On preferential and uniform attachment random graphs

Angélica Pachón

University of Torino
Italy

Berlin Mathematical School
21 August - 1 September 2017

Motivation

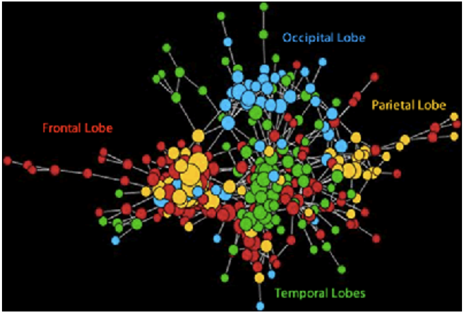

Figure: A human brain networks

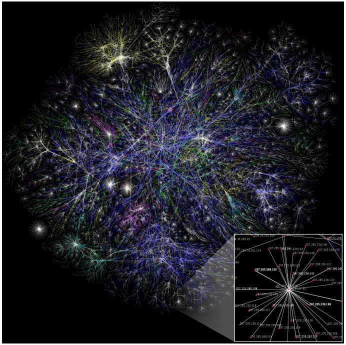

Figure: Internet Traffic Map

Preferential attachment random graphs. Motivation

- Not all things we measure are peaked around a typical value. A classic example of this type of behavior is the sizes of towns and cities

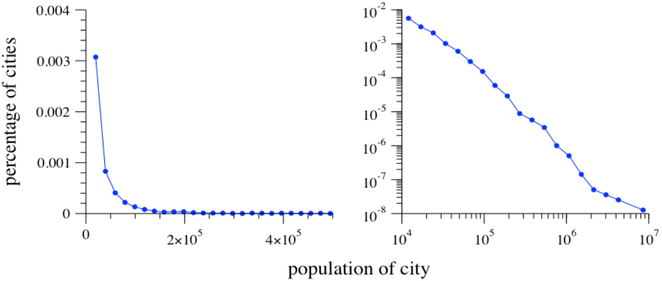

Figure: Left: histogram of the populations of all US cities with population of 10 000 or more. Right: histogram of the same data, but plotted on logarithmic scales.

What does it mean?

Let $p(x)dx$ be the fraction of cities with population between x and $x + dx$. If the histogram is a straight line on log-log scales, then

$$\ln p(x) = -\alpha \ln x + c,$$

where α and c are constants.
Taking the exponential of both sides, this is equivalent to:

$$p(x) = Cx^{-\alpha},$$

with $C = e^c$.

Distributions of this form are said to follow a power law.

Preferential attachment random graphs 1

Barabási–Albert model (BA)

- Start with an initial connected graph of m_0 vertices
- At every step we add a new vertex with $m \leq m_0$ edges
- The probability that a new vertex v_{n+1} will be connected to a vertex $v_j, 1 \leq j \leq n+1$, is proportional to the degree of v_j
- Usually m edges are drawn independently or one by one.
- When $m = 1$ the random graph is a random tree.

Preferential attachment random graphs 2

Formally, let $m \in \mathbb{Z}^+, n \in \mathbb{Z}^+ \cup \{0\}$, and let us define the process $(G_m^t)_{t\geq 1}$, with G_m^1 the graph with a single vertex, without loops. Then,

1. at time $t = n(m+1)+1$ add a new vertex v_{n+1},
2. for $i = 2, \dots, m+1$ at each time $t = n(m+1)+i$ add a direct edge from v_{n+1} to v_j, $j = 1, \dots, n+1$, with

$$\mathbb{P}(v_{n+1} \longrightarrow v_j) = \begin{cases} \frac{d(v_j, t-1)}{T_d(t-1)}, & v_j \neq v_{n+1} \\ \frac{d(v_j, t-1)+1}{T_d(t-1)}, & v_j = v_{n+1}, \end{cases} \tag{1}$$

where $d(v_j, t)$ denotes the degree of v_j at time t and $T_d(t) = \sum_{k=1}^{n+1} d(v_k, t)$.

Preferential attachment

Let $N_{k,t}^{BA}$ be the number of vertices with **degree** equal to k in the BA model, and note that at time $t = n(m+1)$, G_m^t has exactly n vertices.

Theorem (Bollobás, Riordan, Spencer, Tusnády)

Let $m \geq 1$ and $\epsilon > 0$ be fixed, and put $\alpha_{m,k} = \frac{2m(m+1)}{k(k+1)(k+2)}$. Then **whp** *we have*

$$(1-\epsilon)\alpha_{m,k} \leq \frac{N_{m+k,n(m+1)}^{BA}}{n} \leq (1+\epsilon)\alpha_{m,k},$$

for every k in the range $m \leq k \leq n^{1/15}$.

A continuous time birth process: The Yule process

- Let $\{N_\lambda(T) : T \geq 0\}$ be a pure birth process with $N_\lambda(0) = b$, $b \geq 1$, and

$$\mathbb{P}(N_\lambda(T+h) = k+\ell \mid N_\lambda(T) = k) = \begin{cases} k\lambda h + o(h), & \ell = 1, \\ o(h), & \ell > 1, \\ 1 - k\lambda h + o(h), & \ell = 0. \end{cases} \tag{2}$$

The Yule model $\{Y^{a,b}_{\lambda,\beta}(T)\}_{T\geq 0}$

- It makes use of **two related Yule processes**, $\{N_\lambda(T)\}_{T\geq 0}$ and $\{N_\beta(T)\}_{T\geq 0}$, of rates $\lambda > 0$ and $\beta > 0$, respectively, and with initial conditions $N_\lambda(0) = a$ and $N_\beta(0) = b$.
- The **relation** between them is such that when a new individual appears in the Yule process with parameter β, a new Yule process with parameter λ starts.

The method of embedding when $m = 1$

- Let $\{Z_i(T) : T \geq 0\}_{i \geq 1}$ be independent and identically distributed copies of $\{N_1(T) : T \geq 0\}$, a **Yule process** $\{N_\lambda(T)\}_{T\geq 0}$, $\lambda = 1$.
- Let $\{\tau_n\}_{n\geq 0}$ be a convenient sequence of random times, with $\tau_0 = 0$, and so that $\{Z_1(T)\}_{T\geq\tau_0}$, $\{Z_2(T)\}_{T\geq\tau_0}$ and $\{Z_i(T-\tau_{i-2})\}_{T\geq\tau_{i-2}}$, $i \geq 2$.
- Let
$$\tilde{d}(v_i, n) \equiv Z_i(\tau_n - \tau_{i-2}), 1 \leq i \leq n+2$$
and $\tilde{D}_n \equiv \{\tilde{d}(v_i, n), 1 \leq i \leq n+2\}, n \geq 0$.

The method of embedding when $m = 1$

Theorem (Athreya (2007))

Let $\{Z_i(T) : T \geq 0\}_{i \geq 1}$ *and* $\{\tau_n\}_{n \geq 0}$ *be as above. Let*

$$\tilde{d}(v_i, n) \equiv Z_i(\tau_n - \tau_{i-2}), 1 \leq i \leq n+2$$

and $\tilde{D}_n \equiv \{\tilde{d}(v_i, n), 1 \leq i \leq n+2\}, n \geq 0$. *Consider the degree vector sequence for the random graph sequence* $\{G_n\}_{n \geq 0}$, $D_n = \{d(v_i, n), 1 \leq i \leq n+2\}$. *Then the two sequences of random vectors* $\{D_n : n \geq 0\}$ *and* $\{\tilde{D}_n : n \geq 0\}$ *have the same distribution.*

Through this technique of embedding a "discrete" sequence of graphs in a "continuous time" branching process

- K.B. Athreya (2006-2008) studied the empirical degree distribution for the linear, sub-linear and super-linear Preferential attachment model.
- S. Bhamidi (2007) used the results of Aldous (1991) of assymptotic "Fringe distribution" for general families of random trees to study more properties:
  - emprirical degree distribution
  - the size of the subtree rooted at the kth vertex to the born.
  - degree of the root
  - maximum degree
  - the hight of the tree
- S. Dereich, M. Ortgiese (2014): preferential attachment models with fitness

Through this technique of embedding a "discrete" sequence of graphs in a "continuous time" branching process

- K.B. Athreya (2006-2008) studied the empirical degree distribution for the linear, sub-linear and super-linear Preferential attachment model.
- S. Bhamidi (2007) used the results of Aldous (1991) of assymptotic "Fringe distribution" for general families of random trees to study more properties:
  - emprirical degree distribution
  - the size of the subtree rooted at the kth vertex to the born.
  - degree of the root
  - maximum degree
  - the hight of the tree
- S. Dereich, M. Ortgiese (2014): preferential attachment models with fitness

Through this technique of embedding a "discrete" sequence of graphs in a "continuous time" branching process

- K.B. Athreya (2006-2008) studied the empirical degree distribution for the linear, sub-linear and super-linear Preferential attachment model.
- S. Bhamidi (2007) used the results of Aldous (1991) of assymptotic "Fringe distribution" for general families of random trees to study more properties:
  - emprirical degree distribution
  - the size of the subtree rooted at the kth vertex to the born.
  - degree of the root
  - maximum degree
  - the hight of the tree
- S. Dereich, M. Ortgiese (2014): preferential attachment models with fitness

Through this technique of embedding a "discrete" sequence of graphs in a "continuous time" branching process

- K.B. Athreya (2006-2008) studied the empirical degree distribution for the linear, sub-linear and super-linear Preferential attachment model.
- S. Bhamidi (2007) used the results of Aldous (1991) of assymptotic "Fringe distribution" for general families of random trees to study more properties:
  - emprirical degree distribution
  - the size of the subtree rooted at the kth vertex to the born.
  - degree of the root
  - maximum degree
  - the hight of the tree
- S. Dereich, M. Ortgiese (2014): preferential attachment models with fitness

The method of weak convergence

- Consider a Yule model $\{Y^{m,1}_{1/2,1}(T)\}_{T\geq 0}$, that is, the initial conditions for the two Yule processes are $N_\lambda(0) = m$, and $N_\beta(0) = 1$, $\lambda = 1/2$, $\beta = 1$.
- Let $\mathcal{N}^{m,1}_T$ be the size of a individual from $N_\beta(T)$ chosen uniformly at random at time T in $\{Y^{m,1}_{1/2,1}(T)\}_{T\geq 0}$
- Let $deg(v, n)$ denotes the degree of v at time $t = n(m+1)$ (i.e., when there are exactly n vertices) in the **Barabási-Albert model**.

Weak convergence theorem

Theorem (P., Polito, Sacerdote (2016))

For every $i, j \in \mathbb{N}$, $j \geq i$, *and* $w_1 < w_2 < \cdots < w_h \in \mathbb{R}^+$, *there exists a convenient non-decreasing sequence* $z_1, \ldots, z_h \in \mathbb{N}$ *of stopping times, so that*

$$\lim_{i \to \infty} \mathbb{P}[deg(v_j, j + z_1) = k_1, \ldots, deg(v_j, j + z_b) = k_h] \tag{3}$$
$$= \mathbb{P}[N_{1/2}(\ln(1 + w_1)) = k_1, \ldots, N_{1/2}(\ln(1 + w_h)) = k_h],$$

where $N_{1/2}$ *is a Yule process with* $N_{1/2}(0) = m$ *and* $m \leq k_1 \leq \cdots \leq k_b$.

Weak convergence theorem

Let $d^U(V_t)$ indicates the degree of a vertex chosen uniformly at random at time t in the BA model, and $N_{k,t}^{BA}$ be the number of vertices with **degree** equal to k in the BA model.

Theorem (P., Polito, Sacerdote (2016))

Consider a Yule model $\{Y_{1/2,1}^{m,1}(T)\}_{T\geq 0}$. Then for $t = n(m+1)$ we have

$$p_k := \lim_{n\to\infty} \mathbb{P}(d^U(V_t) = k) = \lim_{T\to\infty} \mathbb{P}(\mathcal{N}_T^{m,1} = k). \tag{4}$$

where $\mathcal{N}_T^{m,1}$ is the size of a individual from $N_1(T)$ chosen uniformly at random at time T in $\{Y_{1/2,1}^{m,1}(T)\}_{T\geq 0}$.
Furthermore as $n \to \infty$,

$$\frac{N_{k,t}^{BA}}{n} \to p_k$$

in probability.

Proposition

Consider a Yule model $\{Y^{m,1}_{1/2,1}(T)\}_{T\geq 0}$ *and* $\mathcal{N}^{m,1}_T$ *as above. Then,*

$$p_k = m(m+1)\frac{\Gamma(k)\Gamma(3)}{\Gamma(k+3)}, \tag{5}$$

where $k \geq m$.

Some heuristics I

We interpret the BA model identifying two different processes in discrete time, one for the appearance of directed edges of a fixed vertex and the other for the creation of new vertices.

- For the first one, note that in the BA model, at time when there are n vertices, we have mn directed edges. If $d(v, n)$ denotes the degree of an existing vertex v at time when there are n vertices in the BA model, then by the preferential attachment rule,

$$\mathbb{P}[d(v, n+1) = k+1 \mid d(v, n) = k] \approx \frac{km}{2mn} = \frac{k}{2n}.$$

From this we can see that the distribution of the time interval between the instants at which $d(v, n)$ changes from k to $k+1$ is approximately Geometric with parameter $k/(2n)$.

Some heuristics I

We interpret the BA model identifying two different processes in discrete time, one for the appearance of directed edges of a fixed vertex and the other for the creation of new vertices.

- For the first one, note that in the BA model, at time when there are n vertices, we have mn directed edges. If $d(v,n)$ denotes the degree of an existing vertex v at time when there are n vertices in the BA model, then by the preferential attachment rule,

$$\mathbb{P}[d(v,n+1) = k+1 \mid d(v,n) = k] \approx \frac{km}{2mn} = \frac{k}{2n}.$$

From this we can see that the distribution of the time interval between the instants at which $d(v,n)$ changes from k to $k+1$ is approximately Geometric with parameter $k/(2n)$.

Some heuristics I

We interpret the BA model identifying two different processes in discrete time, one for the appearance of directed edges of a fixed vertex and the other for the creation of new vertices.

- For the first one, note that in the BA model, at time when there are n vertices, we have mn directed edges. If $d(v, n)$ denotes the degree of an existing vertex v at time when there are n vertices in the BA model, then by the preferential attachment rule,

$$\mathbb{P}[d(v, n+1) = k+1 \mid d(v, n) = k] \approx \frac{km}{2mn} = \frac{k}{2n}.$$

From this we can see that the distribution of the time interval between the instants at which $d(v, n)$ changes from k to $k+1$ is approximately Geometric with parameter $k/(2n)$.

Some heuristics II

- For the second one, we see the deterministic process of appearance of vertices in the BA model in a different manner. To do that we wait up to see n vertices in the BA model, $n > i$, $i > 1$ and construct a set of i dependent but identically distributed birth processes (in discrete time). We call this the planted model. Then we consider the following experiment.
  1. Choose one of the i birth processes with probability proportional to its number of vertices it has.
  2. Choose a vertex uniformly at random, among the vertices belonging to the selected birth process.
- We prove that the planted model together with the previous experiment induce the uniform distribution for selecting a vertex on the set of $[n]$ vertices in the BA model. Furthermore, they induce the uniform distribution for choosing a birth process on the set of i birth processes of the planted model.

Some heuristics II

- For the second one, we see the deterministic process of appearance of vertices in the BA model in a different manner. To do that we wait up to see n vertices in the BA model, $n > i$, $i > 1$ and construct a set of i dependent but identically distributed birth processes (in discrete time). We call this the planted model. Then we consider the following experiment.
  1. Choose one of the i birth processes with probability proportional to its number of vertices it has.
  2. Choose a vertex uniformly at random, among the vertices belonging to the selected birth process.
- We prove that the planted model together with the previous experiment induce the uniform distribution for selecting a vertex on the set of $[n]$ vertices in the BA model. Furthermore, they induce the uniform distribution for choosing a birth process on the set of i birth processes of the planted model.

Some heuristics II

- For the second one, we see the deterministic process of appearance of vertices in the BA model in a different manner. To do that we wait up to see n vertices in the BA model, $n > i$, $i > 1$ and construct a set of i dependent but identically distributed birth processes (in discrete time). We call this the planted model. Then we consider the following experiment.
  1. Choose one of the i birth processes with probability proportional to its number of vertices it has.
  2. Choose a vertex uniformly at random, among the vertices belonging to the selected birth process.
- We prove that the planted model together with the previous experiment induce the uniform distribution for selecting a vertex on the set of $[n]$ vertices in the BA model. Furthermore, they induce the uniform distribution for choosing a birth process on the set of i birth processes of the planted model.

The Uniform-Preferential attachment model (UPA)

We propose a **generalization of the Barabási-Albert model** which takes into account two different attachment rules for new nodes of the network. We investigate the degree distribution.

Motivation

- Consider a website where registered members can submit content, such as text posts.
- Furthermore, registered users can vote previous posts to determine their position on the site's pages.
- Hence, the submissions with the most positive votes appear on the front page, together with a fixed number of the most recent posts (see **www.reddit.com**).

This is a network in which nodes are the posts and edges are votes.

The attachment rules:

- Imagine that when a user submits a new post, it also votes on some other previous submissions.
- It is reasonable to assume that the user tends to select and vote either on **the most recent posts** or **the most popular posts**.
- Hence, the user votes the posts according to two different rules:
  - with uniform probability if the user decides to select a post recently published, and
  - with probability proportional to the number of votes, otherwise.

The attachment rules:

- Imagine that when a user submits a new post, it also votes on some other previous submissions.
- It is reasonable to assume that the user tends to select and vote either on **the most recent posts** or **the most popular posts**.
- Hence, the user votes the posts according to two different rules:
  - with uniform probability if the user decides to select a post recently published, and
  - with probability proportional to the number of votes, otherwise.

The UPA model

Formally, suppose that every new node v_{t+1} selects a neighbor either within a limited window of nodes (the $l \in \mathbb{N}$ youngest nodes of the network), or among all nodes $(v_0, \ldots, v_t)$ as follows: Let $0 \leq p \leq 1$ and $l \geq 1$. Then,

(a) At the starting period $t = l$, $l \in \mathbb{N}$, the initial graph G^l has $l + 1$ nodes $(v_0, v_1, \ldots v_l)$, where every node v_j, $1 \leq j \leq l$, is connected to v_0.

(b) Given G^t, at time $t + 1$ add a new node v_{t+1} together with an outgoing edge. Such edge links v_{t+1} with an existing node chosen either within a window, or among all nodes present in the network at time t, as follows:
   - with probability p, v_{t+1} chooses its neighbour in the set $\{v_{t-l+1}, \ldots, v_t\}$, and each node within this window has probability $\frac{1}{l}$ of being chosen.
   - with probability $1 - p$, the neighbour of v_{t+1} is chosen from the set $\{v_0, \ldots, v_t\}$, and each node v_j, $j = 0, \ldots, t$, has probability $\frac{\deg(v_j)}{2t}$ of being chosen.

The UPA model

Formally, suppose that every new node v_{t+1} selects a neighbor either within a limited window of nodes (the $l \in \mathbb{N}$ youngest nodes of the network), or among all nodes $(v_0, \ldots, v_t)$ as follows: Let $0 \leq p \leq 1$ and $l \geq 1$. Then,

(a) At the starting period $t = l$, $l \in \mathbb{N}$, the initial graph G^l has $l + 1$ nodes (v_0, v_1, ... v_l), where every node v_j, $1 \leq j \leq l$, is connected to v_0.

(b) Given G^t, at time $t + 1$ add a new node v_{t+1} together with an outgoing edge. Such edge links v_{t+1} with an existing node chosen either within a window, or among all nodes present in the network at time t, as follows:

- with probability p, v_{t+1} chooses its neighbour in the set $\{v_{t-l+1}, ..., v_t\}$, and each node within this window has probability $\frac{1}{l}$ of being chosen.
- with probability $1 - p$, the neighbour of v_{t+1} is chosen from the set $\{v_0, ..., v_t\}$, and each node v_j, $j = 0, \ldots, t$, has probability $\frac{\deg(v_j)}{2t}$ of being chosen.

The UPA model

Formally, suppose that every new node v_{t+1} selects a neighbor either within a limited window of nodes (the $l \in \mathbb{N}$ youngest nodes of the network), or among all nodes $(v_0, \ldots, v_t)$ as follows: Let $0 \leq p \leq 1$ and $l \geq 1$. Then,

(a) At the starting period $t = l$, $l \in \mathbb{N}$, the initial graph G^l has $l+1$ nodes (v_0, v_1, ... v_l), where every node v_j, $1 \leq j \leq l$, is connected to v_0.

(b) Given G^t, at time $t+1$ add a new node v_{t+1} together with an outgoing edge. Such edge links v_{t+1} with an existing node chosen either within a window, or among all nodes present in the network at time t, as follows:

- with probability p, v_{t+1} chooses its neighbour in the set $\{v_{t-l+1}, \ldots, v_t\}$, and each node within this window has probability $\frac{1}{l}$ of being chosen.
- with probability $1-p$, the neighbour of v_{t+1} is chosen from the set $\{v_0, \ldots, v_t\}$, and each node v_j, $j = 0, \ldots, t$, has probability $\frac{\deg(v_j)}{2t}$ of being chosen.

Asymptotic degree distribution in the UPA model

Theorem

Let $N(k,t)$ denotes the number of nodes in the network with degree k at time t. Then,

$$\frac{N(k,t)}{t} \to P(k) \tag{6}$$

in probability as $t \to \infty$, where for $l = 1$ it holds:

$$P(k) = \begin{cases} \frac{2(1-p)}{3-p} & \text{if } k = 1 \\ \frac{(1-p)^2}{(2-p)(3-p)} + \frac{p}{2-p} & \text{if } k = 2 \\ \left(\frac{2}{1-p}+2\right)\left(\frac{2}{1-p}+1\right) B\left(k, 1+\frac{2}{1-p}\right) P(2) & \text{if } k > 2, \end{cases} \tag{7}$$

where $B(x,y)$ is the Beta function,

Asymptotic degree distribution in the UPA model

Theorem

Let $N(k,t)$ denotes the number of nodes in the network with degree k at time t. Then,

$$\frac{N(k,t)}{t} \to P(k) \tag{6}$$

in probability as $t \to \infty$, where for $l = 1$ it holds:

$$P(k) = \begin{cases} \frac{2(1-p)}{3-p} & \text{if } k = 1 \\ \frac{(1-p)^2}{(2-p)(3-p)} + \frac{p}{2-p} & \text{if } k = 2 \\ \left(\frac{2}{1-p} + 2\right)\left(\frac{2}{1-p} + 1\right) B\left(k, 1 + \frac{2}{1-p}\right) P(2) & \text{if } k > 2, \end{cases} \tag{7}$$

where $B(x,y)$ is the Beta function,

while for $l > 1$ we have:

$$P(k) = \begin{cases} \frac{2}{(3-p)} \left(1 - \frac{p}{l}\right)^l & \text{ik } k = 1 \\ \frac{2}{2+k(1-p)} \left(\frac{p}{l}(H_{k-1} - H_k) + \frac{(1-p)(k-1)}{2} P(k-1)\right) & \text{if } k = 2, \ldots, l+1 \\ \frac{B\left(k, l+2+\frac{2}{1-p}\right)}{B\left(l+1, k+1+\frac{2}{1-p}\right)} P(l+1) & \text{if } k > l+1, \end{cases}$$

where

$$H_k = \begin{cases} \left(\frac{p}{l}\right)^{k-1} \sum_{m=1}^{l-(k-1)} \binom{l-m}{l-m-(k-1)} \left(1 - \frac{p}{l}\right)^{l-m-(k-1)} & \text{if } k = 1, \ldots, l. \\ 0 & \text{if } k > l. \end{cases} \tag{9}$$

while for $l > 1$ we have:

$$P(k) = \begin{cases} \frac{2}{(3-p)}\left(1-\frac{p}{l}\right)^{l} & \text{ik } k = 1 \\ \frac{2}{2+k(1-p)}\left(\frac{p}{l}(H_{k-1}-H_k) + \frac{(1-p)(k-1)}{2}P(k-1)\right) & \text{if } k = 2,\ldots,l+1 \\ \frac{B\left(k,l+2+\frac{2}{1-p}\right)}{B\left(l+1,k+1+\frac{2}{1-p}\right)}P(l+1) & \text{if } k > l+1, \end{cases}$$

where

$$H_k = \begin{cases} \left(\frac{p}{l}\right)^{k-1}\sum_{m=1}^{l-(k-1)}\binom{l-m}{l-m-(k-1)}\left(1-\frac{p}{l}\right)^{l-m-(k-1)} & \text{if } k = 1,\ldots,l. \\ 0 & \text{if } k > l. \end{cases} \tag{9}$$

Corollary

As $k \to \infty$, *for* $l = 1$

$$\frac{N(k,t)}{t} \sim C_p \left[k^{-\left(1+\frac{2}{1-p}\right)} - \frac{3-p}{(1-p)^2} k^{-\left(2+\frac{2}{1-p}\right)} \right], \tag{10}$$

where $C_p = \Gamma(1 + {}^2/_{(1-p)})({}^2/_{(1-p)} + 2)({}^2/_{(1-p)} + 1)P(2)$, *and for* $l > 1$,

$$\frac{N(k,t)}{t} \sim C_{p,l} \left[k^{-\left(1+\frac{2}{1-p}\right)} - \frac{3-p}{(1-p)^2} k^{-\left(2+\frac{2}{1-p}\right)} \right], \tag{11}$$

where $C_{p,l} = \Gamma(l + 2 + {}^2/_{(1-p)})\big(\Gamma(l+1)\big)^{-1} P(l+1)$.

Ideas of the proof

To prove the previos theorem we pursue the following steps:

(1) we determine recursivelly $\mathbb{E}[N(k,t)], k = 1, 2, \ldots;$

(2) we prove the existence of $P(k) := \lim_{t\to\infty} \mathbb{E}[N(t,k)]/t$,

(3) we determine an explicit expression for $P(k)$,

(4) we use the Azuma-Hoeffding Inequality to prove convergence in probability of $N(k,t)/t$ to $P(k)$.

Future projects

- On the relation between generalizations of Yule's model and random graphs.

  - To explore variants of random graphs with birth and death mechanism in the creation of new links or vertices; migration effects; attachment rules proportional to given functions, Non-Markov hypotesis and long-range dependence.