The branching process approximation and applications

(1) Coupling the network to a branching process
(2) Survival of multitype killed branching random walks
(3) Existence and size of the giant component
(4) Empirical component size distribution
(5) Robustness of the network

Coupling the network to a branching process

Take a concave function $f\colon \mathbb{N}\cup\{0\}\to(0,\infty)$ with $f(0)\le 1$ and

$$\Delta f(k) := f(k+1) - f(k) < 1 \text{ for all } k\in\mathbb{N}.$$

Coupling the network to a branching process

Take a concave function $f\colon \mathbb{N}\cup\{0\}\to(0,\infty)$ with $f(0)\leq 1$ and

$$\Delta f(k) := f(k+1) - f(k) < 1 \text{ for all } k \in \mathbb{N}.$$

Model evolution: At time $N = 1$, we have a single vertex labeled 1. In each time step $N \to N+1$ we

- add a new vertex labeled $N+1$, and
- for each $n \leq N$ independently introduce an oriented edge from the new vertex $N+1$ to the old vertex n with probability

$$\frac{f(\text{indegree of } n \text{ at time } N)}{N}.$$

Coupling the network to a branching process

Take a concave function $f\colon \mathbb{N} \cup \{0\} \to (0, \infty)$ with $f(0) \leq 1$ and

$$\Delta f(k) := f(k+1) - f(k) < 1 \text{ for all } k \in \mathbb{N}.$$

Model evolution: At time $N = 1$, we have a single vertex labeled 1. In each time step $N \to N + 1$ we

- add a new vertex labeled $N + 1$, and
- for each $n \leq N$ independently introduce an oriented edge from the new vertex $N + 1$ to the old vertex n with probability

$$\frac{f(\text{indegree of } n \text{ at time } N)}{N}.$$

All edges are ordered from the younger to the older vertex. For the questions of interest, edges may be considered as unordered.

Coupling the network to a branching process

We ask

- For which attachment functions f is there a giant component?
- What proportion of vertices lies in the giant component?
- For which attachment functions f is the network robust?
- What is the asymptotic behaviour of the empirical component size distribution?

Coupling the network to a branching process

We ask

- For which attachment functions f is there a giant component?
- What proportion of vertices lies in the giant component?
- For which attachment functions f is the network robust?
- What is the asymptotic behaviour of the empirical component size distribution?

Example: If $f(k) = \beta$ there is no preferential attachment, the model is a dynamical version of the Erdős-Rényi model first studied by Dubins. In this case Shepp has shown that a giant component exists if and only if $\beta > \frac{1}{4}$.

Coupling the network to a branching process

We ask

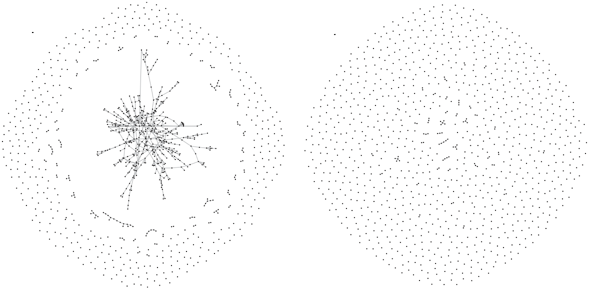
- For which attachment functions f is there a giant component?
- What proportion of vertices lies in the giant component?
- For which attachment functions f is the network robust?
- What is the asymptotic behaviour of the empirical component size distribution?

Example: If $f(k) = \gamma k + \beta$ there is linear preferential attachment. We expect similar behaviour as in the case of preferential attachment with fixed outdegree. In this case Bollobas and Riordan have shown robustness if $\delta = 0$, loosely corresponding to $\gamma = \frac{1}{2}$ in our model.

Coupling the network to a branching process

We ask

- For which attachment functions f is there a giant component?
- What proportion of vertices lies in the giant component?
- For which attachment functions f is the network robust?
- What is the asymptotic behaviour of the empirical component size distribution?



Simulation of the model with $f(k) = \frac{1}{2}\sqrt{k} + x$, for $x = \frac{2}{5}, \frac{1}{10}$ and 1000 vertices, generated by Christian Mönch using the *Network Workbench Tool*.

Coupling the network to a branching process

We ask

- For which attachment functions f is there a giant component?
- What proportion of vertices lies in the giant component?
- For which attachment functions f is the network robust?
- What is the asymptotic behaviour of the empirical component size distribution?

The answer to these questions are based on an approximation of the neighbourhood of a uniformly chosen vertex by the genealogy of a killed branching random walk. Before stating our results, I will describe this approximation.

Coupling the network to a branching process

Particle positions are on the real line and types are given by the relative position of their father. Define the pure birth process $(Z_t : t \geq 0)$ by its generator

$$Lg(k) = f(k)\,\Delta g(k).$$

Coupling the network to a branching process

Particle positions are on the real line and types are given by the relative position of their father. Define the pure birth process $(Z_t \colon t \geq 0)$ by its generator

$$Lg(k) = f(k)\,\Delta g(k).$$

A particle which has its parent to its left generates offspring

- to its right with relative positions at the jumps of the process $(Z_t \colon t \geq 0)$;
- to its left with relative positions distributed according to the Poisson process Π on $(-\infty, 0]$ with intensity measure $e^t\,\mathbb{E}[f(Z_{-t})]\,dt$.

Coupling the network to a branching process

For $\tau > 0$ we let $(Z_t^{(\tau)}: t \geq 0)$ be the pure birth process $(Z_t: t \geq 0)$ conditioned to have a birth at time τ.

Coupling the network to a branching process

For $\tau > 0$ we let $(Z_t^{(\tau)} \colon t \geq 0)$ be the pure birth process $(Z_t \colon t \geq 0)$ conditioned to have a birth at time τ.

A particle which has its parent at distance τ to its right generates offspring

- to its right with relative positions at the jumps of $(Z_t^{(\tau)} - \mathbf{1}[\tau, \infty)(t) \colon t \geq 0)$.
- to its left in the same manner as before.

Coupling the network to a branching process

For $\tau > 0$ we let $(Z_t^{(\tau)} \colon t \geq 0)$ be the pure birth process $(Z_t \colon t \geq 0)$ conditioned to have a birth at time τ.

A particle which has its parent at distance τ to its right generates offspring

- to its right with relative positions at the jumps of $(Z_t^{(\tau)} - \mathbf{1}[\tau, \infty)(t) \colon t \geq 0)$.
- to its left in the same manner as before.

We start the branching random walk with one initial particle in location $-X$, where X is standard exponential and kill particles and their offspring if their position is to the right of the origin.

Coupling the network to a branching process

Denote by

- T the total number of individuals in the killed branching random walk,
- $C_N(v)$ the size of the component in the network containing the vertex v.

Coupling the network to a branching process

Denote by

- T the total number of individuals in the killed branching random walk,
- $C_N(v)$ the size of the component in the network containing the vertex v.

Proposition 1

Suppose that (c_N) is a sequence of integers with

$$\lim_{N\to\infty} \frac{c_N}{\log N \log\log N} = 0.$$

Then one can couple

- the network with N vertices together with a uniformly chosen vertex V, and
- the killed branching random walk

such that, with probability tending to one,

$$C_N(V) \wedge c_N = T \wedge c_N.$$

Coupling the network to a branching process

From the branching process approximation we get a first criterion for the existence and size of a giant component.

Coupling the network to a branching process

From the branching process approximation we get a first criterion for the existence and size of a giant component.

Theorem 3

The proportion of vertices in the largest component of the network converges to the survival probability $p(f)$ of the killed branching random walk, while the proportion of vertices in the second largest component converges to zero, in probability.

In particular, there exists a giant component if and only if the killed branching random walk is supercritical, i.e. $p(f) > 0$.

Coupling the network to a branching process

From the branching process approximation we get a first criterion for the existence and size of a giant component.

Theorem 3

The proportion of vertices in the largest component of the network converges to the survival probability $p(f)$ of the killed branching random walk, while the proportion of vertices in the second largest component converges to zero, in probability.

In particular, there exists a giant component if and only if the killed branching random walk is supercritical, i.e. $p(f) > 0$.

- Although the branching process approximation is only local we can derive a global result. Crucial for this is a sprinkling argument.

Coupling the network to a branching process

From the branching process approximation we get a first criterion for the existence and size of a giant component.

Theorem 3

The proportion of vertices in the largest component of the network converges to the survival probability $p(f)$ of the killed branching random walk, while the proportion of vertices in the second largest component converges to zero, in probability.

In particular, there exists a giant component if and only if the killed branching random walk is supercritical, i.e. $p(f) > 0$.

- Although the branching process approximation is only local we can derive a global result. Crucial for this is a sprinkling argument.
- Real networks do not look locally like trees, they are more clustered.

Coupling the network to a branching process

This is the sprinkling argument required:

Proposition 2

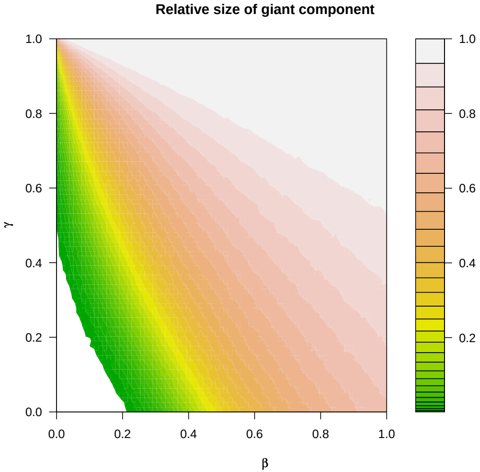
Let $0 < \varepsilon < f(0)$ and let $(\mathcal{G}_N^{(\varepsilon)})$ be the preferential attachment graphs with attachment rule $f - \varepsilon$, and $C_N^{(\varepsilon)}(v)$ the size of the component in $\mathcal{G}_N^{(\varepsilon)}$ containing the vertex v.
If, with high probability,

$$\sum_{v=1}^{N} \mathbf{1}\{C_N^{(\varepsilon)}(v) \geq c_N\} \geq \kappa N$$

then there exists a coupling of $\mathcal{G}_N$ with $\mathcal{G}_N^{(\varepsilon)}$ such that

- $\mathcal{G}_N^{(\varepsilon)} \leq \mathcal{G}_N$, and
- all connected components of $\mathcal{G}_N^{(\varepsilon)}$ with at least c_N vertices belong to one connected component in $\mathcal{G}_N$ with at least κN vertices.

Coupling the network to a branching process



Simulation for the linear case $f(k) = \gamma k + \beta$.

Survival of multitype killed branching random walks

Survival criteria for multitype killed branching random walks can be expressed in terms of the principal eigenvalue of an associated operator on a compact type space.

Survival of multitype killed branching random walks

Survival criteria for multitype killed branching random walks can be expressed in terms of the principal eigenvalue of an associated operator on a compact type space.

Describe the type space as

$$S := \{\ell\} \cup [0, \infty],$$

and let M_τ be the intensity measure of the spatial offspring distribution on the real line, for a particle of type $\tau \in S$.

Survival of multitype killed branching random walks

Survival criteria for multitype killed branching random walks can be expressed in terms of the principal eigenvalue of an associated operator on a compact type space.

Describe the type space as

$$S := \{\ell\} \cup [0, \infty],$$

and let M_τ be the intensity measure of the spatial offspring distribution on the real line, for a particle of type $\tau \in S$.

Given $0 < \alpha < 1$ we define a score operator A_α on the Banach space $C(S)$ by

$$A_\alpha g(\tau) = \int_{-\infty}^{0} g(-t) e^{-\alpha t}\, M_\tau(dt) + \int_{0}^{\infty} g(\ell)\, e^{-\alpha t}\, M_\tau(dt).$$

Survival of multitype killed branching random walks

Survival criteria for multitype killed branching random walks can be expressed in terms of the principal eigenvalue of an associated operator on a compact type space.

Describe the type space as

$$S := \{\ell\} \cup [0, \infty],$$

and let M_τ be the intensity measure of the spatial offspring distribution on the real line, for a particle of type $\tau \in S$.

Given $0 < \alpha < 1$ we define a score operator A_α on the Banach space $C(S)$ by

$$A_\alpha g(\tau) = \int_{-\infty}^{0} g(-t) e^{-\alpha t} \, M_\tau(dt) + \int_{0}^{\infty} g(\ell) \, e^{-\alpha t} \, M_\tau(dt).$$

As $M_\tau \leq M_{\tau'}$ for all $\tau \geq \tau' \geq 0$ the value $A_\alpha g(\infty)$ can be defined by taking a limit.

Survival of multitype killed branching random walks

The operator A_α is a well-defined compact operator iff $A_\alpha 1(0) < \infty$.

Survival of multitype killed branching random walks

The operator A_α is a well-defined compact operator iff $A_\alpha 1(0) < \infty$.

Proposition 2

The killed branching random walk suffers almost sure extinction iff there exists $0 < \alpha < 1$ such that A_α is a well-defined operator with spectral radius $\rho(A_\alpha) \leq 1$.

Survival of multitype killed branching random walks

The operator A_α is a well-defined compact operator iff $A_\alpha 1(0) < \infty$.

Proposition 2

The killed branching random walk suffers almost sure extinction iff there exists $0 < \alpha < 1$ such that A_α is a well-defined operator with spectral radius $\rho(A_\alpha) \leq 1$.

Proof of sufficiency:

- There exists a positive eigenvector $v \in C(S)$ corresponding to the principal eigenvalue $\rho(A_\alpha) \leq 1$.
- Starting with one particle of type τ, the score at generation n given by

$$X_n := \sum_{\substack{\text{particles at } x \\ \text{of type } t}} e^{-\alpha x} \frac{v(t)}{v(\tau)}$$

  is a nonnegative supermartingale, which converges almost surely.
- Hence the position of the leftmost particle diverges to $+\infty$, and the killed branching process dies out almost surely.

Survival of multitype killed branching random walks

Proof of necessity:

- Fix α and let v, ν be the principal eigenvectors of A_α and its dual operator.
- Starting with one particle of type τ, the process

$$W_n^{(\tau)} := \rho(A_\alpha)^{-n} \sum_{\substack{\text{particles at } x \\ \text{of type } t}} e^{-\alpha x} \frac{v(t)}{v(\tau)}$$

  is a nonnegative martingale, converging almost surely to some $W^{(\tau)}$.
- If $\rho(A_\alpha) > 1$ for all α, then $W^{(\tau)} > 0$ almost surely, for some α.
- Almost surely with respect to

$$dQ = \int \nu(d\tau)\, v(\tau)\, W_\tau\, dP_\tau,$$

  picking a particle in generation n according to

$$\mu(x_n) = \rho(A_\alpha)^{-n} \frac{v(t_n)}{v(t_0)} e^{-\alpha x_n} \frac{W^{(t_n)}(x_n)}{W^{(t_0)}(x_0)}$$

  yields that $\lim x_n/n = -\log \rho(A_\alpha)' < 0$.
- Hence there is a positive probability that an ancestral line goes to $-\infty$ and the killed branching process survives.

Existence and size of the giant component

In the case of a linear attachment rule $f(n) = \gamma n + \beta$ it turns out that the spatial offspring distribution is independent of the numerical value of the type. Hence the type space can be collapsed into $S = \{\ell, r\}$.

Existence and size of the giant component

In the case of a linear attachment rule $f(n) = \gamma n + \beta$ it turns out that the spatial offspring distribution is independent of the numerical value of the type. Hence the type space can be collapsed into $S = \{\ell, r\}$.

Now A_α is well-defined iff $\gamma < \alpha < 1 - \gamma$ and in this case becomes the matrix

$$A = \begin{pmatrix} \frac{\beta}{\alpha-\gamma} & \frac{\beta}{1-\alpha-\gamma} \\ \frac{\beta+\gamma}{\alpha-\gamma} & \frac{\beta}{1-\alpha-\gamma} \end{pmatrix}$$

Existence and size of the giant component

In the case of a linear attachment rule $f(n) = \gamma n + \beta$ it turns out that the spatial offspring distribution is independent of the numerical value of the type. Hence the type space can be collapsed into $S = \{\ell, r\}$.

Now A_α is well-defined iff $\gamma < \alpha < 1 - \gamma$ and in this case becomes the matrix

$$A = \begin{pmatrix} \frac{\beta}{\alpha-\gamma} & \frac{\beta}{1-\alpha-\gamma} \\ \frac{\beta+\gamma}{\alpha-\gamma} & \frac{\beta}{1-\alpha-\gamma} \end{pmatrix}$$

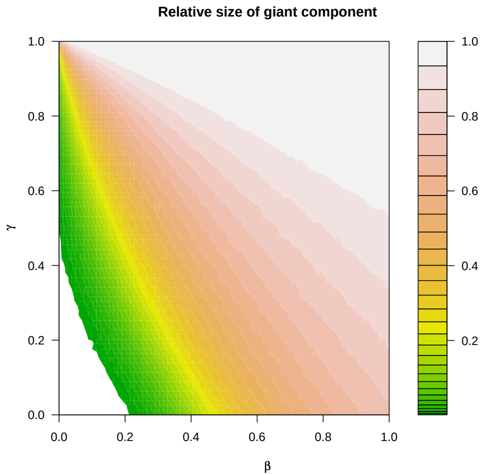
Hence our result becomes completely explicit in the linear case.

Theorem 4

A giant component exists if and only if

$$\gamma \geq \frac{1}{2} \text{ or } \beta > \frac{(\frac{1}{2} - \gamma)^2}{1 - \gamma}.$$

Existence and size of the giant component



Simulation for the linear case $f(k) = \gamma k + \beta$.

Existence and size of the giant component

In the nonlinear case monotone dependence of M_τ on the type allows necessary and sufficient conditions for existence of a giant component, which are often close.

Existence and size of the giant component

In the nonlinear case monotone dependence of M_τ on the type allows necessary and sufficient conditions for existence of a giant component, which are often close.

Theorem 5

If

$$2 \sum_{k=0}^{\infty} \prod_{j=0}^{k} \frac{f(j)}{\frac{1}{2} + f(j)} > 1,$$

then there exists a giant component.

Existence and size of the giant component

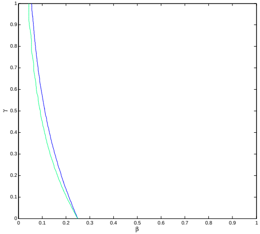
In the nonlinear case monotone dependence of M_τ on the type allows necessary and sufficient conditions for existence of a giant component, which are often close.

Theorem 5

If

$$2\sum_{k=0}^{\infty}\prod_{j=0}^{k}\frac{f(j)}{\frac{1}{2}+f(j)} > 1,$$

then there exists a giant component.



Bounds for the boundary between phases of nonexistence/existence of the giant component for $f(k) = \gamma\sqrt{k} + \beta$ in the (β, γ)–plane.

Empirical component size distribution

Denote by

- T the total number of individuals in the killed branching random walk,
- $C_N(v)$ the size of the component in the network containing the vertex v.

Empirical component size distribution

Denote by

- T the total number of individuals in the killed branching random walk,
- $C_N(v)$ the size of the component in the network containing the vertex v.

Theorem 6

For every $k \in \mathbb{N}$,

$$\frac{1}{N}\sum_{v=1}^{N} \mathbf{1}\{C_N(v) = k\} \longrightarrow P\{T = k\} \quad \text{in probability.}$$

Robustness of the network

Recall that, given $\mathcal{G}_N$ and a deletion parameter $q < 1$ we obtain the percolated network $\mathcal{G}_N(q)$ by removing every edge of $\mathcal{G}_N$ independently with probability q. The network is robust if the giant component survives for every $q < 1$.

Robustness of the network

Recall that, given $\mathcal{G}_N$ and a deletion parameter $q < 1$ we obtain the percolated network $\mathcal{G}_N(q)$ by removing every edge of $\mathcal{G}_N$ independently with probability q. The network is robust if the giant component survives for every $q < 1$.

Theorem 7

For any attachment function f, the network is robust if and only if

$$\gamma := \lim_{n\to\infty} \frac{f(n)}{n} \geq \frac{1}{2}.$$

Robustness of the network

Recall that, given $\mathcal{G}_N$ and a deletion parameter $q < 1$ we obtain the percolated network $\mathcal{G}_N(q)$ by removing every edge of $\mathcal{G}_N$ independently with probability q. The network is robust if the giant component survives for every $q < 1$.

Theorem 7

For any attachment function f, the network is robust if and only if

$$\gamma := \lim_{n\to\infty} \frac{f(n)}{n} \geq \frac{1}{2}.$$

The condition is also equivalent to

$$\tau := \frac{\gamma + 1}{\gamma} \leq 3$$

which confirms the claim of nonrigorous network science.

Robustness of the network

Recall that, given $\mathcal{G}_N$ and a deletion parameter $q < 1$ we obtain the percolated network $\mathcal{G}_N(q)$ by removing every edge of $\mathcal{G}_N$ independently with probability q. The network is robust if the giant component survives for every $q < 1$.

Theorem 7

For any attachment function f, the network is robust if and only if

$$\gamma := \lim_{n\to\infty} \frac{f(n)}{n} \geq \frac{1}{2}.$$

The network is robust if and only if the killed branching random walk has infinite mean growth conditional on survival. This corresponds to the situation that the operator A_α is ill-defined for any $0 < \alpha < 1$.

Robustness of the network

Recall that, given $\mathcal{G}_N$ and a deletion parameter $q < 1$ we obtain the percolated network $\mathcal{G}_N(q)$ by removing every edge of $\mathcal{G}_N$ independently with probability q.
The network is robust if the giant component survives for every $q < 1$.

Theorem 7

For any attachment function f, the network is robust if and only if

$$\gamma := \lim_{n\to\infty} \frac{f(n)}{n} \geq \frac{1}{2}.$$

Robustness was first rigorously verified by Bollobas and Riordan for the preferential attachment model with fixed outdegree and $\delta = 0$, corresponding to the linear case of our model with $\gamma = \frac{1}{2}$.

Robustness of the network

Precise criteria for the existence of a giant component in the percolated network can be given in terms of the operators A_α, and become explicit in the linear case.

Robustness of the network

Precise criteria for the existence of a giant component in the percolated network can be given in terms of the operators A_α, and become explicit in the linear case.

Theorem 8

Suppose f is an attachment function such that the network is not robust. Then the percolated network $\mathcal{G}_N(q)$ has a giant component if and only if

$$q < 1 - \frac{1}{\min_\alpha \rho(A_\alpha)}.$$

In the linear case $f(k) = \gamma k + \beta$, $0 < \gamma < \frac{1}{2}$, the network has a giant component if and only if

$$q < 1 - \left(\frac{1}{2\gamma} - 1\right)\left(\sqrt{1 + \frac{\gamma}{\beta}} - 1\right).$$

Robustness of the network

Precise criteria for the existence of a giant component in the percolated network can be given in terms of the operators A_α, and become explicit in the linear case.

Theorem 8

Suppose f is an attachment function such that the network is not robust. Then the percolated network $\mathcal{G}_N(q)$ has a giant component if and only if

$$q < 1 - \frac{1}{\min_\alpha \rho(A_\alpha)}.$$

In the linear case $f(k) = \gamma k + \beta$, $0 < \gamma < \frac{1}{2}$, the network has a giant component if and only if

$$q < 1 - \left(\tfrac{1}{2\gamma} - 1\right)\left(\sqrt{1 + \tfrac{\gamma}{\beta}} - 1\right).$$

The corresponding problem for preferential attachment models with fixed outdegree seems to be still open.

Preferential attachment networks

A small selection of references:

- Albert, Barabasi, Jeong. Error and attack tolerance of complex networks. Nature 406, 378-382 (2000)
- Biggins. Martingale convergence in the branching random walk. Journal of Applied Probability, 14, 25-37 (1977)
- Bollobas, Riordan. Robustness and vulnerability of scale- free random graphs. Internet Mathematics 1, 1-35 (2003)
- Bollobas, Janson, Riordan. The phase transition in inhomogeneous random graphs. Random Structures & Algorithms 31, 3-122 (2007)
- Dereich, Mörters. Random networks with sublinear preferential attachment: The giant component. Submitted to Annals of Probability.
- Shepp. Connectedness of certain random graphs. Israel Journal of Mathematics 67, 23–33 (1989)