Percolation on preferential attachment networks

Peter Mörters

University of Bath

joint work with Steffen Dereich (Marburg)

Preferential attachment networks

Networks consists of a large, but finite, number of nodes connected by links.
In the modern world, networks are ubiquitous:

- social and communication networks,
- world wide web and internet,
- scientific and other collaboration graphs, . . .

Preferential attachment networks

Networks consists of a large, but finite, number of nodes connected by links.
In the modern world, networks are ubiquitous:

- social and communication networks,
- world wide web and internet,
- scientific and other collaboration graphs, . . .

Preferential attachment is a principle that aims to explain the emergent features of these networks. It was made popular by Barabási and Albert (1999).

Preferential attachment networks

Networks consists of a large, but finite, number of nodes connected by links.
In the modern world, networks are ubiquitous:

- social and communication networks,
- world wide web and internet,
- scientific and other collaboration graphs, . . .

Preferential attachment is a principle that aims to explain the emergent features of these networks. It was made popular by Barabási and Albert (1999).

Networks are built dynamically by adding vertices one-by-one. When a new vertex is introduced, it is linked by edges to a fixed or random number of existing vertices with a probability proportional to an increasing function f of their degree. The higher the degree of a vertex, the more likely it is to establish further links.

Our variant of the model

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

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

Our variant of the model

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

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

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

Our variant of the model

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

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

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

Our variant of the model

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

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

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 $m \le n$ independently introduce an oriented edge from the new vertex $n+1$ to the old vertex m with probability

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

Our variant of the model

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

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

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 $m\le n$ independently introduce an oriented edge from the new vertex $n+1$ to the old vertex m with probability

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

Example:

Our variant of the model

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

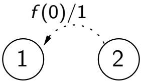
$$\Delta f(k) := f(k+1) - f(k) < 1 \qquad \text{for all } k\in\mathbb{N}.$$

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 $m\le n$ independently introduce an oriented edge from the new vertex $n+1$ to the old vertex m with probability

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

Example:



Our variant of the model

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

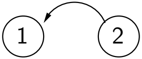
$$\Delta f(k) := f(k+1) - f(k) < 1 \qquad \text{for all } k \in \mathbb{N}.$$

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 $m \le n$ independently introduce an oriented edge from the new vertex $n+1$ to the old vertex m with probability

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

Example:



Our variant of the model

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

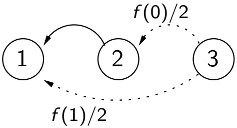
$$\Delta f(k) := f(k+1) - f(k) < 1 \quad \text{for all } k\in\mathbb{N}.$$

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 $m\le n$ independently introduce an oriented edge from the new vertex $n+1$ to the old vertex m with probability

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

Example:



Our variant of the model

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

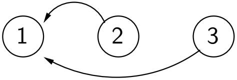
$$\Delta f(k) := f(k+1) - f(k) < 1 \qquad \text{for all } k\in\mathbb{N}.$$

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 $m\le n$ independently introduce an oriented edge from the new vertex $n+1$ to the old vertex m with probability

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

Example:



Our variant of the model

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

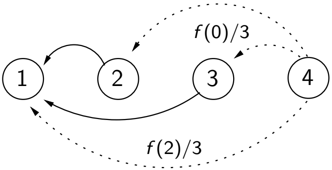
$$\Delta f(k) := f(k+1)-f(k) < 1 \quad \text{for all } k\in\mathbb{N}.$$

Model evolution: At time $n=1$, we have a single vertex (labeled 1). In each time step $n\longrightarrow n+1$ we

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

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

Example:



Our variant of the model

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

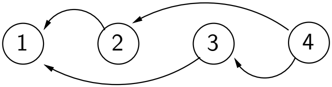
$$\Delta f(k) := f(k+1) - f(k) < 1 \quad \text{for all } k\in\mathbb{N}.$$

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 $m\le n$ independently introduce an oriented edge from the new vertex $n+1$ to the old vertex m with probability

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

Example:



All edges are ordered from the younger to the older vertex. For the questions of interest, edges may be considered as unordered. We denote the resulting increasing sequence of graphs by $(\mathcal{G}_N)$.

Power law exponents

The empirical degree distribution of $\mathcal{G}_N$ is given by

$$X_N(k) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{\text{degree of vertex } i = k\}.$$

Power law exponents

The empirical degree distribution of $\mathcal{G}_N$ is given by

$$X_N(k) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{\text{degree of vertex } i = k\}.$$

Theorem 1: (M, Dereich 2008)

There exists a probability distribution μ such that

$$\lim_{N \uparrow \infty} X_N = \mu \quad \text{in probability.}$$

Moreover, the limit $\gamma := \lim_{k \uparrow \infty} \frac{f(k)}{k}$ exists and, if $\gamma > 0$, the network is scale-free in the sense that

$$\lim_{k \uparrow \infty} \frac{-\log \mu(k)}{\log k} = \frac{1+\gamma}{\gamma} =: \tau < \infty.$$

Power law exponents

The empirical degree distribution of $\mathcal{G}_N$ is given by

$$X_N(k) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{\text{degree of vertex } i = k\}.$$

Theorem 1: (M, Dereich 2008)

There exists a probability distribution μ such that

$$\lim_{N \uparrow \infty} X_N = \mu \qquad \text{in probability.}$$

Moreover, the limit $\gamma := \lim_{k \uparrow \infty} \frac{f(k)}{k}$ exists and, if $\gamma > 0$, the network is scale-free in the sense that

$$\lim_{k \uparrow \infty} \frac{-\log \mu(k)}{\log k} = \frac{1+\gamma}{\gamma} =: \tau < \infty.$$

Physicists believe that many of the emerging properties of the network depend only on τ and not on other features of the model.

Percolation: Definitions

- Let $\mathcal{C}_N \subset \mathcal{G}_N$ be the largest connected component of the graph. We say that $(\mathcal{G}_N)$ has a giant component if

$$\lim_{N\to\infty} \frac{\#\mathcal{C}_N}{N} = p > 0 \quad \text{in probability.}$$

Percolation: Definitions

- Let $\mathcal{C}_N \subset \mathcal{G}_N$ be the largest connected component of the graph. We say that $(\mathcal{G}_N)$ has a giant component if

$$\lim_{N\to\infty} \frac{\#\mathcal{C}_N}{N} = p > 0 \quad \text{in probability.}$$

- Given $\mathcal{G}_N$ and a retention parameter p we obtain the percolated graph $\mathcal{G}_N(p)$ by removing every edge of $\mathcal{G}_N$ independently with probability $q := 1 - p$.

Percolation: Definitions

- Let $\mathcal{C}_N \subset \mathcal{G}_N$ be the largest connected component of the graph. We say that $(\mathcal{G}_N)$ has a giant component if

$$\lim_{N\to\infty} \frac{\#\mathcal{C}_N}{N} = p > 0 \quad \text{in probability.}$$

- Given $\mathcal{G}_N$ and a retention parameter p we obtain the percolated graph $\mathcal{G}_N(p)$ by removing every edge of $\mathcal{G}_N$ independently with probability $q := 1 - p$.
- We say the network survives pecolation with parameter p if and only if the network $(\mathcal{G}_N(p))$ has a giant component.

Percolation: Definitions

- Let $\mathcal{C}_N \subset \mathcal{G}_N$ be the largest connected component of the graph. We say that $(\mathcal{G}_N)$ has a giant component if

$$\lim_{N\to\infty} \frac{\#\mathcal{C}_N}{N} = p > 0 \quad \text{in probability.}$$

- Given $\mathcal{G}_N$ and a retention parameter p we obtain the percolated graph $\mathcal{G}_N(p)$ by removing every edge of $\mathcal{G}_N$ independently with probability $q := 1 - p$.
- We say the network survives pecolation with parameter p if and only if the network $(\mathcal{G}_N(p))$ has a giant component.
- The network $(\mathcal{G}_N)$ is robust if the network survives percolation for every retention parameter $0 < p \leq 1$, i.e. if the critical retention parameter is zero.

Questions:

- For which attachment rules f does a giant component exist?
- For which attachment rules f is the network robust?
- Given a non-robust network, for which p does it survive percolation?

Percolation

Questions:

- For which attachment rules f does a giant component exist?
- For which attachment rules f is the network robust?
- Given a non-robust network, for which p does it survive percolation?

Theorem 2: (M, Dereich 2010)

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

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

Percolation

Questions:

- For which attachment rules f does a giant component exist?
- For which attachment rules f is the network robust?
- Given a non-robust network, for which p does it survive percolation?

Theorem 2: (M, Dereich 2010)

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

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

where τ is the power-law exponent of the network.

Percolation

Precise criteria for existence of a giant component and survival of the network under percolation can be given in terms of the principal eigenvalue of a compact operator. They become explicit if f is linear.

Percolation

Precise criteria for existence of a giant component and survival of the network under percolation can be given in terms of the principal eigenvalue of a compact operator. They become explicit if f is linear.

Theorem 3: (M, Dereich 2010)

Suppose the attachment function is linear, i.e.

$$f(k) = \gamma k + \beta, \qquad \text{with } 0 \leq \gamma < 1.$$

Then a giant component exists if and only if

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

and if $\gamma < \frac{1}{2}$ the network survives percolation with retention parameter p if and only if

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

Framework of the proof

- We couple the neighbourhood of a uniformly chosen vertex with the genealogy of a killed branching random walk. This coupling is successful with high probability if up to $\log N$ vertices are explored.

Framework of the proof

- We couple the neighbourhood of a uniformly chosen vertex with the genealogy of a killed branching random walk. This coupling is successful with high probability if up to $\log N$ vertices are explored.
- A sprinkling argument is used to establish the existence of a giant component from the local information. This means that we consider the approximating trees for a network with a slightly reduced edge density . If they survive percolation with positive probability, the remaining edges will connect the surviving trees with high probability, and hence a giant component exists.

Coupling the network to a branching process

We get a first criterion for the existence and size of a giant component.

Coupling the network to a branching process

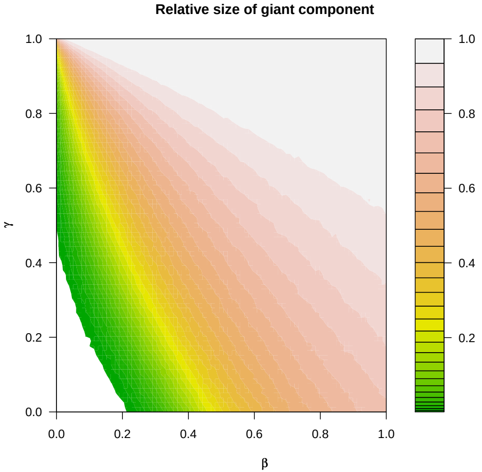
We get a first criterion for the existence and size of a giant component.

Proposition

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. $p(f) > 0$.

Size of the giant component



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

Framework of the proof

- We couple the neighbourhood of a uniformly chosen vertex with the genealogy of a killed branching random walk. This coupling is successful with high probability if up to $\log N$ vertices are explored.
- A sprinkling argument is used to establish the existence of a giant component from the local information. This means that we consider the approximating trees for a network with a slightly reduced edge density . If they survive percolation with positive probability, the remaining edges will connect the surviving trees with high probability, and hence a giant component exists.

Framework of the proof

- We couple the neighbourhood of a uniformly chosen vertex with the genealogy of a killed branching random walk. This coupling is successful with high probability if up to $\log N$ vertices are explored.
- A sprinkling argument is used to establish the existence of a giant component from the local information. This means that we consider the approximating trees for a network with a slightly reduced edge density . If they survive percolation with positive probability, the remaining edges will connect the surviving trees with high probability, and hence a giant component exists.
- The underlying branching random walk has infinite offspring and a uncountable but compact typespace. Martingale arguments are used to show that the killed branching random walk survives percolation if $1/p$ is strictly smaller than the largest eigenvalue of an associated score operator. This operator is unbounded if the network is robust and compact otherwise.

Framework of the proof

- We couple the neighbourhood of a uniformly chosen vertex with the genealogy of a killed branching random walk. This coupling is successful with high probability if up to $\log N$ vertices are explored.
- A sprinkling argument is used to establish the existence of a giant component from the local information. This means that we consider the approximating trees for a network with a slightly reduced edge density . If they survive percolation with positive probability, the remaining edges will connect the surviving trees with high probability, and hence a giant component exists.
- The underlying branching random walk has infinite offspring and a uncountable but compact typespace. Martingale arguments are used to show that the killed branching random walk survives percolation if $1/p$ is strictly smaller than the largest eigenvalue of an associated score operator. This operator is unbounded if the network is robust and compact otherwise.
- In the linear case the typespace degenerates to have just two elements. In this case eigenvalue calculations can be carried out explicitly.