The age-dependent random connection model

joint work with Peter Gracar and Peter Mörters

Arne Grauer, Lukas Lüchtrath

02.04.2019

Universität zu Köln

Introduction

Networks arising in different contexts, be it social, communication, technological or biological, have strikingly similar features.

Introduction

Networks arising in different contexts, be it social, communication, technological or biological, have strikingly similar features.

Such features we will focus on in this talk are:

- Networks are scale-free, i.e. the proportion of vertices of degree k converges against a number $\mu(k)$ and for some power law exponent τ we have

$$\mu(k) = k^{-\tau + o(1)}, \quad \text{as } k \uparrow \infty.$$

Introduction

Networks arising in different contexts, be it social, communication, technological or biological, have strikingly similar features.

Such features we will focus on in this talk are:

- Networks are scale-free.
- Networks show strong clustering, i.e. two vertices, sampled from the neighbourhood of a typical vertex, have an asymptotically positive probability of being connected by an edge.

Introduction

Probabilistic methodology:

- Build a network as a growing sequence of random graphs defined from simple interaction principles,
- Prove the emerging features in form of limit theorems.

Introduction

Probabilistic methodology:

- Build a network as a growing sequence of random graphs defined from simple interaction principles,
- Prove the emerging features in form of limit theorems.

The simple building principles for our network are:

- The network is build dynamically by adding vertices successively,
- when a new vertex is added, it prefers to link to existing vertices that are either
  - powerful,
  - or similar to the new vertex.

Introduction

The idea of preferential attachement to powerful vertices was introduced by *Barabási and Albert (1999)*. They suggested to make the connection probability proportional to the degree of the vertex. We speak about degree-based preferential attachement.

Introduction

The idea of preferential attachement to powerful vertices was introduced by *Barabási and Albert (1999)*. They suggested to make the connection probability proportional to the degree of the vertex. We speak about degree-based preferential attachement.

The idea of preferential attachement to similar vertices is realized by embedding the graphs into space and giving preference to short edges.

Introduction

The idea of preferential attachement to powerful vertices was introduced by *Barabási and Albert (1999)*. They suggested to make the connection probability proportional to the degree of the vertex. We speak about degree-based preferential attachement.

The idea of preferential attachement to similar vertices is realized by embedding the graphs into space and giving preference to short edges.
Various (degree-based) spatial preferential attachment models were studied by: *Manna and Sen (2002), Flaxman, Frieze and Vera (2006), Aiello, Bonato, Cooper, Janssen and Pralat (2009), Jacob, Mörters (2015, 2016), Jordan (2010, 2012), Janssen, Pralat, Wilson (2012), Jordan and Wade (2013), ...*

Introduction

Problem

Those models are complicated since the actual degree of a vertex depends in a complex way on the the rest of the graph. Therefore, we have complicated but (on a large scale) inessential correlations between edges.

Introduction

Problem

Those models are complicated since the actual degree of a vertex depends in a complex way on the the rest of the graph. Therefore, we have complicated but (on a large scale) inessential correlations between edges.

Solution

Preferential attachment to old vertices. The age of a vertex is a given quantity that is highly linked to its degree so we remove the complicated correlations between the edges. We speak about age-based preferential attachment.

The age-based spatial preferential attachment model

We build the graph dynamically in continuous time. At $t = 0$, we start with the empty graph G_0. Then

The age-based spatial preferential attachment model

We build the graph dynamically in continuous time. At $t = 0$, we start with the empty graph G_0. Then

- vertices arrive according to a standard Poisson process in time and are placed independently uniformly on the d-dimensional torus $\mathbb{T}_1^d = (-1/2, 1/2]^d$.

The age-based spatial preferential attachment model

We build the graph dynamically in continuous time. At $t = 0$, we start with the empty graph G_0. Then

- vertices arrive according to a standard Poisson process in time and are placed independently uniformly on the d-dimensional torus $\mathbb{T}_1^d = (-1/2, 1/2]^d$.
- A new vertex (x, t), born a time t and placed at position x, forms an edge to each existing vertex (y, s) independently with probability

$$\varphi\left(\frac{t \cdot d(x,y)^d}{\beta\left(\frac{t}{s}\right)^\gamma}\right).$$

We call this graph G_t.

Rescaling

- (G_t) is a dynamic process of graphs defined from simple interacting principles.

Rescaling

- (G_t) is a dynamic process of graphs defined from simple interacting principles.
- *Missing:* the limit theorems.

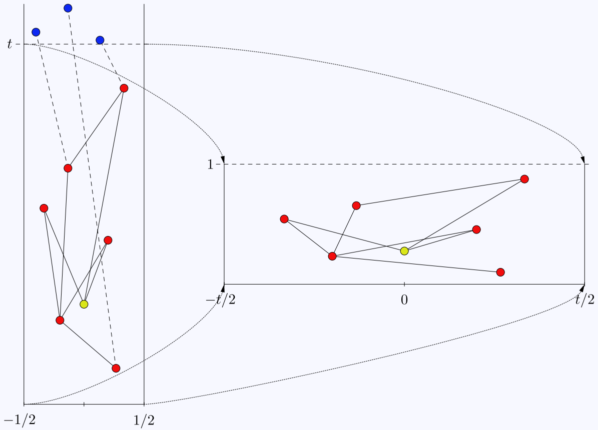
Rescaling

- (G_t) is a dynamic process of graphs defined from simple interacting principles.
- *Missing:* the limit theorems.

For finite $t > 0$, we define the rescaling mapping

$$h_t : \begin{array}{ccc} \mathbb{T}_1^d \times (0, t] & \longrightarrow & \mathbb{T}_t^d \times (0, 1], \\ (x, s) & \longmapsto & \left(t^{1/d} x, s/t\right). \end{array}$$

Rescaling



Rescaling

As seen on the picture, we can construct a new graph process (G^t).

Rescaling

As seen on the picture, we can construct a new graph process (G^t).

The two processes (G_t) and (G^t) have the same one dimensional marginals, i.e. for every fixed t, G_t and G^t have the same law.

Note this is only true for the one-dimensional marginals. In general, the behaviour of the process (G^t) is strikingly different from (G_t).

Rescaling

(G^t) has a natural limit G^∞.

- Place points according to a Poisson point process on $\mathbb{R}^d$ and mark them by independent uniformly $(0, 1]$ distributed 'birth times'.

Rescaling

(G^t) has a natural limit G^∞.

- Place points according to a Poisson point process on $\mathbb{R}^d$ and mark them by independent uniformly $(0,1]$ distributed 'birth times'.
- Given points and birth times, independently connect two points in position x with birth time u, resp. position y with birth time s, with probability

$$\varphi(\beta^{-1}(s \vee u)^{1-\gamma}(s \wedge u)^{\gamma} \cdot |x-y|^d).$$

Rescaling

(G^t) has a natural limit G^∞.

- Place points according to a Poisson point process on $\mathbb{R}^d$ and mark them by independent uniformly $(0,1]$ distributed 'birth times'.
- Given points and birth times, independently connect two points in position x with birth time u, resp. position y with birth time s, with probability

$$\varphi(\beta^{-1}(s \vee u)^{1-\gamma}(s \wedge u)^{\gamma} \cdot |x-y|^d).$$

We call G^∞ The age-dependent random connection model.

Weak local limit

Define G_0^∞ the palm version of G^∞. That is to add a root vertex at the origin with uniformly distributed 'birth time' U and connect it to the graph as before.

Weak local limit

Define G_0^∞ the palm version of G^∞. That is to add a root vertex at the origin with uniformly distributed 'birth time' U and connect it to the graph as before.

Let be H be a nonnegative, uniformly integrable functional, acting on locally finite rooted graphs and only depending on a bounded graph neighbourhood of the root.

We denote the class of those functionals by $\mathcal{H}$.

Weak local limit

Theorem

In probability, the graph sequence (G_t) converges weakly locally to G_0^∞ in the sense that for any such H

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\mathbf{x}\in G_t} H(\theta_{\mathbf{x}} G_t) = \mathbb{E} H(G_0^\infty) \qquad \text{in probability,}$$

where $\theta_{\mathbf{x}}$ acts on points $\mathbf{y} = (y, s)$ as $\theta_{\mathbf{x}}(\mathbf{y}) = (y - x, s)$ and on graphs accordingly.

Neighbourhoods and Degree distributions

We consider edges as oriented from the younger to the older vertex.

- Indegree of a vertex $\mathbf{x}$ is the number of younger vertices that connect to it,
- Outdegree of $\mathbf{x}$ is the number of older vertices that it connected to.

Neighbourhoods and Degree distributions

We consider edges as oriented from the younger to the older vertex.

- Indegree of a vertex **x** is the number of younger vertices that connect to it,
- Outdegree of **x** is the number of older vertices that it connected to.

Aim: Study the asymptotic degree distribution of the age-based spatial preferential model.

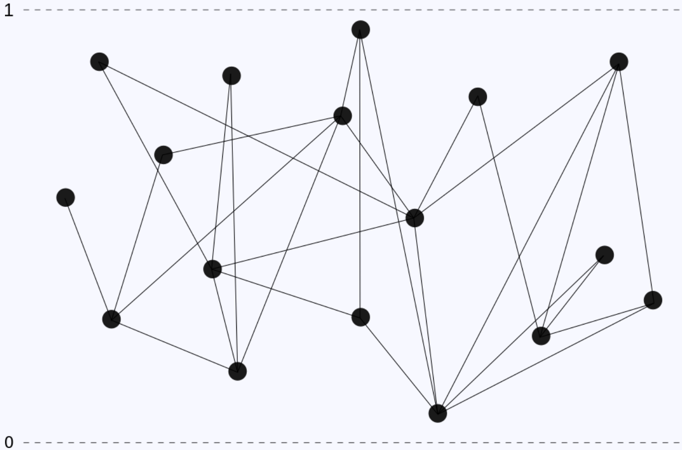
Neighbourhoods and Degree distributions

We consider edges as oriented from the younger to the older vertex.

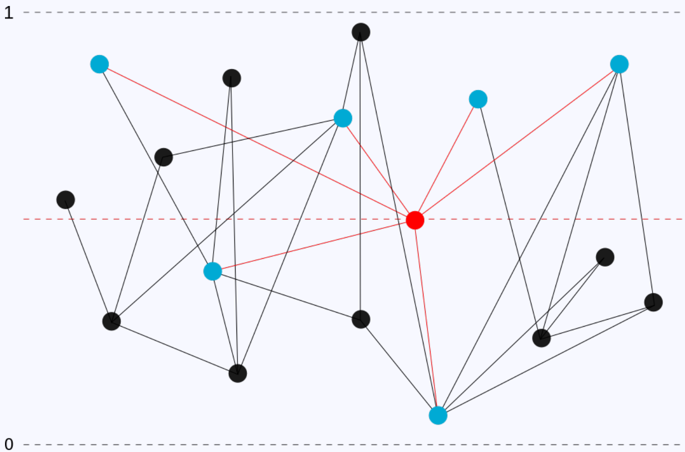
- Indegree of a vertex $\mathbf{x}$ is the number of younger vertices that connect to it,
- Outdegree of $\mathbf{x}$ is the number of older vertices that it connected to.

Aim: Study the asymptotic degree distribution of the age-based spatial preferential model.
To do this, consider the neighbourhood of a fixed vertex in the age-dependent random connection model G^{∞}.

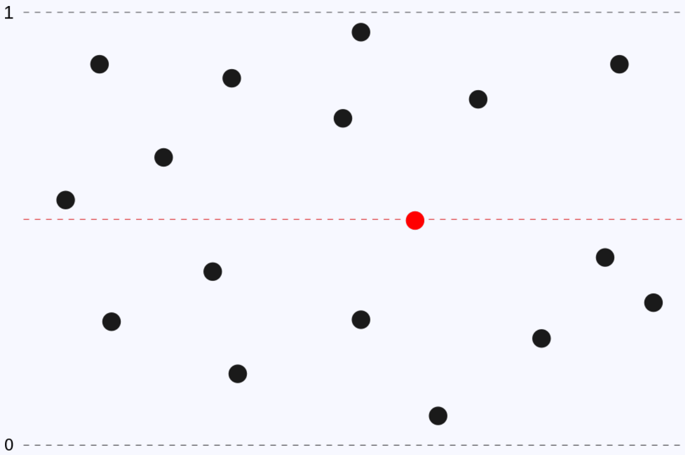
Neighbourhoods and Degree distributions



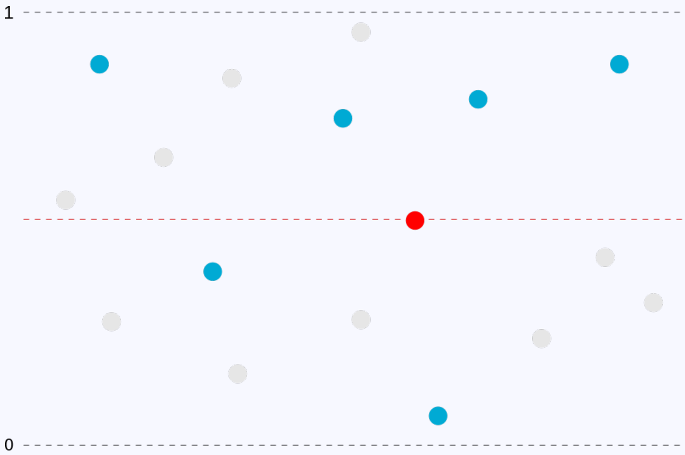
Neighbourhoods and Degree distributions



Neighbourhoods and Degree distributions



Neighbourhoods and Degree distributions



Neighbourhoods and Degree distributions

Proposition

- The older neighbours $\mathcal{Y}_{\mathbf{x}}(G^\infty)$ of a vertex $\mathbf{x} = (x, u)$ in G^∞ form a Poisson point process on $\mathbb{R}^d \times [0, u)$ with intensity measure

$$\lambda_{\mathcal{Y}_{\mathbf{x}}(G^\infty)} := \varphi\left(\beta^{-1} u \left(\frac{s}{u}\right)^\gamma |x - y|^d\right) dy\, ds.$$

Neighbourhoods and Degree distributions

Proposition

- The older neighbours $\mathcal{Y}_{\mathbf{x}}(G^\infty)$ of a vertex $\mathbf{x} = (x, u)$ in G^∞ form a Poisson point process on $\mathbb{R}^d \times [0, u)$ with intensity measure

$$\lambda_{\mathcal{Y}_{\mathbf{x}}(G^\infty)} := \varphi\left(\beta^{-1} u \left(\frac{s}{u}\right)^{\gamma} |x - y|^d\right) \, dy \, ds.$$

- The younger neighbours $\mathcal{Z}_{\mathbf{x}}^\infty(s_0, G^\infty)$ of $\mathbf{x} = (x, u)$ in G^∞ at time $s_0 \in (u, 1]$ form a Poisson point process on $\mathbb{R}^d \times (u, s_0]$ with intensity measure

$$\lambda_{\mathcal{Z}_{\mathbf{x}}^\infty(s_0, G^\infty)} := \varphi\left(\beta^{-1} s \left(\frac{u}{s}\right)^{\gamma} |x - y|^d\right) \, dy \, ds.$$

Neighbourhoods and Degree distributions

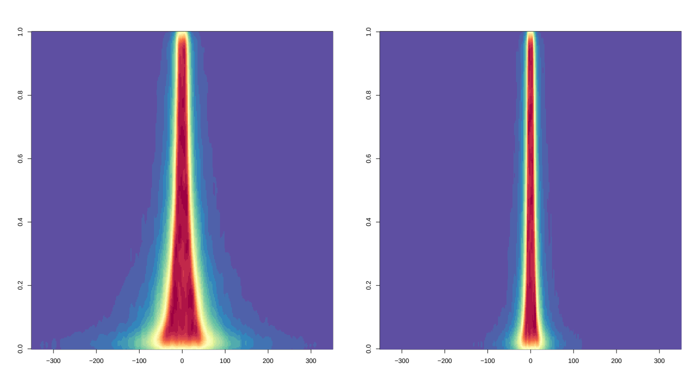


Figure: Heatmaps of the neighbourhood of a relatively old root (left, birth time 0.2) and of a relatively young root (right, birth time 0.8) in G_0^∞ with $\beta = 5$, $\gamma = 1/3$ and $\varphi(x) = 1 \wedge x^{-2}$.

Neighbourhoods and Degree distributions

The empirical outdegree distribution ν_t of the graph G_t is defined by

$$\nu_t := \frac{1}{t} \sum_{\mathbf{x} \in G_t} \mathbb{1}_{\{|\mathcal{Y}_{\mathbf{x}}(G_t)| = k\}} \qquad \text{for } k \in \mathbb{N}.$$

Neighbourhoods and Degree distributions

The empirical outdegree distribution ν_t of the graph G_t is defined by

$$\nu_t := \frac{1}{t} \sum_{\mathbf{x} \in G_t} \mathbb{1}_{\{|\mathcal{Y}_{\mathbf{x}}(G_t)|=k\}} \qquad \text{for } k \in \mathbb{N}.$$

Theorem

For any function $g : \mathbb{N}_0 \to [0, \infty)$ growing no faster than exponentially we have

$$\frac{1}{t} \sum_{\mathbf{x} \in G_t} g(|\mathcal{Y}_{\mathbf{x}}(G_t)|) = \int g \, d\nu_t \longrightarrow \int g \, d\nu,$$

in probability, as $t \to \infty$, where ν is the Poisson distribution with parameter $\beta/(1-\gamma)$.

Neighbourhoods and Degree distributions

The empirical indegree distribution μ_t of the graph G_t is defined by

$$\mu_t(k) = \frac{1}{t} \sum_{\mathbf{x} \in G_t} \mathbb{1}_{\{|\mathcal{Z}_{\mathbf{x}}(t, G_t)| = k\}}.$$

Neighbourhoods and Degree distributions

The empirical indegree distribution μ_t of the graph G_t is defined by

$$\mu_t(k) = \frac{1}{t} \sum_{\mathbf{x} \in G_t} \mathbb{1}_{\{|\mathcal{Z}_{\mathbf{x}}(t, G_t)| = k\}}.$$

Theorem

For any function $g : \mathbb{N}_0 \to [0, \infty)$ growing no faster than linearly we have

$$\frac{1}{t} \sum_{\mathbf{x} \in G_t} g\big(|\mathcal{Z}_{\mathbf{x}}(t, G_t)|\big) = \int g \, d\mu_t \longrightarrow \int g \, d\mu,$$

in probability, as $t \to \infty$, where μ is the mixed Poisson distribution with density $f(\lambda) = \beta^{1/\gamma}(\gamma\lambda + \beta)^{-(1+1/\gamma)}$ for $\lambda > 0$.

Neighbourhoods and Degree distributions

The empirical indegree distribution μ_t of the graph G_t is defined by

$$\mu_t(k) = \frac{1}{t} \sum_{\mathbf{x} \in G_t} \mathbb{1}_{\{|\mathcal{Z}_{\mathbf{x}}(t, G_t)| = k\}}.$$

Theorem

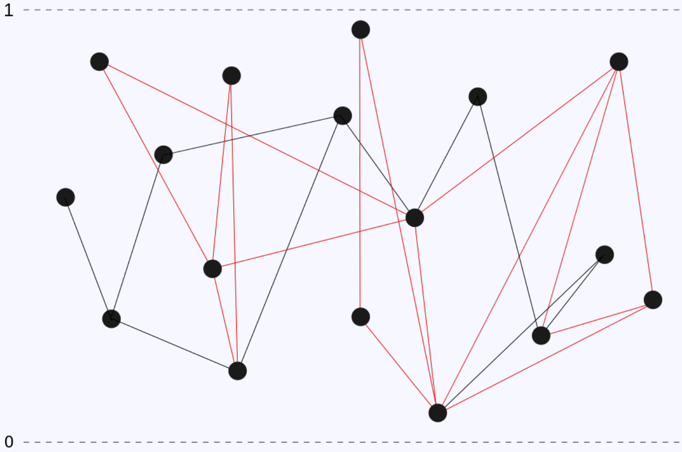
For any function $g : \mathbb{N}_0 \to [0, \infty)$ growing no faster than linearly we have

$$\frac{1}{t} \sum_{\mathbf{x} \in G_t} g\big(|\mathcal{Z}_{\mathbf{x}}(t, G_t)|\big) = \int g \, d\mu_t \longrightarrow \int g \, d\mu,$$

in probability, as $t \to \infty$, where μ is the mixed Poisson distribution with density $f(\lambda) = \beta^{1/\gamma}(\gamma\lambda + \beta)^{-(1+1/\gamma)}$ for $\lambda > 0$.

Notice that $\mu(k) = k^{-(1+\frac{1}{\gamma})+o(1)}$ as $k \uparrow \infty$.

Clustering



Clustering

For a finite graph, we call a pair of edges in G a wedge if they share an endpoint (called its tip). The global clustering coefficient or transitivity of G is defined by

$$c^{\text{glob}}(G) := 3\frac{\text{Number of triangles in } G}{\text{Number of wedges in } G},$$

if there is at least one wedge in G and $c^{\text{glob}}(G) := 0$ otherwise.

Clustering

For a vertex $\mathbf{x}$ with at least two neighbours, the local clustering coefficient is defined by

$$c_{\mathbf{x}}^{\text{loc}}(G) := \frac{\text{Number of triangles in } G \text{ containing vertex } \mathbf{x}}{\text{Number of wedges with tip } \mathbf{x} \text{ in } G}.$$

Clustering

For a vertex $\mathbf{x}$ with at least two neighbours, the local clustering coefficient is defined by

$$c_{\mathbf{x}}^{\text{loc}}(G) := \frac{\text{Number of triangles in } G \text{ containing vertex } \mathbf{x}}{\text{Number of wedges with tip } \mathbf{x} \text{ in } G}.$$

Let $V_2(G)$ be the set of vertices in G with degree at least two, and define the average clustering coefficient by

$$c^{\text{av}}(G) := \frac{1}{|V_2(G)|} \sum_{\mathbf{x} \in V_2(G)} c_{\mathbf{x}}^{\text{loc}}(G),$$

if $V_2(G) \neq \emptyset$ and as $c^{\text{av}}(G) := 0$ otherwise.

Clustering

- The average clustering coefficient places more weight on the low degree nodes, while transitivity places more weight on high degree nodes.

Clustering

- The average clustering coefficient places more weight on the low degree nodes, while transitivity places more weight on high degree nodes.
- Pick two edges sharing a vertex, uniformly from all wedges in the graph. $c^{\text{glob}}(G)$ is the probability that the two endvertices are connected.

Clustering

- The average clustering coefficient places more weight on the low degree nodes, while transitivity places more weight on high degree nodes.
- Pick two edges sharing a vertex, uniformly from all wedges in the graph. $c^{\text{glob}}(G)$ is the probability that the two endvertices are connected.
- Pick a vertex uniformly at random and condition on the event that it has degree greater equal two. Pick two of its neigbours, uniformly at random. $c^{\text{av}}(G)$ is the probability that these neighbours are connected.

Clustering

Theorem

For the global clustering coefficient, there exists a number $c_\infty^{glob} \geq 0$ *such that*

$$c^{glob}(G_t) \longrightarrow c_\infty^{glob}$$

in probability, as $t \to \infty$*. The limiting global clustering coefficient* c_∞^{glob} *is positive if and only if* $\gamma < 1/2$*.*

Clustering

Theorem

For the average clustering coefficient we have

$$c^{av}(G_t) \longrightarrow \int_0^1 \mathbb{P}\{(X_u^{(1)}, S_u^{(1)}) \leftrightarrow (X_u^{(2)}, S_u^{(2)})\}\, \pi(du),$$

in probability as $t \to \infty$, where $(X_u^{(1)}, S_u^{(1)})$ and $(X_u^{(2)}, S_u^{(2)})$ are two independent random variables on $\mathbb{R}^d \times [0,1]$ distributed according to the normalised intensity measure of the neighbourhood of $(0,u)$, and π is the probability measure on $[0,1]$ with density proportional to the probability that $(0,u)$ has at least two neighbours.

Clustering

Proof

For a finite rooted graph G define $H(G) = c_{\mathbf{x}}^{\text{loc}}(G)$ if the root $\mathbf{x}$ has degree at least two, and $H(G) = 0$ otherwise.

Clustering

Proof

For a finite rooted graph G define $H(G) = c_{\mathbf{x}}^{\text{loc}}(G)$ if the root $\mathbf{x}$ has degree at least two, and $H(G) = 0$ otherwise.

- H depends only on the neighbours of $\mathbf{x}$ and their neighbours,
- as H is bounded, $H \in \mathcal{H}$.

Clustering

Proof

For a finite rooted graph G define $H(G) = c_{\mathbf{x}}^{\mathrm{loc}}(G)$ if the root $\mathbf{x}$ has degree at least two, and $H(G) = 0$ otherwise.

- H depends only on the neighbours of $\mathbf{x}$ and their neighbours,
- as H is bounded, $H \in \mathcal{H}$.

We get

$$\frac{1}{t} \sum_{\mathbf{x} \in G_t} H(\theta_{\mathbf{x}} G_t) \longrightarrow \mathbb{E}\left[H(G_0^{\infty})\right]$$

in probability, as $t \to \infty$,

Clustering

Proof

For a finite rooted graph G define $H(G) = c_{\mathbf{x}}^{\text{loc}}(G)$ if the root $\mathbf{x}$ has degree at least two, and $H(G) = 0$ otherwise.

- H depends only on the neighbours of $\mathbf{x}$ and their neighbours,
- as H is bounded, $H \in \mathcal{H}$.

We get

$$\frac{1}{t} \sum_{\mathbf{x} \in G_t} H(\theta_{\mathbf{x}} G_t) \longrightarrow \mathbb{E}\left[H(G_0^\infty)\right]$$

in probability, as $t \to \infty$, with

$$\mathbb{E}\left[H(G_0^\infty)\right] = \int_0^1 \sum_{k \geq 2} \mathbb{E}\left[\frac{2}{k(k-1)} \sum_{(x,s) \leftrightarrow (0,u)} \sum_{\substack{(y,v) \leftrightarrow (0,u) \\ v < s}} \mathbb{1}_{\{(x,s) \leftrightarrow (y,v)\}} \mathbb{1}_{\{(0,u) \text{ has degree } k\}}\right] du.$$

Clustering

Proof.

Conditioned on the degree of $(0, u)$, the neighbours of $(0, u)$ form a Binomial process with sampling distribution given by the normalized intensity measure of the neighbourhood.

$$\mathbb{E}\left[H(G_0^\infty)\right] = \int_0^1 \mathbb{P}\left\{(X_u^{(1)}, S_u^{(1)}) \leftrightarrow (X_u^{(2)}, S_u^{(2)})\right\} \mathbb{P}\left\{(0,u) \text{ has degree greater equal two}\right\} du,$$

where $(X_u^{(1)}, S_u^{(1)})$ and $(X_u^{(2)}, S_u^{(2)})$ are independent and identically distributed. □

Clustering

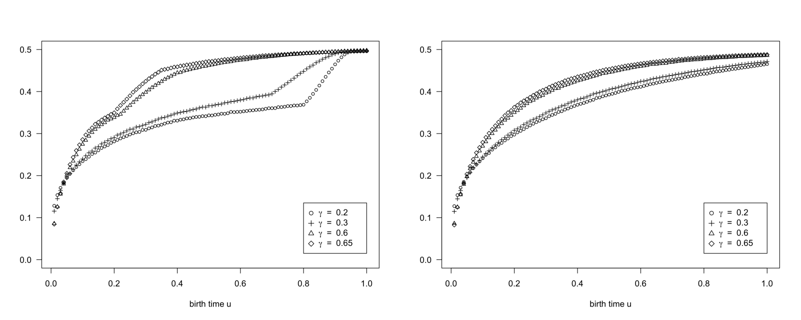


Figure: Local clustering coefficient of a vertex $(0, u)$. The left plot corresponds to the case with low edge density ($c_{\mathrm{ed}} = 0.1$), while the right plot corresponds to high edge density ($c_{\mathrm{ed}} = 10$).

Clustering

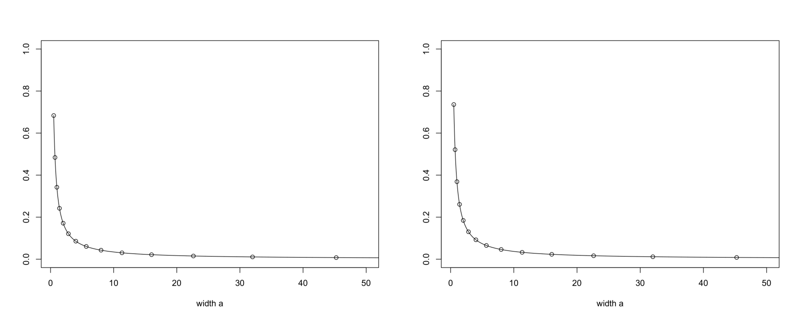


Figure: Average clustering coefficient for the network with profile function $\varphi = \frac{1}{2a}\mathbb{1}_{[0,a]}$ for $\gamma = 0.3$ in the left resp. $\gamma = 0.6$ in the right graphs.

Literatur

- The age-dependent random connection model, *arXiv:1810.03429 , 2018*
- Emergence of scaling in random networks, *Science, 286(5439):509– 512, 1999*
- Spatial preferential attachment networks: power laws and clustering coefficients, *Ann. Appl. Probab., 25(2):632–662, 2015*

THANK YOU!