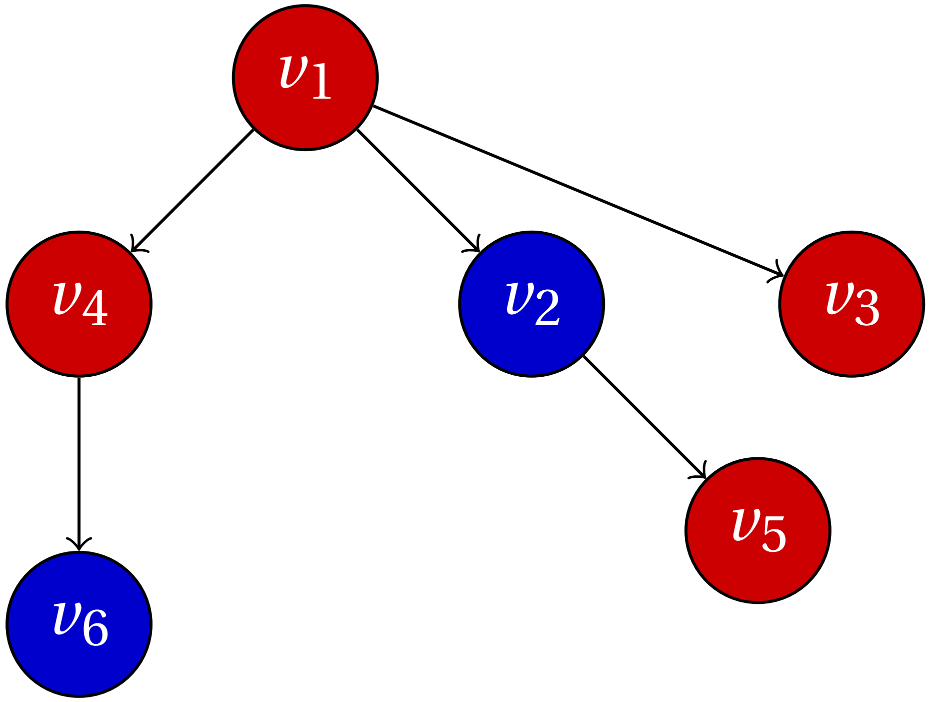
- At birth vertex is painted red with probability p and painted blue with probability $1-p$



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

Surgery: From BP Model to Superstar Model

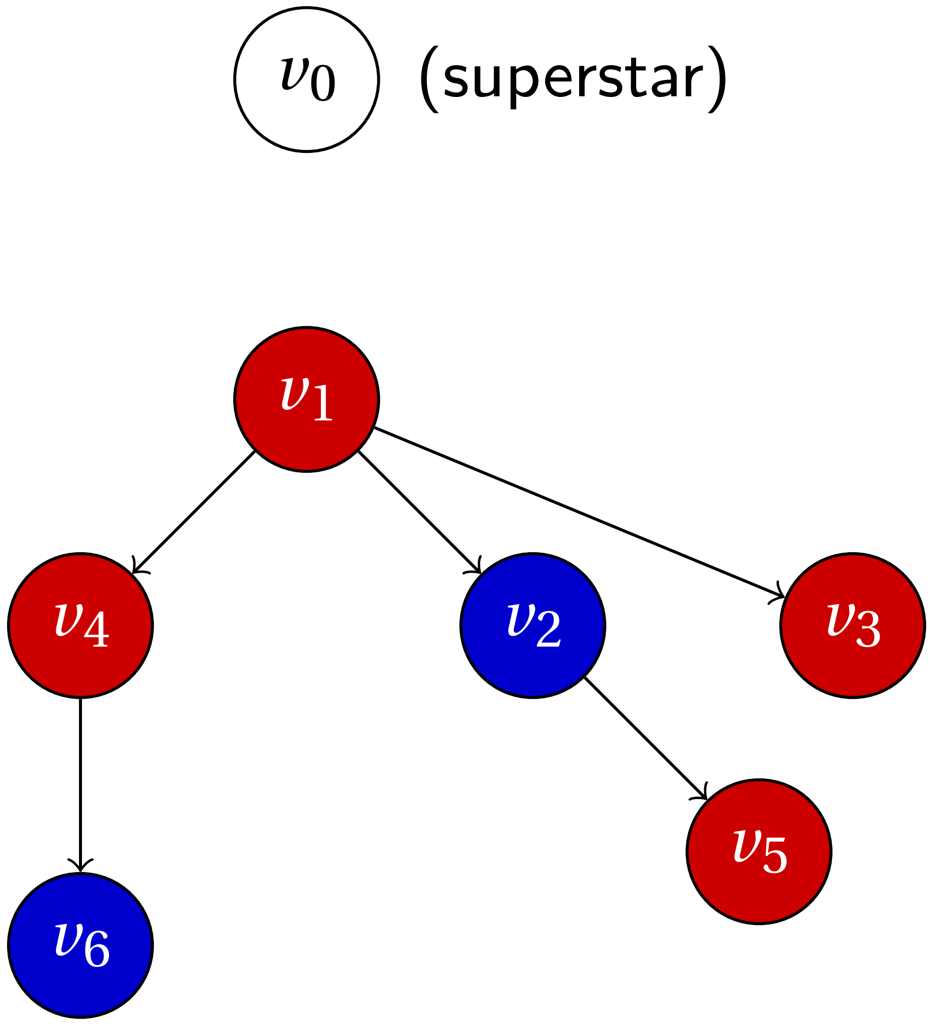
$\mathscr{F}(\tau_6)$



Surgery: From BP Model to Superstar Model

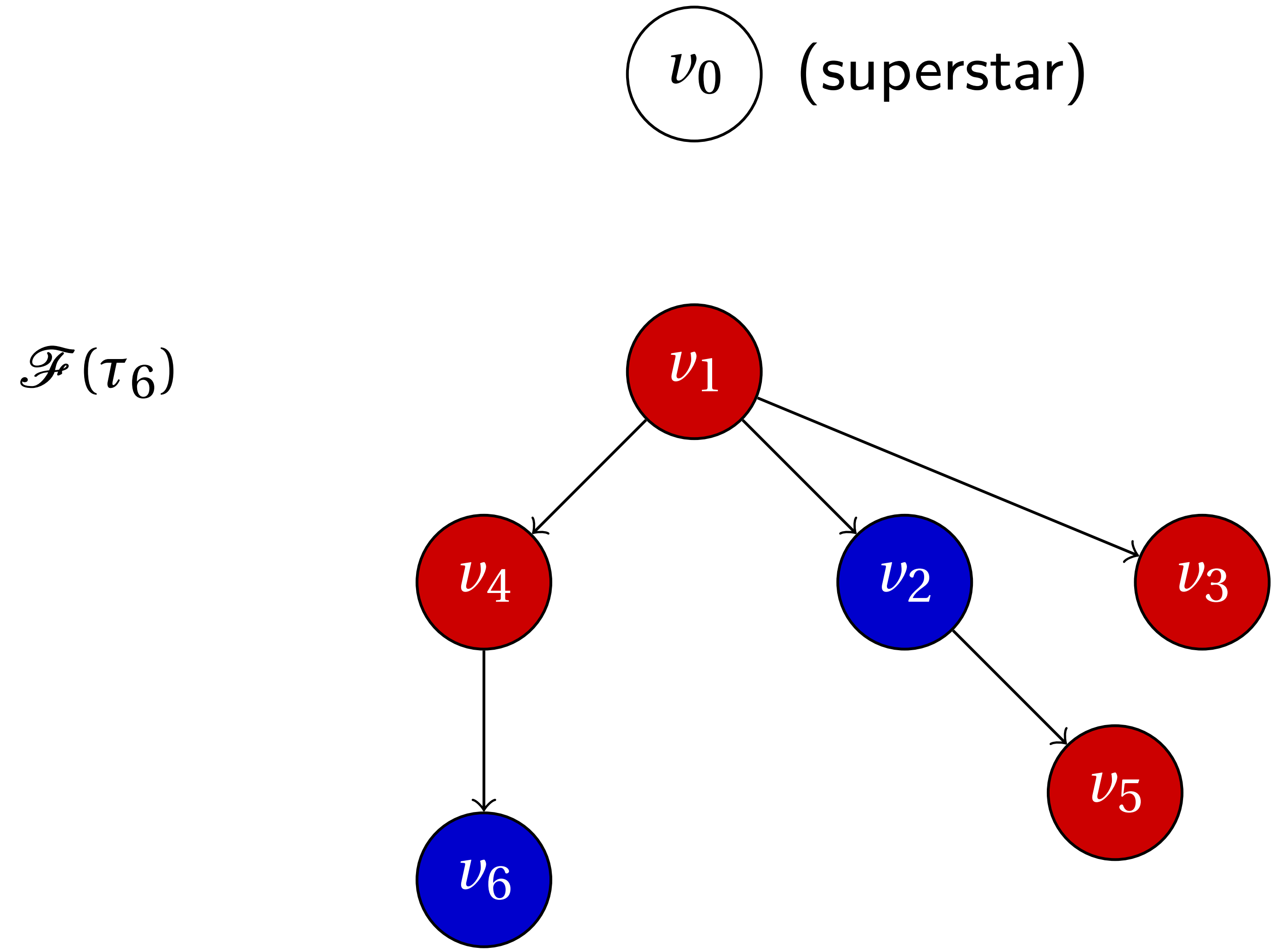
- Add an exogenous superstar vertex v_0 to the vertex set

$\mathscr{F}(\tau_6)$



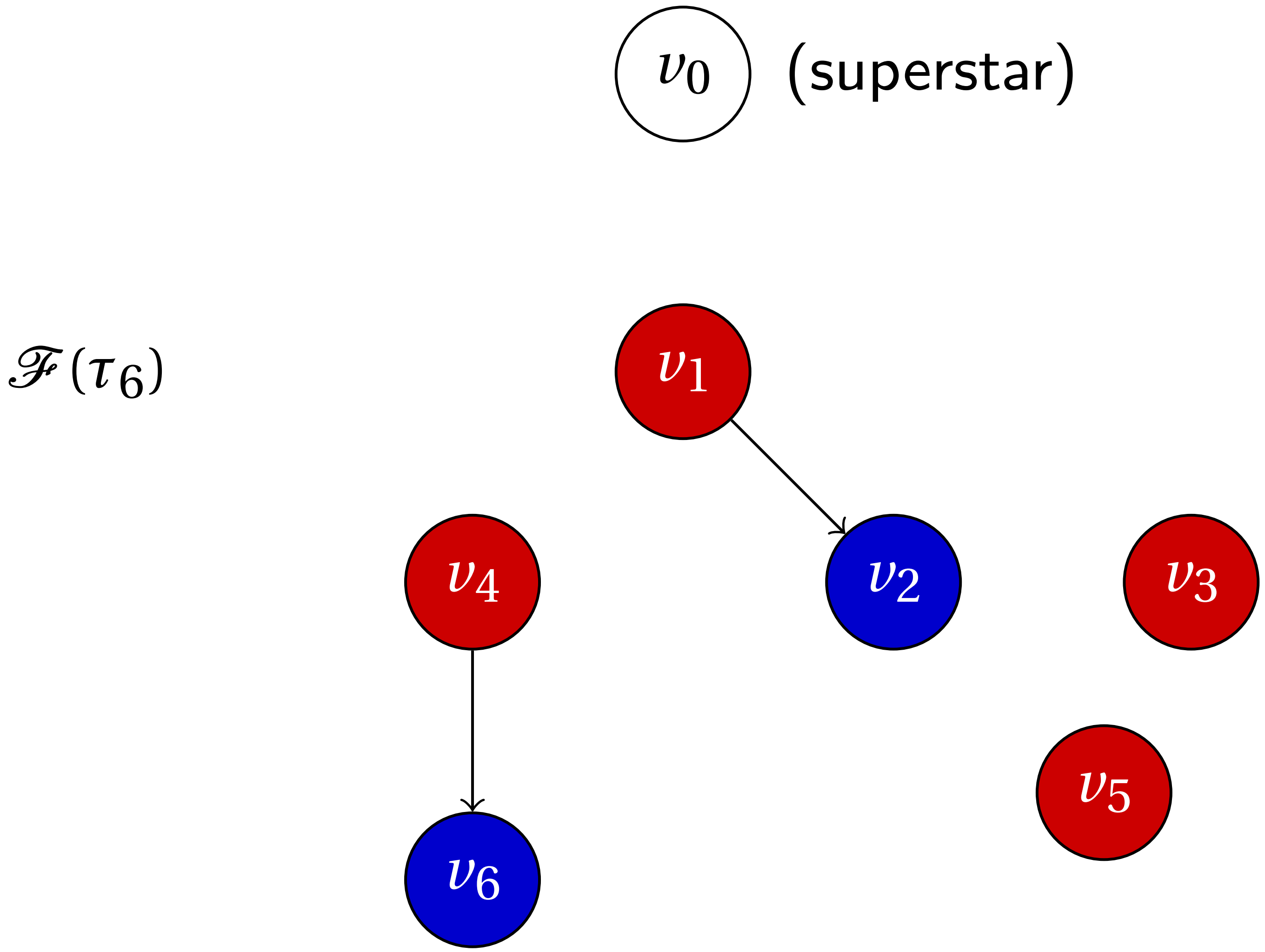
Surgery: From BP Model to Superstar Model

- 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



Surgery: From BP Model to Superstar Model

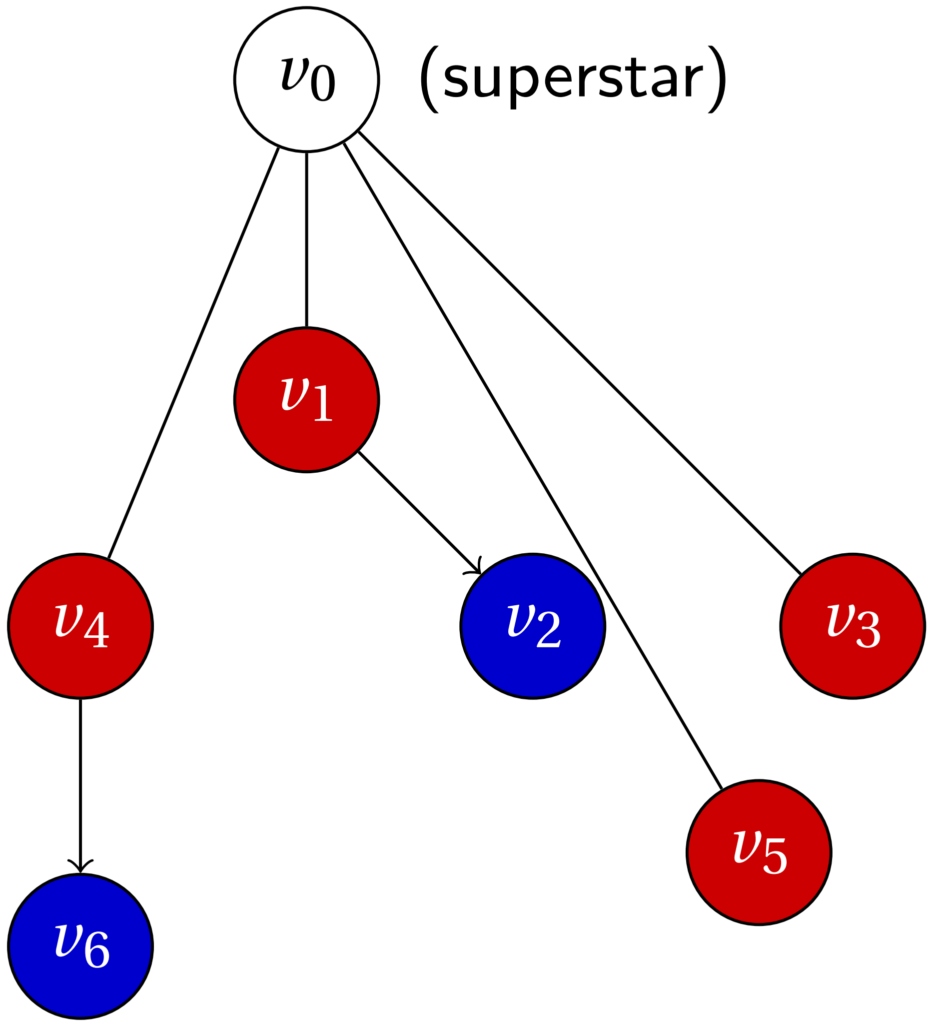
- Add an exogenous superstar vertex v_0 to the vertex set
- For each red vertex remove the edge from parent and create an undirected edge to the superstar vertex v_0



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

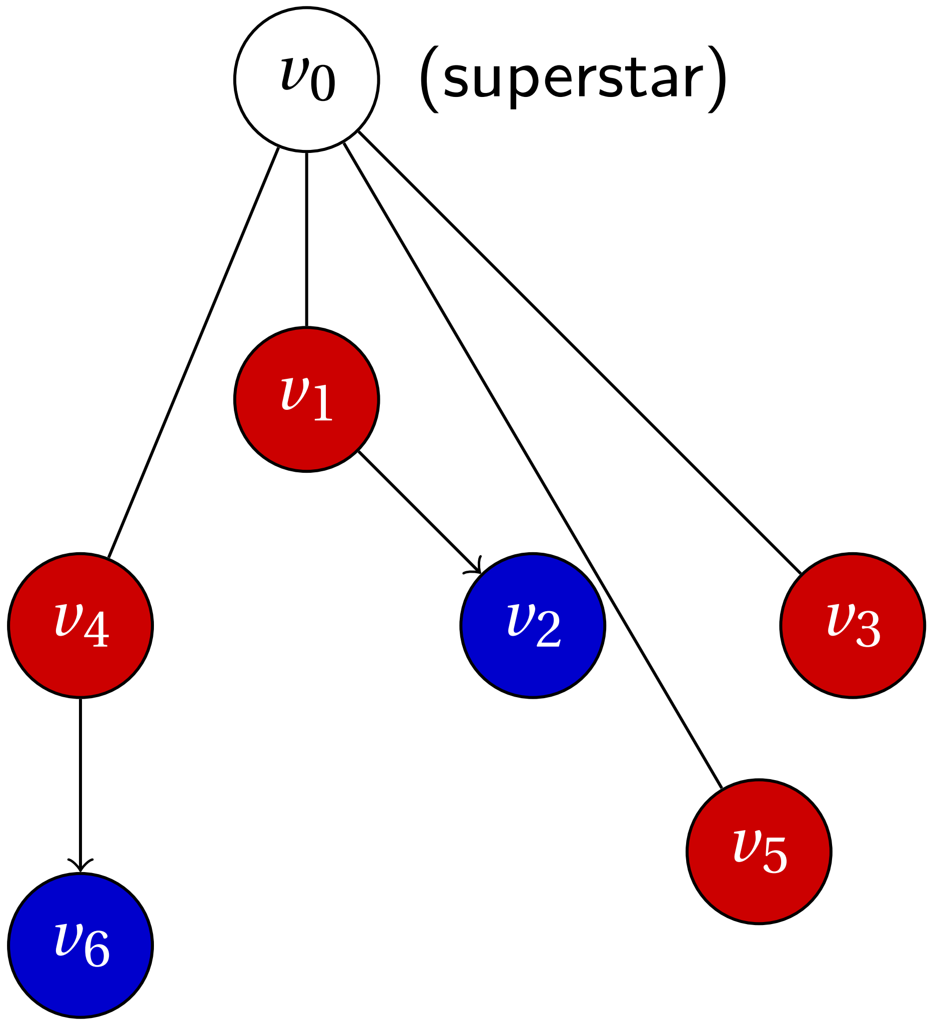
$\mathscr{F}(\tau_6)$



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

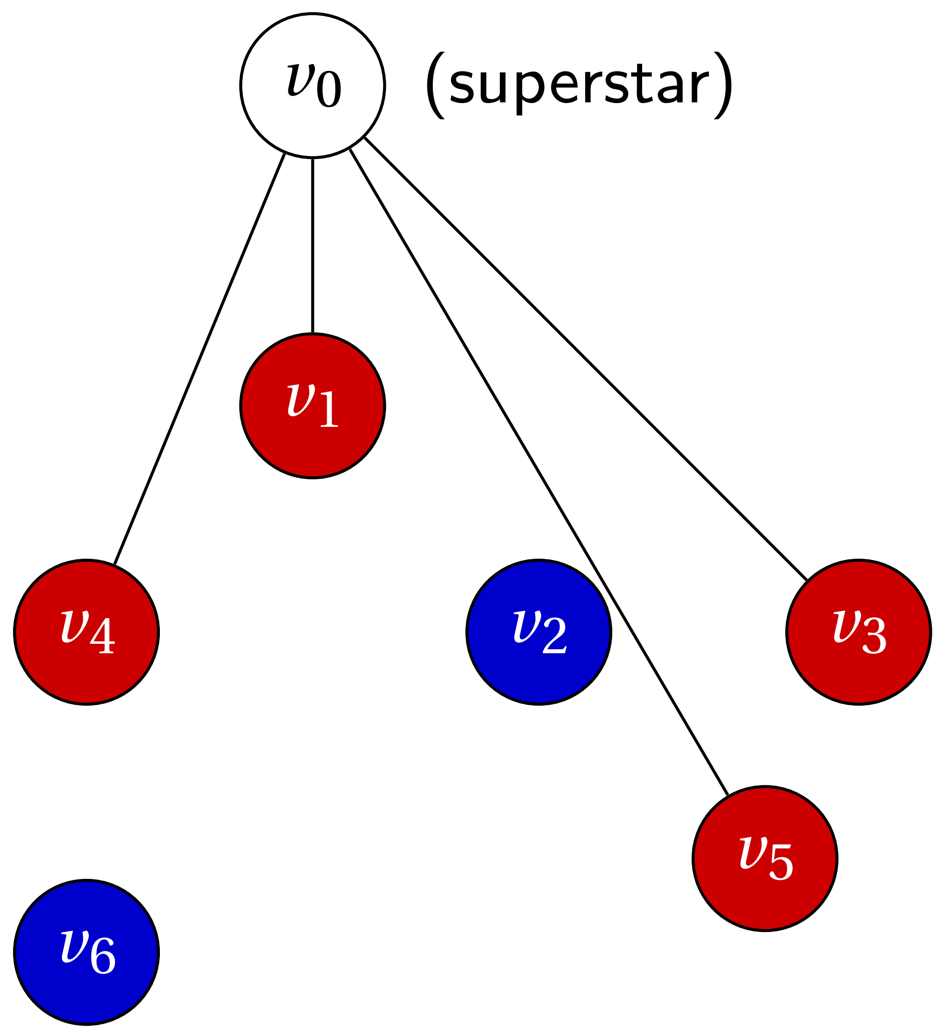
$\mathscr{F}(\tau_6)$



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

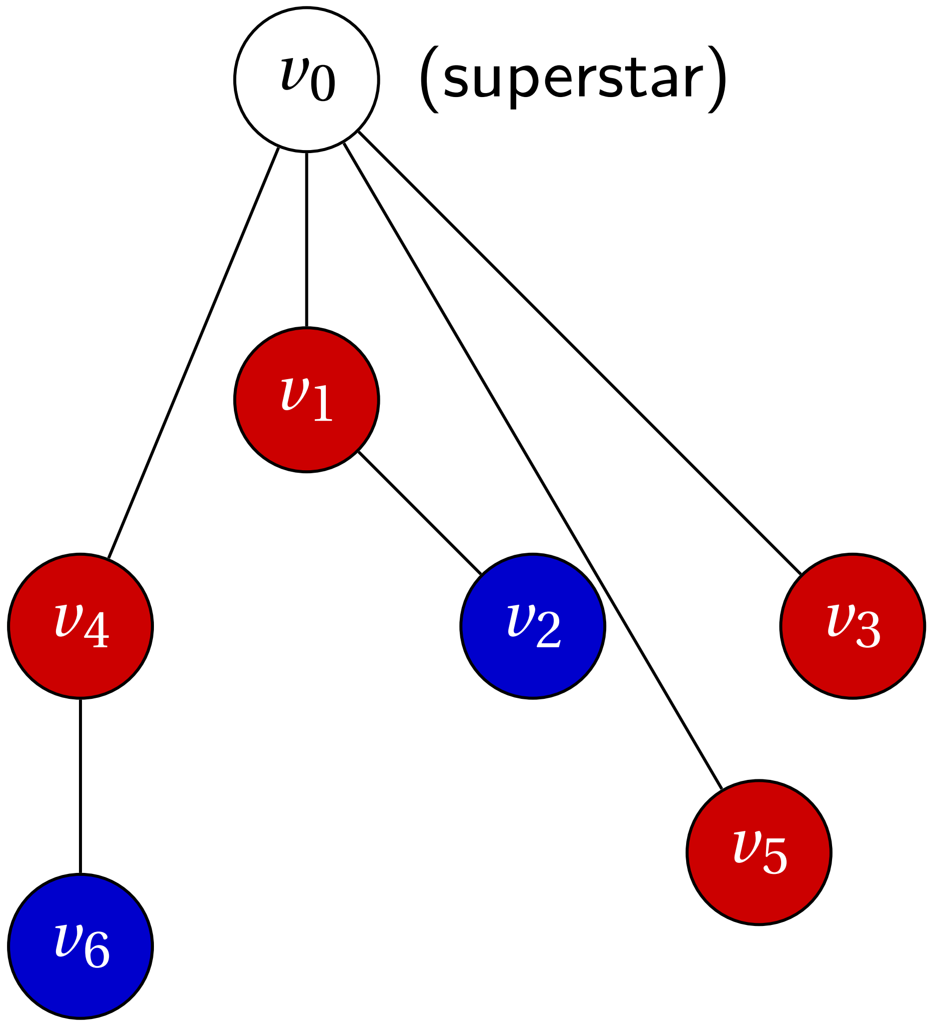
$\mathscr{F}(\tau_6)$



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

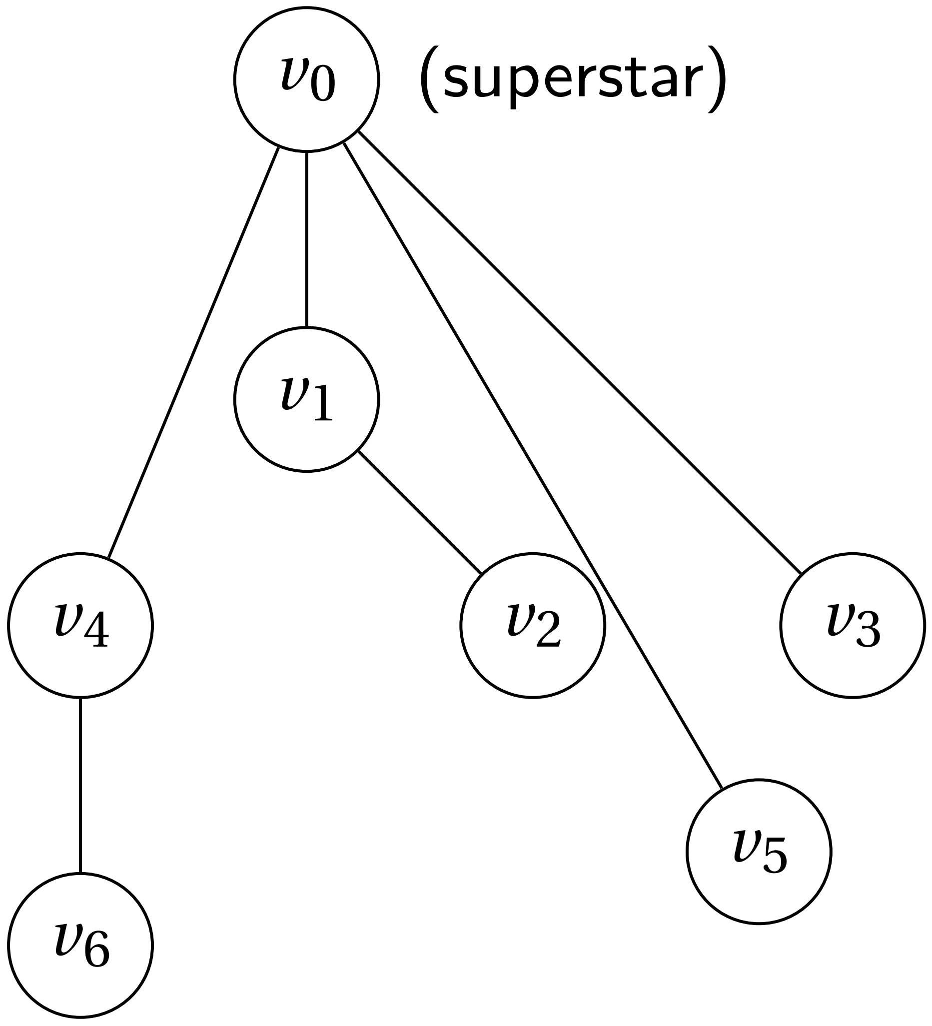
$\mathscr{F}(\tau_6)$



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

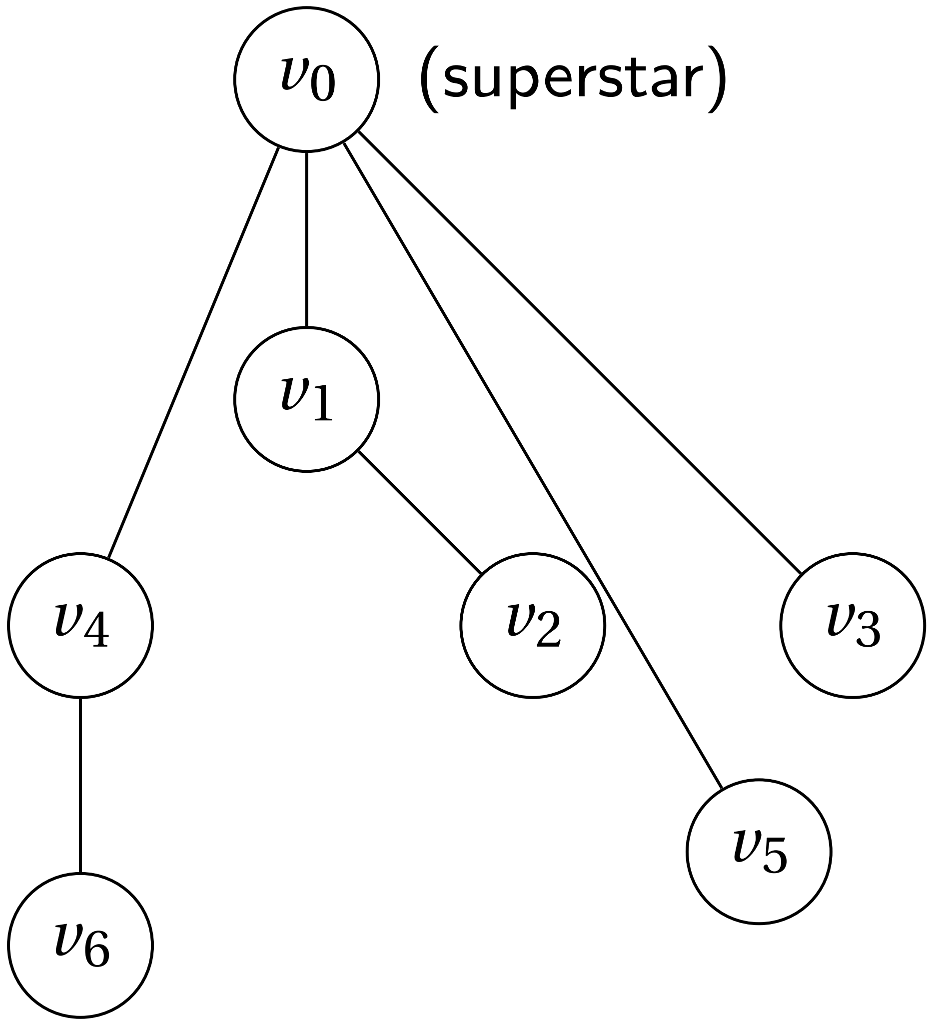
$\mathscr{F}(\tau_6)$



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

$S(\tau_6)$



Relating the BP Construction with the Superstar Model

Relating the BP Construction with the Superstar Model

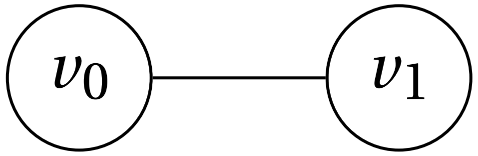
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}

Relating the BP Construction with the Superstar Model

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

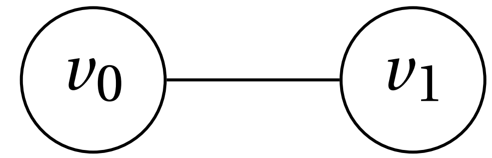
Relating the BP Construction with the Superstar Model

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



Relating the BP Construction with the Superstar Model

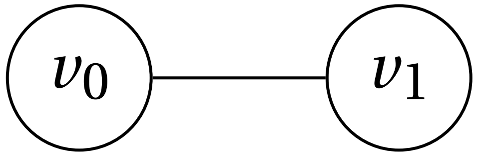
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}
- Base case: $S(\tau_1) = G_2$



- Need to show that $S(\tau_n)$ and G_{n+1} have same probabilistic evolution

Relating the BP Construction with the Superstar Model

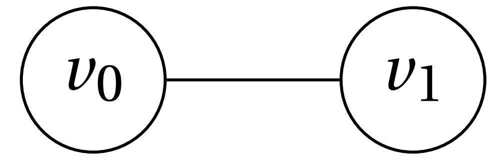
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}
- Base case: $S(\tau_1) = G_2$



- 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

Relating the BP Construction with the Superstar Model

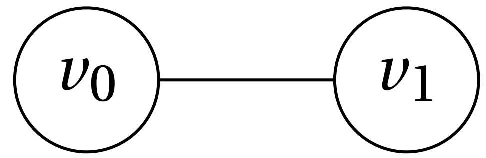
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}
- Base case: $S(\tau_1) = G_2$



- 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

Relating the BP Construction with the Superstar Model

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

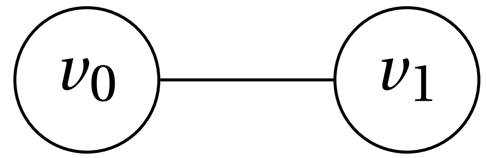


- Need to show that $S(\tau_n)$ and G_{n+1} have same probabilistic evolution
- Superstar: probability of joining superstar = probability of red vertex being born = p
- Same probability for S and G
- Non-superstars: degree of vertex = number of blue children + 1

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

Relating the BP Construction with the Superstar Model

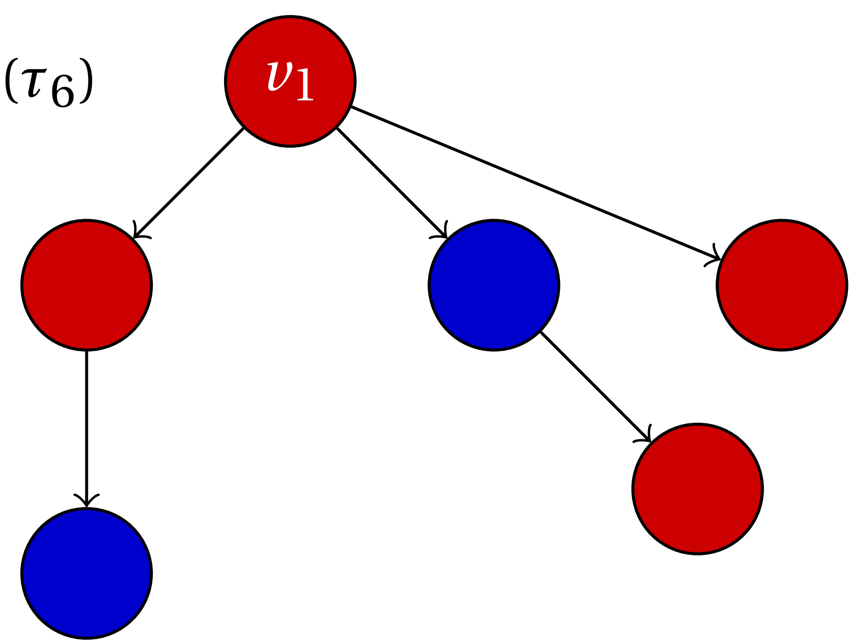
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}
- Base case: $S(\tau_1) = G_2$

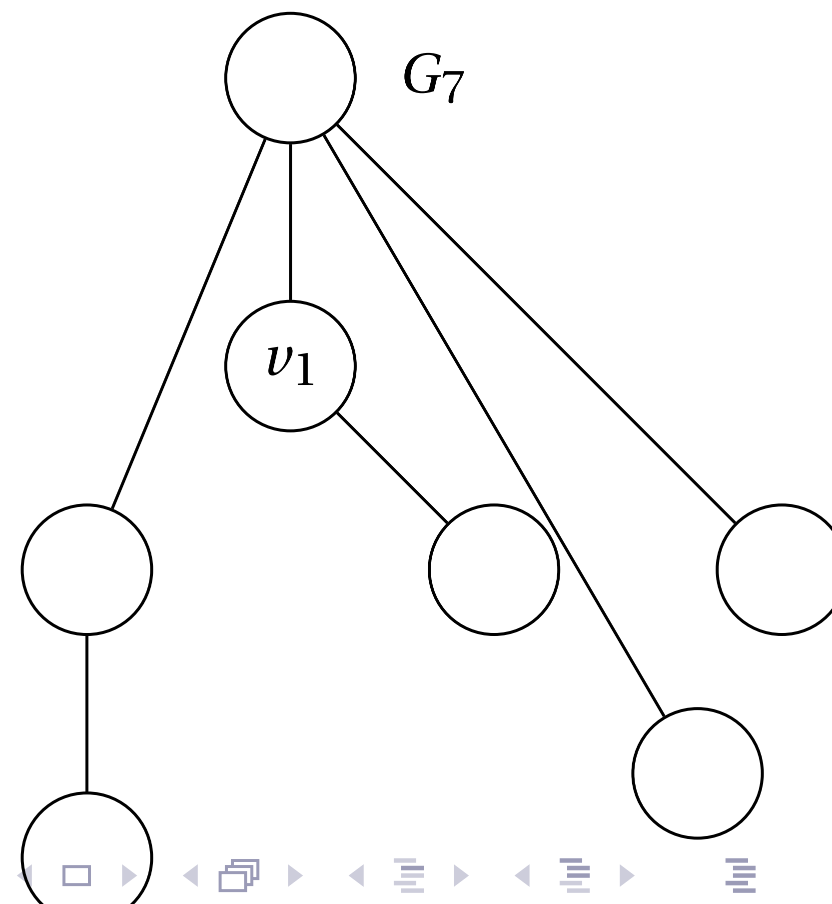


- Need to show that $S(\tau_n)$ and G_{n+1} have same probabilistic evolution
- Superstar: probability of joining superstar = probability of red vertex being born = p
- Same probability for S and G
- Non-superstars: degree of vertex = number of blue children + 1

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

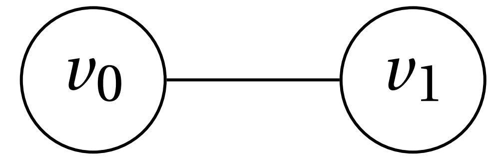
$\mathscr{F}(\tau_6)$





Relating the BP Construction with the Superstar Model

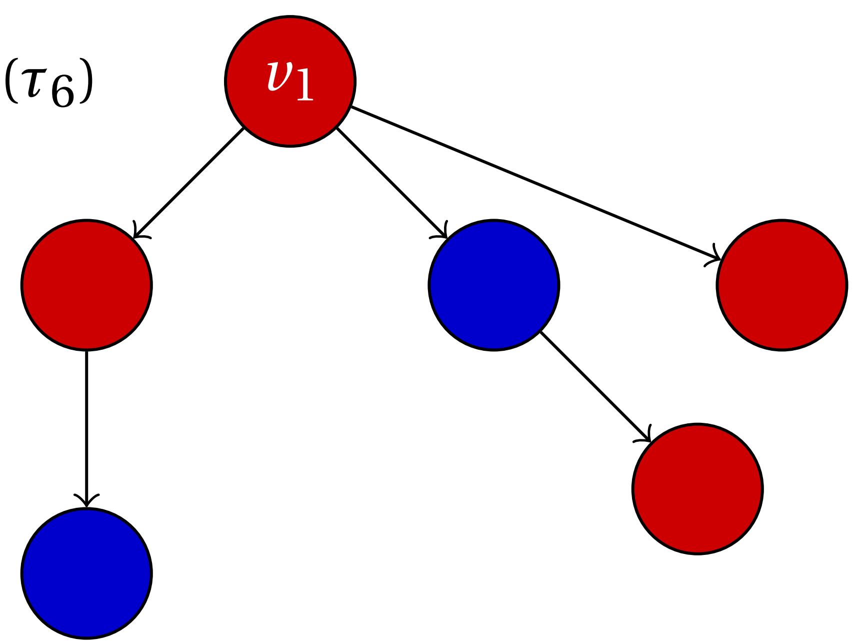
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}
- Base case: $S(\tau_1) = G_2$

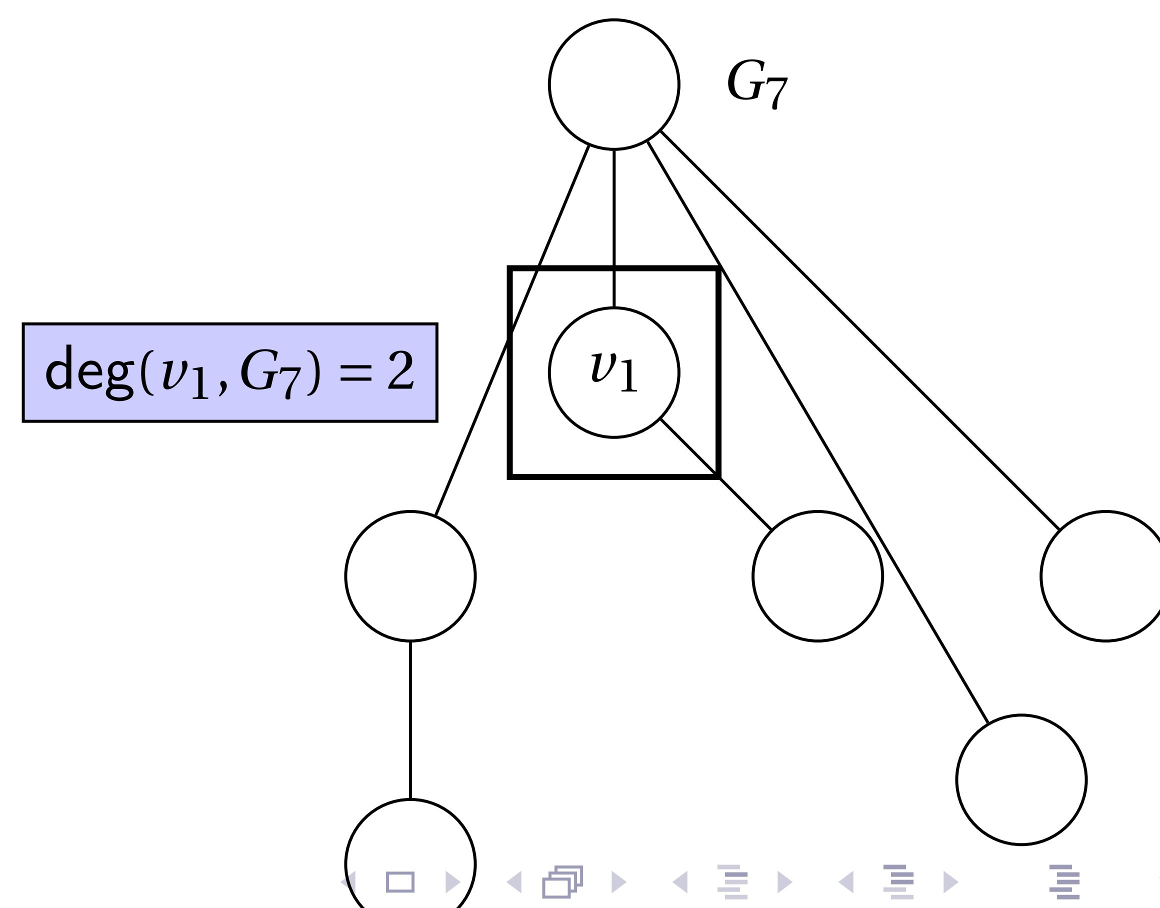


- Need to show that $S(\tau_n)$ and G_{n+1} have same probabilistic evolution
- Superstar: probability of joining superstar = probability of red vertex being born = p
- Same probability for S and G
- Non-superstars: degree of vertex = number of blue children + 1

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

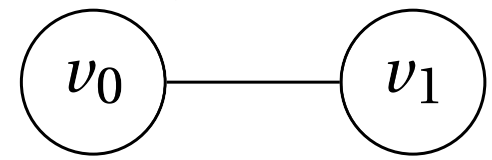
$\mathscr{F}(\tau_6)$





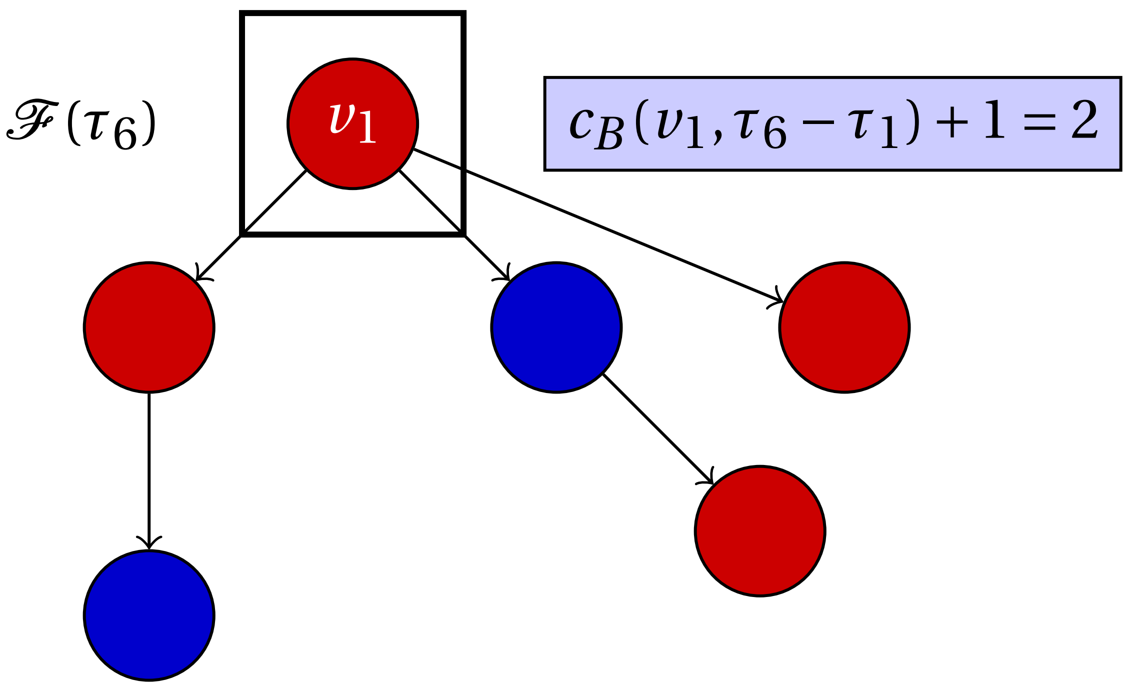
Relating the BP Construction with the Superstar Model

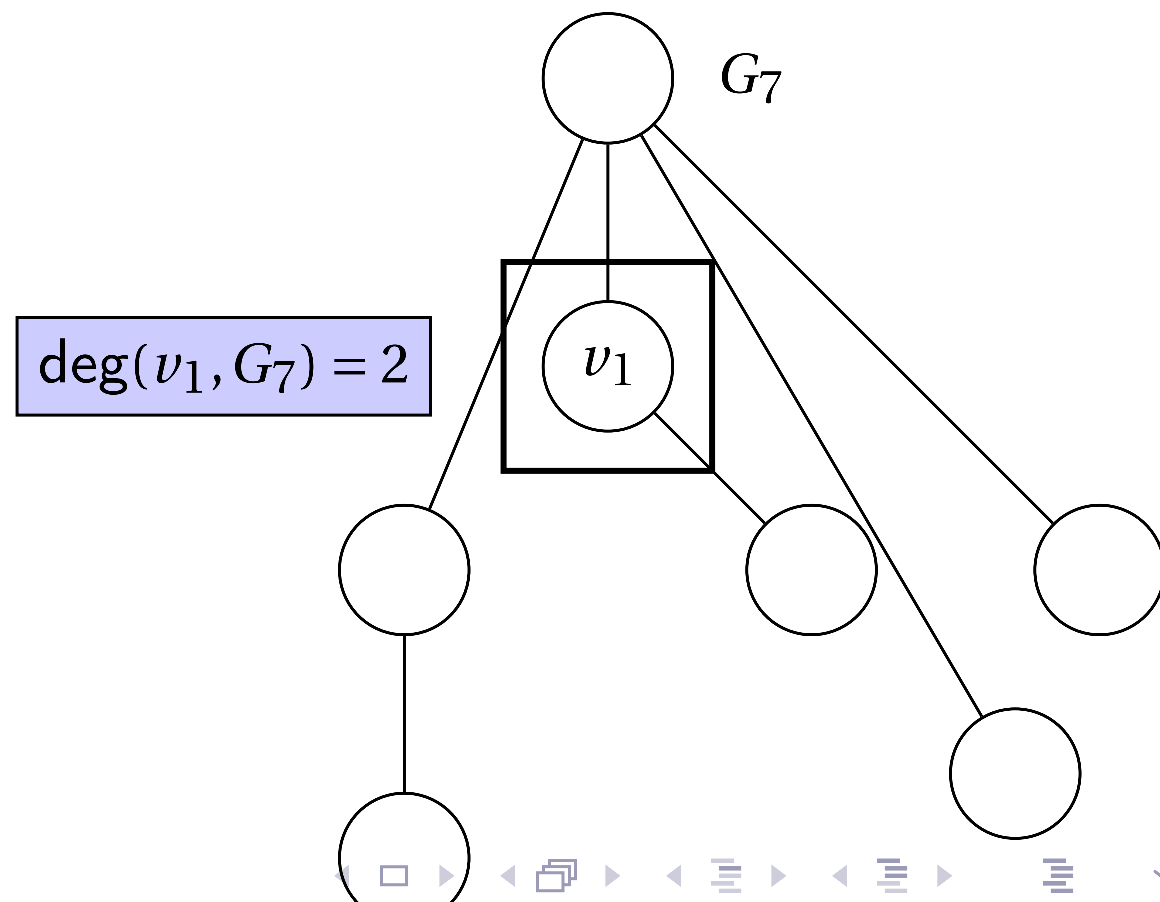
- Claim: $S(\tau_n)$ is "probabilistically the same" as G_{n+1}
- Base case: $S(\tau_1) = G_2$



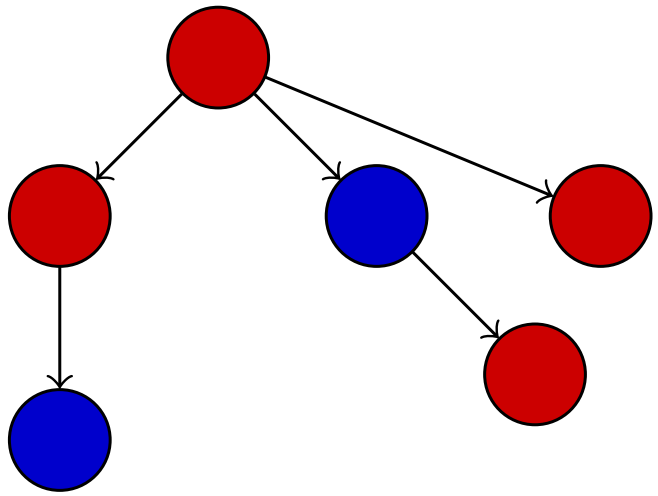
- Need to show that $S(\tau_n)$ and G_{n+1} have same probabilistic evolution
- Superstar: probability of joining superstar = probability of red vertex being born = p
- Same probability for S and G
- Non-superstars: degree of vertex = number of blue children + 1

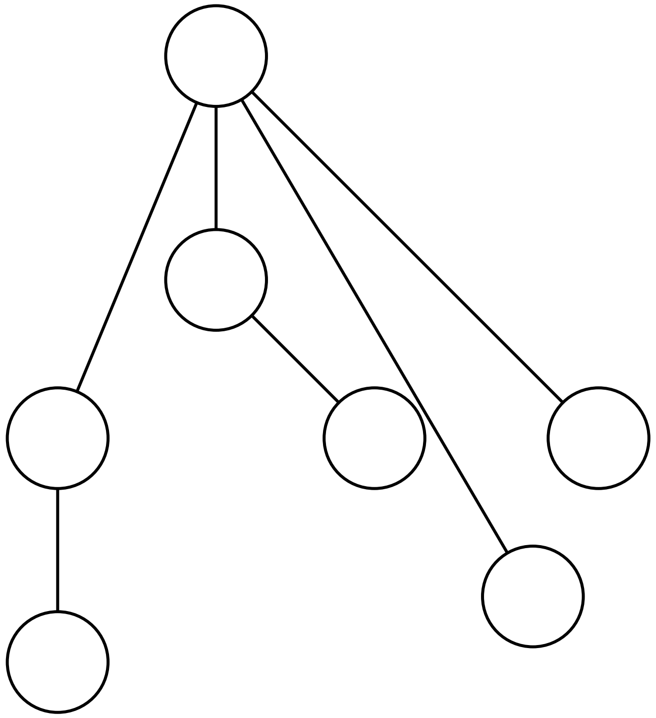
$$\deg(v_k, G_{n+1}) = c_B(v_k, \tau_n - \tau_k) + 1$$



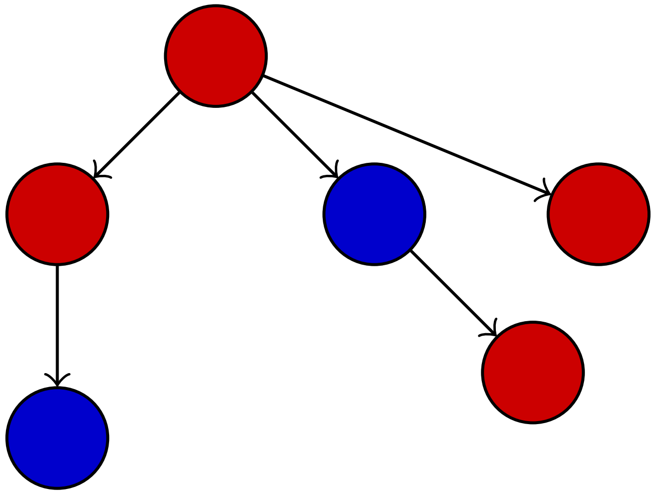


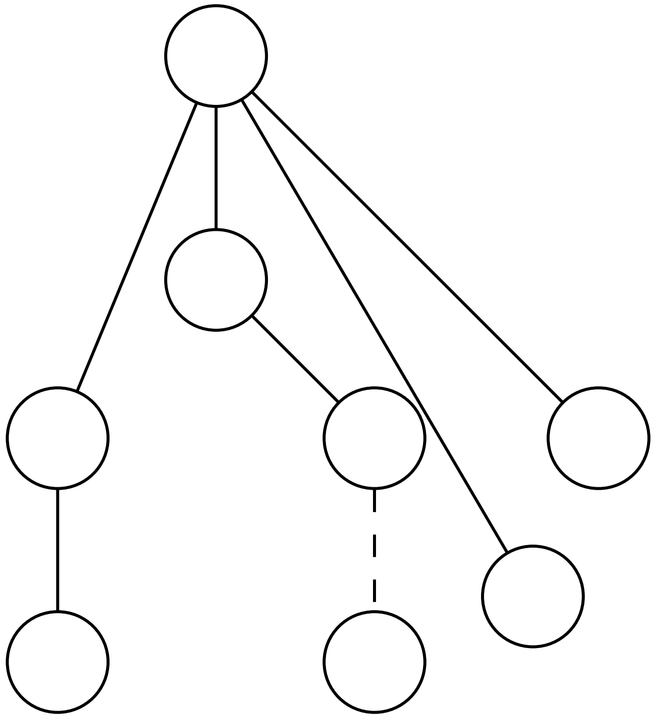
Further Linking of the BP Model and the Superstar Model



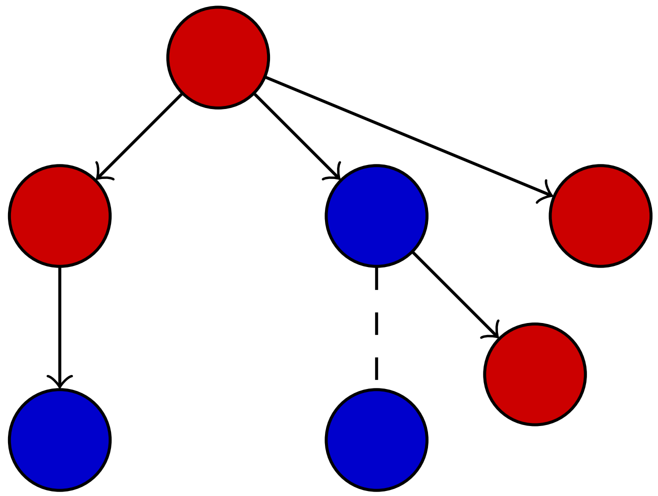


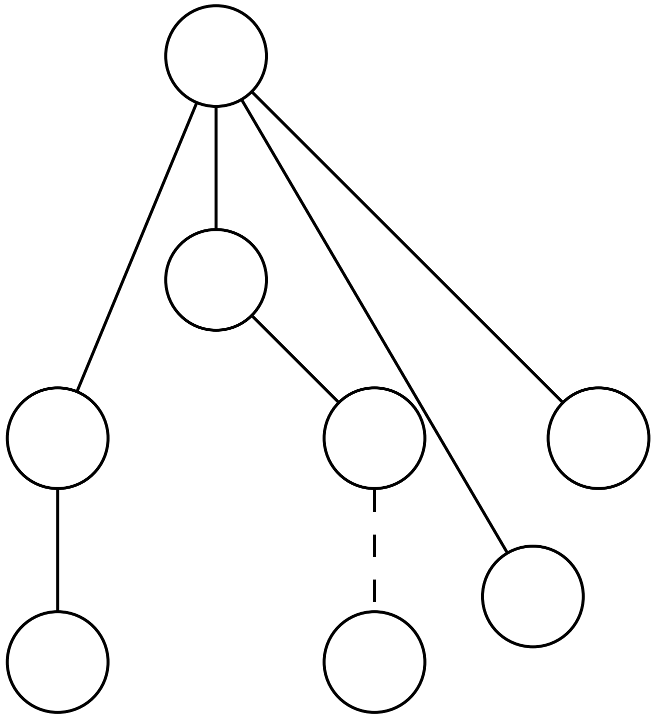
Further Linking of the BP Model and the Superstar Model



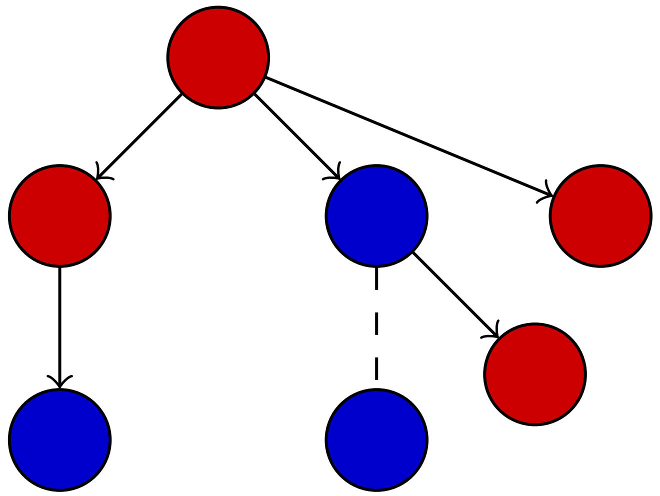


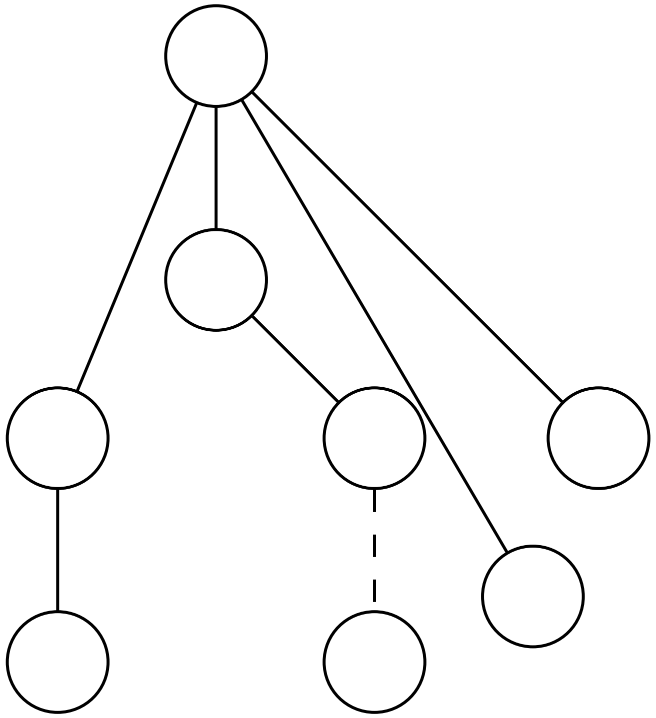
Further Linking of the BP Model and the Superstar Model





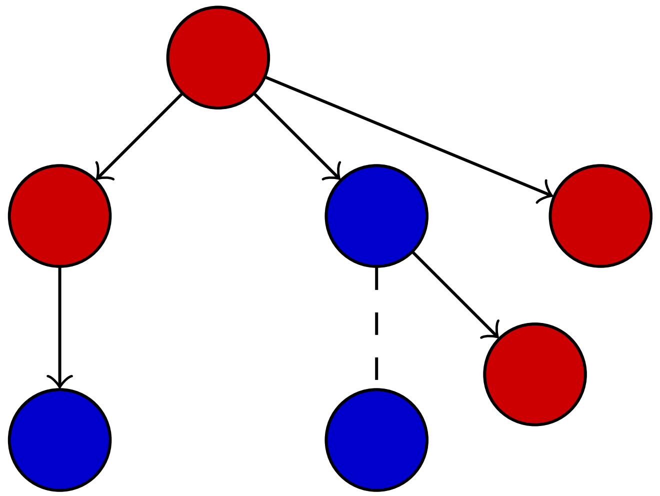
Further Linking of the BP Model and the Superstar Model

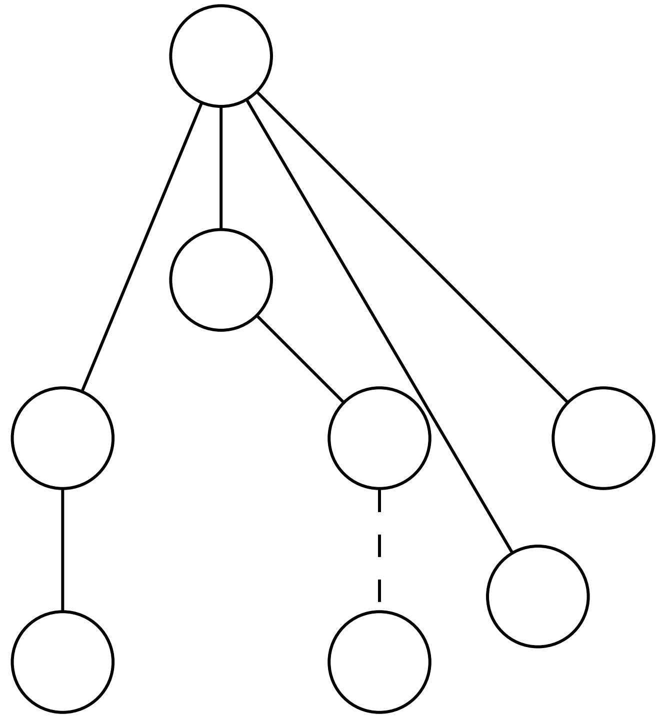




$$\mathbb{P}\left(v_n \text{ joins } v_k | G_n\right) \quad = \quad \mathbb{P}\left(v_n \text{ is blue and born to } v_k | \mathscr{F}(\tau_{n-1})\right)$$

Further Linking of the BP Model and the Superstar Model

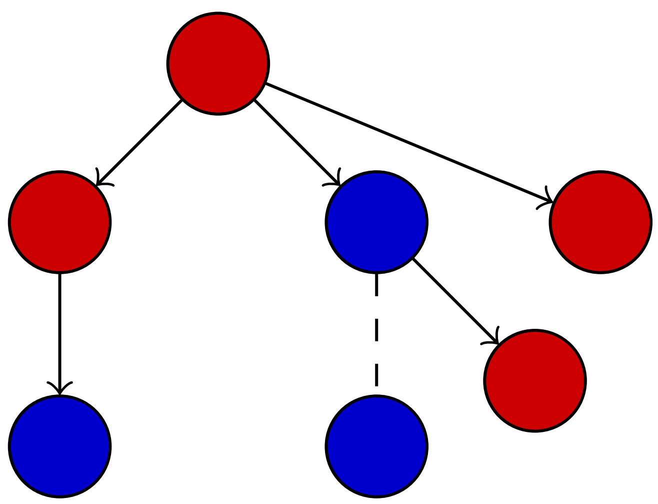


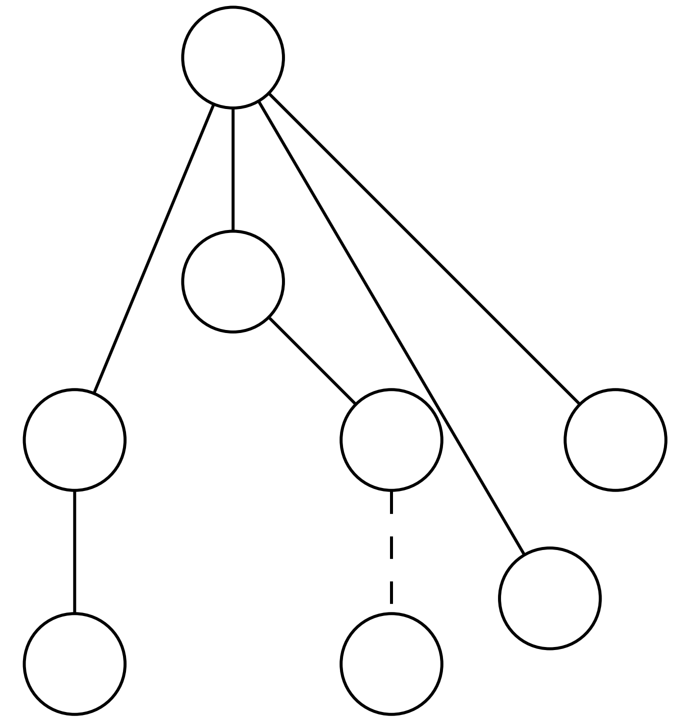


$$\mathbb{P}\left(v_n \text{ joins } v_k | G_n\right) = \mathbb{P}\left(v_n \text{ is blue and born to } v_k | \mathscr{F}(\tau_{n-1})\right)$$

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

Further Linking of the BP Model and the Superstar Model

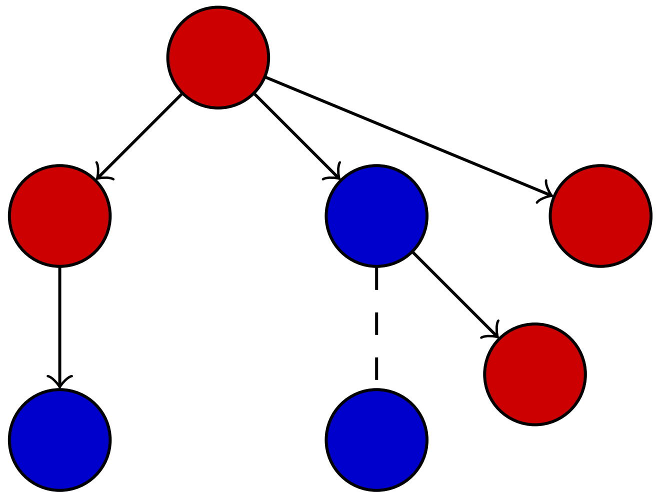


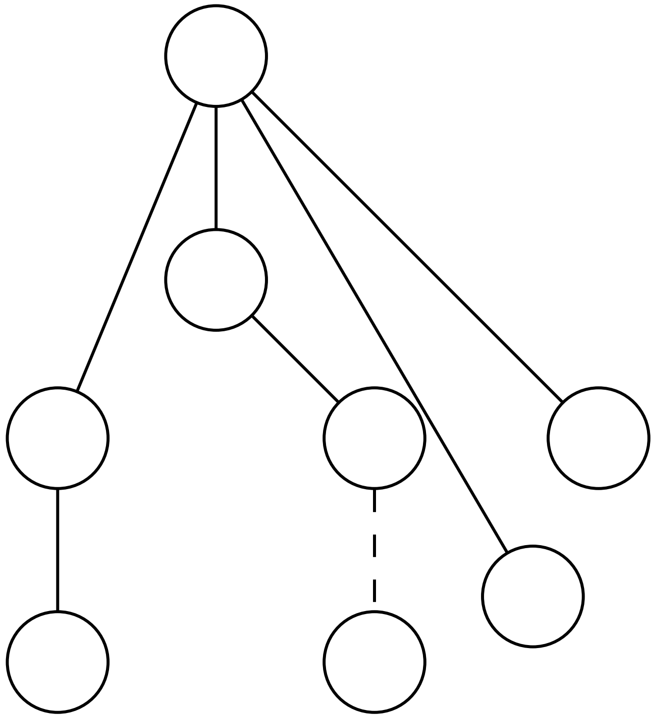


$$\mathbb{P}\left(v_n \text{ joins } v_k | G_n\right) = \mathbb{P}\left(v_n \text{ is blue and born to } v_k | \mathscr{F}(\tau_{n-1})\right)$$

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

Further Linking of the BP Model and the Superstar Model





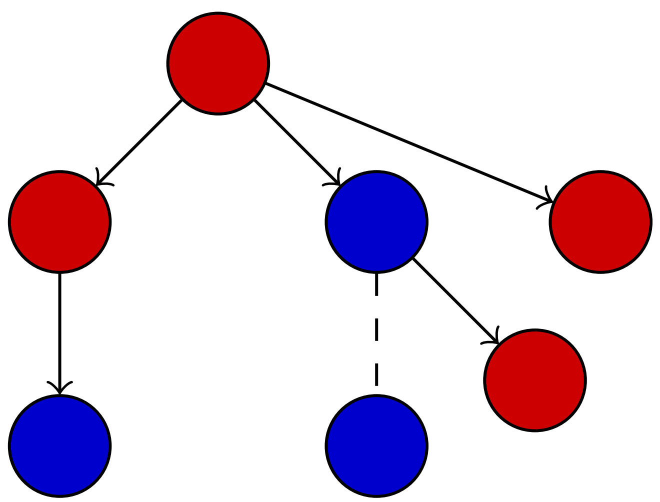
$$\mathbb{P}\left(v_n \text{ joins } v_k | G_n\right) = \mathbb{P}\left(v_n \text{ is blue and born to } v_k | \mathscr{F}(\tau_{n-1})\right)$$

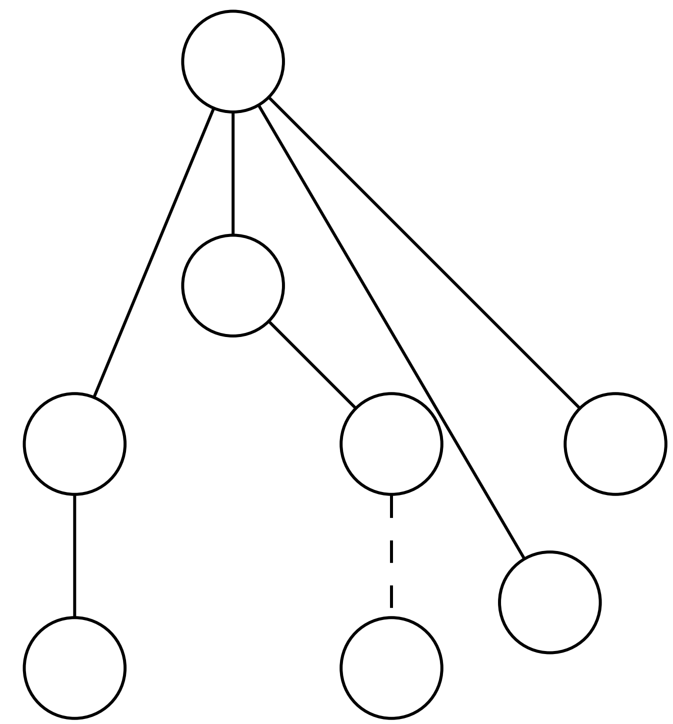
$$\mathbb{P}\left(v_n \text{ joins } v_k | G_n\right) = (1-p)\frac{\deg(v_k, G_n)}{\sum_{v_j \in G_n \setminus v_0} \deg(v_j, G_n)}$$

$$= (1-p)\frac{\deg(v_k, G_n)}{2(n-1) - \deg(v_0, G_n)}$$

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

Further Linking of the BP Model and the Superstar Model





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

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

$$= (1-p)\frac{\deg(v_k, G_n)}{2(n-1) - \deg(v_0, G_n)}$$

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

$\deg(v_k, G_n)$

Power of choice in random trees [D'Souza, Mitzenmacher]

Model: Motivation and construction

- Usual pref. attachment: Basic assumption: every new vertex has knowledge of entire network
- Each stage new vertex chooses 2 vertices uniformly at random
- Connect to vertex with maximal degree **amongst** the ones chosen (breaking ties with probability $1/2$)
- Model which incorporates randomness as well as limited choice
- Let $\mathcal{T}_n$ denote the tree on n vertices

Power of choice in random trees [D'Souza, Mitzenmacher]

Model: Motivation and construction

- Usual pref. attachment: Basic assumption: every new vertex has knowledge of entire network
- Each stage new vertex chooses 2 vertices uniformly at random
- Connect to vertex with maximal degree **amongst** the ones chosen (breaking ties with probability $1/2$)
- Model which incorporates randomness as well as limited choice
- Let $\mathscr{T}_n$ denote the tree on n vertices

Theorem (Angel, Pemantle, SB)

There exists a rooted limiting random tree $\mathscr{T}_\infty$, described by Jagers-Nerman stable age distribution theory such that such that $\mathscr{T}_n$ converges locally $\mathscr{T}_\infty$.

Main idea

- Pick a vertex uniformly at random: Chance it is a leaf in $\mathscr{T}_n$?

Main idea

- Pick a vertex uniformly at random: Chance it is a leaf in $\mathscr{T}_n$?
- Assume stabilizes to some p_0

Main idea

- Pick a vertex uniformly at random: Chance it is a leaf in $\mathcal{T}_n$?
- Assume stabilizes to some p_0
- Number of times queried: Poisson with mean

$$\sum_{i=N+1}^{n} \frac{2}{i} \sim -2\log\frac{N}{n} \approx 2\exp(1)$$

Main idea

- Pick a vertex uniformly at random: Chance it is a leaf in $\mathcal{T}_n$?
- Assume stabilizes to some p_0
- Number of times queried: Poisson with mean

$$\sum_{i=N+1}^{n} \frac{2}{i} \sim -2\log\frac{N}{n} \approx 2\exp(1)$$

- So have interval $[0, T]$ with $T \sim \exp(1/2)$ where queries come at uniform times
- If still a leaf, for each query, no connection made which happens with probability $1 - p_0 + p_0/2 = 1 - p_0/2$.

Main idea

- Pick a vertex uniformly at random: Chance it is a leaf in $\mathscr{T}_n$?
- Assume stabilizes to some p_0
- Number of times queried: Poisson with mean

$$\sum_{i=N+1}^{n} \frac{2}{i} \sim -2\log\frac{N}{n} \approx 2\exp(1)$$

- So have interval $[0,T]$ with $T \sim \exp(1/2)$ where queries come at uniform times
- If still a leaf, for each query, no connection made which happens with probability $1 - p_0 + p_0/2 = 1 - p_0/2$.
- Rate one poisson process, marking each with probability $p_0/2$, time of first point: $X_0 \sim \exp(p_0/2)$
- So probability not a leaf: $1 - p_0 = \mathbb{P}(T > X_0)$

Description of the limit tree

Recursive construction of the degree

- Let p_0 limiting fraction of leaves
- Define $q_0 = p_0/2$
- Then p_0 obtained by doing the following: Let $T \sim \exp(1/2)$ and $X_0 \sim \exp(q_0)$. Then
$$1 - p_0 = \mathbb{P}(T > X_0)$$
- $$p_0 = \frac{\sqrt{5}-1}{2}$$
- General, having obtained p_k, get p_{k+1} by solving
$$1 - (p_0 + \cdots + p_{k+1}) = \mathbb{P}(X_0 + \cdots X_{k+1} > T)$$
where
$$X_{k+1} \sim \exp(p_0 + \cdots + p_k + \frac{p_{k+1}}{2})$$

Description of $\mathscr{T}_\infty$

- After having obtained p_i, let $L_i = \sum_{j=0}^{i} X_j$
- Consider the point process $\mathscr{P}_{\max} = (L_0, L_1, \ldots)$
- Define

$$\mu_{\max}(0,t) = \mathbb{E}(\# i : L_i < t)$$

$$\nu_{\max}(dx) = \exp(-\frac{x}{2})\mu(dx)$$

Description of $\mathscr{T}_\infty$

- After having obtained p_i, let $L_i = \sum_{j=0}^{i} X_j$
- Consider the point process $\mathscr{P}_{\max} = (L_0, L_1, \ldots)$
- Define

$$\mu_{\max}(0,t) = \mathbb{E}(\# i : L_i < t)$$

$$\nu_{\max}(dx) = \exp(-\frac{x}{2})\mu(dx)$$

Theorem

- Then $\mathscr{T}_\infty$ is the Jagers-Nerman stable age distribution tree with offspring distribution $\mathscr{P}_{\max}$, age distribution $\exp(1/2)$ and time to nearest ancestor $\nu_{\max}$
- Implies convergence of global functionals as well such as the spectral distribution of adjacency matrix

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).

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).

Next lecture

Back to critical random graphs: suppose we were interested in the metric structure of maximal components. What can we say? Why should one care.