4.3 Preferential attachment model

Both inhomogeneous random graphs and configuration model are *static* in the sense that, given the sequences $\boldsymbol{w}$ or $\boldsymbol{d}$, we can produce a realization of a random graph with the required degree distribution. They do not explain why this degree distribution appear, what are the internal laws that force the degree distribution to take the required form. In this section I will study a model that is actually given in the form os a random graph process, which results in the power law degree distribution. I start with a heuristic derivation of this law in terms of number of species per genera, that was originally analyzed by Yule[2].

4.3.1 A heuristic derivation of Yule's distribution

A genus is a taxonomic unit in biology that consists of several species. If we have k species in some genus, then at the speciation even one of them can become 2, hence $k+1$ species within the same genera, or, a rarer event, a completely new species in a completely new genus appear (say, once per every m speciation events). Let the unit of time in this model be exactly the event of appearance of a new genus, and let $n_{k,t}$ denote the number of genera with k species when there at the time moment t, that is when there are t genera, $t=\sum_{k=1}^{\infty} n_{k,t}$. Every time unite there are 30 speciation events that lead to an increase of some genus by 1 and one speciation event that leads to a new genus. It is reasonable to assume that the probability that a speciation even happens in a genus with k species is proportional to k:

$$\frac{km}{\sum_{k=1}^{\infty} kn_{k,t}} = \frac{km}{(m+1)t}\,,$$

where the factor m comes from the fact that m such events happens independently. Now we can produce simple bookkeeping

$$n_{k,t+1} = n_{k,t} - \frac{km}{m+1}\frac{n_{k,t}}{t} + \frac{(k-1)m}{m+1}\frac{n_{k-1,t}}{t} = n_{k,t} + \frac{m}{m+1}\left((k-1)\frac{n_{k-1,t}}{t} - k\frac{n_{k,t}}{t}\right),$$

which simply counts how the number of genera with k species can change during one time unit. The last equality holds for any k except for $k=1$, where we have

$$n_{1,t+1} = n_{1,t} + 1 - \frac{m}{m+1}\frac{n_{1,t}}{t}\,.$$

Now I assume (this requires proof, of course) that there are limits

$$\lim_{t\to\infty}\frac{n_{k,t}}{t} = p_k,$$

independent of t. I also have

$$\lim_{t\to\infty}(n_{k,t+1}-n_{k,t}) = \lim_{t\to\infty}\left(t\frac{n_{k,t+1}}{t+1} - t\frac{n_{k,t}}{t} + \frac{n_{k,t+1}}{t+1}\right) = p_k,$$

which implies that for the sequence $(p_k)_{k\geq 1}$, I find

$$p_1 = 1 - \frac{m}{m+1}p_1 \implies p_1 = \frac{m+1}{2m+1}\,,$$

and

$$p_k = \frac{m}{m+1}\big((k-1)p_{k-1} - kp_k\big) \implies p_k = \frac{k-1}{1+k+1/m}p_{k-1}\,.$$

[2]Yule, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character, 21-87.

This yields that

$$
\begin{aligned}
p_k &= \frac{(k-1)(k-2)\dots 1}{(k+1+1/m)(k+1/m)\dots(3+1/m)} p_1 \\
&= \left(1+\frac{1}{m}\right) \frac{(k-1)\dots 1}{(k+1+1/m)\dots(2+1/m)} \\
&= \left(1+\frac{1}{m}\right) \frac{\Gamma(k)\Gamma(2+1/m)}{\Gamma(2+k+1/m)},
\end{aligned}
$$

where

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-1} \mathrm{d}t$$

is the gamma-function, which generalized the notion of the factorial for non-integer values.

Problem 4.14. Show that

$$\Gamma(x) = (x-1)\Gamma(x-1), \quad \Gamma(1) = 1.$$

Therefore

$$\Gamma(k+1) = k!$$

for integer k.

Euler's beta function is defined

$$\mathrm{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)},$$

and hence I finally get my solution

$$p_k = \left(1+\frac{1}{m}\right) \mathrm{B}(k, 2+1/m).$$

For large values of x Stirling's approximation

$$\Gamma(x) \sim \sqrt{2\pi} e^{-x} x^{x-1/2}$$

can be used to show that

$$\mathrm{B}(x, y) \sim x^{-y}\Gamma(y),$$

that is it has a power law tail with exponent y.

This finishes the proof that our model of speciation produces a power law distribution with exponent

$$\alpha = 2 + \frac{1}{m}.$$

Note that the main component that allowed to get a power law distribution was the principle that in a genus with many species it is much probable to have a speciation event. This principle of "rich get richer" was also used by Barabási and Albert[3], which (again in a heuristic manner) I will present next.

[3] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.

4.3.2 Preferential attachment model for the WWW

Here I present a heuristic derivation of the power law distribution via the so-called principle of *preferential attachment* (in this and the following subsections I follow mainly this paper[4]). Consider the World Wide Web, with can be represented as a directed graph, where the vertices correspond to the web pages and there is an edge from vertex i to vertex j if there is a hyperlink from page i to page j. Now each page can be characterized by the number of hyperlinks to this page, in the graph theoretic language this number is called the *in-degree*. Therefore, we can talk about the in-degree distribution and it turns out that this distribution can be closely approximated by a power law. Assume that a new web page is created. It is reasonable to expect that this new page will link to some popular web pages, i.e., the chance that a new web page is connected to a web-page with in-degree k should be proportional to k, and this is what is usually called the preferential attachment principle.

Here is an informal argument to formalize the preferential attachment. Let $x_j(t)$ be the number of web pages with in-degree j when there are t pages total. Then, for $j \geq 1$ the probability that $x_j(t)$ increases is simply

$$\alpha \frac{x_{j-1}(t)}{t} + (1-\alpha)\frac{(j-1)x_{j-1}(t)}{t}\,,$$

if we assume that new web page appears with only one link to existing pages, and this one link is chosen randomly among all t pages with probability α and with probability $1-\alpha$ this one link is chosen randomly but with probabilities proportional to the existing in-degrees. Similarly, the probability that $x_j(t)$ decreases is

$$\alpha \frac{x_j(t)}{t} + (1-\alpha)\frac{jx_j(t)}{t}\,.$$

Therefore, for $j \geq 1$,

$$\dot{x}_j = \frac{\alpha(x_{j-1}-x_j) + (1-\alpha)\big((j-1)x_{j-1} - jx_j\big)}{t}\,.$$

The case $j=0$ should be treated differently since each new web page has in-degree 0, and therefore

$$\dot{x}_0 = 1 - \frac{\alpha x_0}{t}\,.$$

We obtained a non-autonomous system of linear ordinary differential equations. Since time unit in the model is appearance of one new web page, we can assume that the limiting stationary state should have the form

$$x_j(t) = c_j t,$$

where c_j is a constant, which specifies which fraction of the total number of pages the pages with in-degree j constitute.

We have for x_0

$$\dot{x}_0 = c_0 = 1 - \alpha c_0 \implies c_0 = \frac{1}{1+\alpha}\,.$$

For general j

$$c_j(1+\alpha+j(1-\alpha)) = c_{j-1}(\alpha + (j-1)(1-\alpha)).$$

We can determine c_j exactly using the above recurrence, but for our goal it is enough to note that

$$\frac{c_j}{c_{j-1}} = 1 - \frac{2-\alpha}{1+\alpha+j(1-\alpha)} \sim 1 - \left(\frac{2-\alpha}{1-\alpha}\right)\frac{1}{j}\,.$$

This yields that asymptotically

$$c_j \sim Cj^{-\frac{2-\alpha}{1-\alpha}}\,,$$

[4]Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. Internet mathematics, 1(2), 226-251.

for some constant C. To see this, note that the last expression means

$$\frac{c_j}{c_{j-1}} \sim \left(\frac{j-1}{j}\right)^{\frac{2-\alpha}{1-\alpha}} \sim 1 - \left(\frac{2-\alpha}{1-\alpha}\right)\frac{1}{j}\,,$$

as required.

4.3.3 Conditional expectations

To rigorously analyze preferential attachment model I will need the notions of conditional expectation.

In this section I discuss the necessary theory. Actually, we will need only one small instance of this notion, but it is very advisable, going into research literature on the random graphs, to be aware of the conditional expectation and its properties.

Recall that with a random experiment associated a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, here Ω is a finite set of outcomes, $\mathcal{F}$ is an *algebra* of events, and P is a probability measure. We discussed Ω and P, but almost nothing was said about $\mathcal{F}$. The notion of *algebra* means that if $A, B \in \mathcal{F}$ then $A \cap B, A \cup B, A \setminus B, B \setminus A$ are also in $\mathcal{F}$, and also $\Omega \in \mathcal{F}$. Here are some examples of algebras on Ω:

1. $\{\emptyset, \Omega\}$;
2. $\{\emptyset, A, \overline{A}, \Omega\}$;
3. The set of all subsets of Ω, this is sometimes called the power set and denoted 2^Ω.

Problem 4.15. Can you construct an algebra that includes two events A and B such that $B \neq \overline{A}$?

Consider a *decomposition* $\mathscr{D}$ of Ω, i.e., a set of events $\mathscr{D} = \{D_1, \ldots, D_n\}$ such that D_j are pairwise disjoint, all together they sum to $\Omega = D_1 \cup \ldots \cup D_n$, and $\mathsf{P}(D_j) > 0$. Events D_j are sometimes called *atoms* of $\mathscr{D}$ for all j. If we consider the unions of the sets in $\mathscr{D}$, the resulting collection of sets, together with the empty set, forms an *algebra* (why?), which is called the algebra induced by $\mathscr{D}$ and denoted $\alpha(\mathscr{D})$ (for example, the algebra from example 2 above induced by decomposition $\{A, \overline{A}\}$; which decomposition induce algebra in point 3?). Hence, if we have a decomposition $\mathscr{D}$, then we have an algebra associated with it. The converse is also true. Let $\mathscr{B}$ be an algebra of subsets of finite sample space Ω. Then there exists a unique decomposition $\mathscr{D}$, whose atoms are the elements of $\mathscr{B}$ such that $\mathscr{B} = \alpha(\mathscr{D})$ (prove it. A good idea is to start thinking about the the problem above and the decomposition that can generate an algebra containing two sets A and B).

Let $\mathscr{D}_1$ and $\mathscr{D}_2$ be two decompositions. We say that $\mathscr{D}_2$ is finer than $\mathscr{D}_1$ and denote $D_1 \preccurlyeq D_2$ if $\alpha(\mathscr{D}_1) \subseteq \alpha(\mathscr{D}_2)$.

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, and let $\mathscr{D}$ be a decomposition such that $D_i \in \mathcal{F}$ for each i. Let $A \in \mathcal{F}$ be an event. Then we can consider conditional probabilities $\mathsf{P}(A \mid D_i)$, $i = 1, \ldots, n$, of the event A with respect to events D_i. Now consider a random variable X, which takes the values $\mathsf{P}(A \mid D_i)$ with probabilities $\mathsf{P}(D_i)$. Formally,

$$X(\omega) = \sum_{i=1}^{n} \mathsf{P}(A \mid D_i)\, \mathbf{1}_{D_i}(\omega).$$

This random variable is denoted as $\mathsf{P}(A \mid \mathscr{D})$ to emphasize that it is associated with the decomposition $\mathscr{D}$ and called *conditional probability with respect to decomposition $\mathscr{D}$.*

Problem 4.16. Show that if $A \cap B = \emptyset$, then

$$\mathsf{P}(A \cup B \mid \mathscr{D}) = \mathsf{P}(A \mid \mathscr{D}) + \mathsf{P}(B \mid \mathscr{D}).$$

Also show that if $\mathscr{D} = \{\Omega\}$ (the trivial decomposition) then

$$\mathsf{P}(A \mid \{\Omega\}) = \mathsf{P}(A).$$

Show that using the notion of the conditional probability with respect to a decomposition, the formula of total probability can be written as

$$\mathsf{E}(\mathsf{P}(A \mid \mathscr{D})) = \mathsf{P}(A).$$

Now assume that we start with a random variable Y that takes values $y_1, \ldots, y_n$:

$$Y(\omega) = \sum_{i=1}^{n} y_i \, \mathbf{1}_{D_i}(\omega),$$

where $D_i = \{\omega \colon Y(\omega) = y_i\}$ is the decomposition induced by the random variable Y, which I will denote as $\mathscr{D}_Y$ (you should actually prove that for any surjective function $f \colon V \to W$ the operation f^{-1} defined as $f^{-1}(w) = \{v \in V \colon f(v) = w\}$ induces a decomposition of the set V). Hence, we can actually consider the conditional probability with respect to random variable Y:

$$\mathsf{P}(A \mid Y) := \mathsf{P}(A \mid \mathscr{D}_Y),$$

where $\mathscr{D}$ is the decomposition induced by Y.

By analogy we can consider

$$\mathsf{P}(A \mid Y_1, \ldots, Y_m),$$

as the conditional probability with respect to decomposition induced by the random variables $Y_1, \ldots, Y_m$.

Problem 4.17. Show that if X and Y are two independent identically distributed random variables, each taking values 0 and 1 with probabilities q and p, then

$$\mathsf{P}(X + Y = k \mid Y) = \begin{cases} q(1 - Y), & k = 0, \\ p(1 - Y) + qY, & k = 1, \\ pY, & k = 2. \end{cases}$$

Recall that if we are given a random variable

$$X(\omega) = \sum_{i=1}^{k} x_i \, \mathbf{1}_{A_i}(\omega),$$

then its expectation is

$$\mathsf{E}X = \sum_{i=1}^{k} x_i \, \mathsf{P}(A_i).$$

Here $A_i = \{\omega \in \Omega \colon X(\omega) = x_i\}$. If, additionally, a decomposition $\mathscr{D} = \{D_1, \ldots, D_n\}$ is given, then it is natural to define the *conditional expectation* with respect to decomposition $\mathscr{D}$:

$$\begin{aligned} \mathsf{E}(X \mid \mathscr{D}) := \sum_{i=1}^{k} x_i \, \mathsf{P}(A_i \mid \mathscr{D}) &= \\ = \sum_{i=1}^{k} x_i \sum_{j=1}^{n} \mathsf{P}(A_i \mid D_j) \, \mathbf{1}_{D_j} &= \\ = \sum_{j=1}^{n} \mathbf{1}_{D_j} \sum_{i=1}^{k} x_i \, \mathsf{P}(A_i \mid D_j). \end{aligned}$$

According to this definition the conditional expectation $\mathsf{E}(X \mid \mathscr{D})$ is a random variable, which at the sample points belonging to the set D_j takes the values $\sum_{i=1}^k x_i \, \mathsf{P}(A_i \mid D_j)$. This shows that we could have defined first the conditional expectations with respect to D_j:

$$\mathsf{E}(X \mid D_j) = \sum_{i=1}^{k} x_i \, \mathsf{P}(A_i \mid D_j) = \frac{\mathsf{E}(X \, \mathbf{1}_{D_j})}{\mathsf{P}(D_j)} \,,$$

and then define

$$\mathsf{E}(X \mid \mathscr{D})(\omega) := \sum_{j=1}^{n} \mathsf{E}(X \mid D_j) \, \mathbf{1}_{D_j}(\omega).$$

Problem 4.18. Show that

1. $\mathsf{E}(aX + bY \mid \mathscr{D}) = a \, \mathsf{E}(X \mid \mathscr{D}) + b \, \mathsf{E}(Y \mid \mathscr{D}), \quad a, b \in \mathbf{R}.$
2. $\mathsf{E}(X \mid \Omega) = \mathsf{E}(X).$
3. $\mathsf{E}(C \mid \mathscr{D}) = C, \quad C \in \mathbf{R}.$
4. If $X = \mathbf{1}_A$ then $\mathsf{E}(X \mid \mathscr{D}) = \mathsf{P}(A \mid \mathscr{D}).$
5. $\mathsf{E}(\mathsf{E}(X \mid \mathscr{D})) = \mathsf{E}X.$

Similarly to the conditional probability with respect to decomposition $\mathscr{D}$, if $\mathscr{D}$ is induced by the random variables $Y_1, \ldots, Y_m$, then we have the conditional expectation

$$\mathsf{E}(X \mid Y_1, \ldots, Y_m) := \mathsf{E}(X \mid \mathscr{D}_{Y_1, \ldots, Y_m})$$

with respect to the random variables $Y_1, \ldots, Y_m$.

Sometimes the conditional expectation with respect to a random variable is defined in the following way. Consider two random variables X and Y, and their *probability mass function*

$$p_{XY}(x_i, y_j) = p_{ij} := \mathsf{P}(X = x_i, Y = y_j), \quad i = 1, \ldots, k \,, j = 1 \ldots, n.$$

Then the *marginal distributions* are

$$p_X(x_i) := \mathsf{P}(X = x_i) = \sum_{j=1}^{n} p_{ij}, \quad p_Y(y_j) := \mathsf{P}(Y = y_j) = \sum_{i=1}^{k} p_{ij}.$$

The conditional probability mass function of X given $Y = y_j$ is defined as

$$p_{X|Y}(x_i \mid y_j) = \mathsf{P}(X = x_i \mid Y = y_j) = \frac{p_{ij}}{p_Y(y_j)} \,.$$

Therefore the conditional expectation $\mathsf{E}(X \mid Y = y_j) = \sum_{i=1}^k x_i p_{X|Y}(x_i \mid y_j)$. Note that we can write the last formula as

$$u(y) = \mathsf{E}(X \mid Y = y) = \sum_{i=1}^{k} x_i p_{X|Y}(x_i \mid y).$$

If $y = Y$ then we obtain the random variable $u(Y)$, which, by definition, is the conational expectation of X given Y, written $\mathsf{E}(X \mid Y)$.

A random variable X is called $\mathscr{D}$-measurable if $\mathscr{D}_X \preccurlyeq \mathscr{D}$, i.e., if it can be represented as

$$X = \sum_{i=1}^{n} x_i \, \mathbf{1}_{D_i},$$

where some x_i can be equal.

Problem 4.19. For which random variable X, it is Ω-measurable?

Obviously, any X is $\mathscr{D}_X$-measurable.

Problem 4.20. Show that if X is $\mathscr{D}$-measurable, then

$$\mathsf{E}(YX \mid \mathscr{D}) = X\,\mathsf{E}(Y \mid \mathscr{D}).$$

Problem 4.21. Prove that if $\mathscr{D}_1$ and $\mathscr{D}_2$ are two decompositions, such that $\mathscr{D}_1 \preccurlyeq \mathscr{D}_2$, then

$$\mathsf{E}(\mathsf{E}(X \mid \mathscr{D}_2) \mid \mathscr{D}_1) = \mathsf{E}(X \mid \mathscr{D}_1).$$

Note that the special case is

$$\mathsf{E}(\mathsf{E}(X \mid Y_1, Y_2) \mid Y_1) = \mathsf{E}(X \mid Y_1).$$

Problem 4.22. Show that

$$\mathsf{E}(X \mid Y) = \mathsf{E}\,X,$$

if X and Y are independent. Also show

$$\mathsf{E}(X \mid X) = X.$$

Note that the first equality can be generalized in the following way: We say that X is independent of $\mathscr{D}$ if X and $\mathbf{1}_{D_i}$ are independent for any i, then

$$\mathsf{E}(X \mid \mathscr{D}) = X.$$

Problem 4.23. Let X and Y be i.i.d. r.v., each of which takes values 0 and 1 with probabilities q and p. Show that

$$\mathsf{E}(X + Y \mid Y) = p + Y.$$

To conclude, in literature the notion of the conditional expectation with respect to sigma algebra is used. For us sigma algebra is the same as algebra (as far as we are dealing only with finite sample spaces). Recall that for each decomposition $\mathscr{D}$ there exists algebra $\mathscr{B} = \alpha(\mathscr{D})$. And the conditional expectation with respect to algebra $\mathscr{B}$ should be understood as

$$\mathsf{E}(X \mid \mathscr{B}) = \mathsf{E}(X \mid \mathscr{D}).$$

Problem 4.24. Let X and Y be random variables. Show that $\inf_f \mathsf{E}\,(X - f(Y))^2$ is attained for $f(Y) = \mathsf{E}(X \mid Y)$. This means that conditional expectation of X given Y is the best mean square estimator of X in terms of Y.

4.3.4 Preferential attachment model and its rigorous analysis

Preferential attachment model

The first difference with the random graph models which we considered[5] before is that now we will study a random graph process: i.e., a sequence of random graphs indexed by a "time" variable t. I will denote this sequence $(G_t)_{t\in\mathbf{N}}$. Explicit evolution of the random graph with time is therefore included in the definition of the model. Sometimes such models of random graphs are called *on-line*, opposite to the *off-line* or static random graphs such as Erdős–Rényi model, the configuration model, or the small world model.

Hence we are studying the sequence $G_0, G_1, \dots, G_t, \dots$ To define the model we need to specify the rules 1) how this sequence starts, and 2) how we can obtain G_{t+1} provided that G_t is given. Let G_0 be an

[5]Jordan, Jonathan. "The degree sequences and spectra of scale-free random graphs." Random Structures & Algorithms 29.2 (2006): 226-242.

initial arbitrary finite graph with n_0 vertices and e_0 edges. If we are given G_t, then G_{t+1} is constructed by adding one additional vertex v to the graph and connecting it to m vertices from G_t, the probability of connection being proportional to the degrees of vertices in G_t. This is what is called the *preferential attachment*. Formally, the probability to connect to vertex w is given by

$$\frac{\deg w}{\sum_{u\in V(G_t)} \deg u},$$

and all the connections are independent. Note that this means that it is possible to have multiple edges. We have that

$$|V(G_t)| = n_0 + t, \quad |E(G_t)| = mt + e_0,$$

therefore

$$\sum_{u\in V(G_t)} \deg u = 2(mt + e_0) = c_t t,$$

where $c_t \to 2m$ as $t \to \infty$.

This description fully specifies the model, which has two parameters: initial graph G_0 and the number m of connections of the new vertex. It turns out that the initial graph is not important for the major conclusion, hence I will use only the parameter m to denote the model: $\mathcal{G}(m) = (G_k)_{k\geq 0}$.

Let $N_{k,t}$ be the random variable that is equal to the number of vertices of degree k in G_t. The goal of this section is to prove that the degree distribution of $N_{k,t}$ follows the power law. For this it is convenient to consider the scaled random variable $N_{k,t}/t$, which basically gives us the proportion of the vertices of degree k (to get the exact proportion we would need to divide by $t + n_0$, but since our results are asymptotical, n_0 will not play any role). The usual strategy to prove that we have asymptotically power law degree distribution is

1. Show that the expectations of $N_{k,t}/t$ converge to the power law.

2. Show that $N_{k,t}$ concentrate around $\mathsf{E}N_{k,t}$.

Before starting the main proof, I consider a useful axillary lemma.

Lemma 4.4. *Let x_t, y_t, η_t, r_t be real numbers satisfying*

$$x_{t+1} - x_t = \eta_{t+1}(y_t - x_t) + r_{t+1}, \quad t \in \mathbf{N},$$

and

1. $\lim_{t\to\infty} y_t = x$;

2. $\eta_t > 0$, $t \in \mathbf{N}$, *and* $\eta_t < 1$ *for sufficiently large* t;

3. $\sum_{t=1}^{\infty} \eta_t = \infty$;

4. $\lim_{t\to\infty} \frac{r_t}{\eta_t} = 0$.

Then

$$\lim_{t\to\infty} x_t = x.$$

Proof. First note that due to *4.* we have

$$\eta_{t+1}\left((y_t - x_t) + \frac{r_{t+1}}{\eta_{t+1}}\right) \to \eta_{t+1}(y_t - x_t), \quad t \to \infty,$$

therefore we can consider

$$x_{t+1} - x_t = \eta_{t+1}(y_t - x_t),$$

from where

$$x_{t+1} = x_t(1-\eta_{t+1}) + \eta_{t+1} y_t.$$

Now assume that we chose such large N that $|y_t - x| < \epsilon/2$ for an arbitrary $\epsilon > 0$ (we can do this due to *1*). Let $x_t > x - \epsilon$ and consider

$$x_{t+1} = x_t(1-\eta_{t+1}) + \eta_{t+1} y_t > x - \epsilon + \eta_{t+1}(\epsilon + y_t - x) > x - \epsilon,$$

where I used *2*. Analogously, assuming that $x_t < x + \epsilon$, we can show that it implies that

$$x_{t+1} < x + \epsilon.$$

If $x_t < x - \epsilon$, then

$$x_{t+1} - x_t > \eta_{t+1}(y_t - x + \epsilon) > \eta_{t+1}\frac{\epsilon}{2},$$

therefore

$$\sum_{t=N}^{\infty}(x_{t+1} - x_t) > \sum_{t=N}^{\infty}\eta_{t+1}\frac{\epsilon}{2} = \infty$$

due to *3*. But the partial sum on the left hand side is given by $x_{N+k} - x_N$, therefore $x_t \to \infty$ which contradicts the assumption that $x_t < x - \epsilon$. Similarly, we can arrive to contradiction starting from $x_t > x + \epsilon$. Therefore, putting everything together, for sufficiently large t,

$$|x_t - x| \leq \epsilon,$$

which finishes the proof. ■

Proof that $(\mathsf{E}N_{k,t})_{k\in\mathbf{N}}$ converges to the power law distribution

Theorem 4.5. *In the random graph process $\mathcal{G}(m)$*

$$\mathsf{E}\left(\frac{N_{k,t}}{t}\right) \to \frac{2m(m+1)}{k(k+1)(k+2)}, \quad \textit{when } t \to \infty,$$

for $k \geq m \geq 1$.

Note that this theorem gives the power law degree distribution with the exponent $\alpha = 3$.

Proof. Consider the sequence of random graphs $(G_0, G_1, \ldots, G_t)$. This sequence is an event in the algebra $\mathscr{G}_t = \alpha(G_0, \ldots, G_t)$. Therefore, we can consider the conditional expectation of $N_{k,t+1}$ with respect to $\mathscr{G}_t$, for which there is a unique decomposition $\mathscr{D}_t$. This algebra is also generated by the decomposition $\mathscr{D}_t = \mathscr{D}_{(N_{k,l})_{k\in\mathbf{N}, l=1,\ldots,t}}$, because of the model definition. Thence, the conditional expectation with respect to $\mathscr{G}_t$ can be expressed as

$$\mathsf{E}(N_{k,t+1} \mid \mathscr{G}_t) = \sum_{d=0}^{m} N_{k-d,t}\binom{m}{d}\left(1 - \frac{k-d}{c_t t}\right)^{m-d}\left(\frac{k-d}{c_t t}\right)^{d} + \delta_{m,k}.$$

This formula can be explained as follows: Given the graph G_t with $N_{k-d,t}$ vertices of degree $k-d$, the expected number of vertices of degree k in G_{t+1} is

$$N_{k-d,t}\binom{m}{d}\left(1 - \frac{k-d}{c_t t}\right)^{m-d}\left(\frac{k-d}{c_t t}\right)^{d},$$

we need to sum though all possible d, and $\delta_{m,k}$ is the Kronecker delta, which is equal to 1 if $m = k$ and to 0 if $m \neq k$. This term accommodates one additional vertex that we add to our graph at the each step.

Next, consider the expression

$$\mathsf{E}\left(\frac{N_{t,k+1}}{t+1} \mid \mathscr{G}_t\right) - \frac{N_{k,t}}{t},$$

which is given by

$$\frac{1}{t+1}\left(\sum_{d=0}^{m}\frac{N_{k-d,t}}{t}t\binom{m}{d}\left(\frac{c_t t-k+d}{c_t t}\right)^{m-d}\left(\frac{k-d}{c_t t}\right)^{d}+\delta_{m,k}(t+1)-\frac{N_{k,t}}{t}(t+1)\right).$$

The idea now is to consider only those terms in this expression, which do not approach zero. This means, e.g, that we need to keep track of mostly $d=0$ and $d=1$ hence, as it should be intuitively obvious, the chance to have two edges that connect the new vertex to the same vertex in G_t are very small. After some rearrangements and using the binomial formula, we find

$$\frac{m}{(t+1)c_t}\left(\frac{N_{k-1,t}}{t}(k-1)+\frac{c_t}{m}\delta_{m,k}-\frac{N_{k,t}}{t}\left(k+\frac{c_t}{m}\right)+\sum_{j=1}^{m}t^{-j}\sum_{d=0}^{j}C_{k,j,d,t}\frac{N_{k-d,t}}{t}\right),$$

where $C_{k,j,d,t}$ are bounded in t. Now we can use the properties of the conditional expectation (specifically property 5 in Problem 4.18), and linearity of the expectations to find, writing $x_{k,t}=\mathsf{E}\left(\frac{N_{k,t}}{t}\right)$, that

$$x_{k+1,t+1}-x_{k,t}=\frac{m}{c_t(t+1)}\left(k+\frac{c_t}{m}\right)\left(\frac{k-1}{k+c_t/m}x_{k-1,t}+\frac{c_t/m}{k+c_t/m}\delta_{mk}-x_{k,t}\right)+r_{t+1},$$

where the form r_{t+1} should be clear.

There are three cases to consider. If $k<m$ and since each new vertex has degree m, there are at most $|V(G_0)|$ vertices of degree k, therefore, when $t\to\infty$,

$$x_{k,t}\to 0,\quad 0\leq k\leq m-1.$$

Second case: $k=m$, here we have $\delta_{mm}=1$. Consider

$$\begin{aligned} x_t &= x_{m,t}=\mathsf{E}\left(\frac{N_{m,t}}{t}\right),\\ y_t &= \frac{m-1}{m+c_t/m}x_{m-1,t}+\frac{c_t/m}{m+c_t/m},\\ \eta_{t+1} &= \frac{m}{c_t(t+1)}\left(m+\frac{c_t}{m}\right). \end{aligned}$$

Note that all the conditions of the lemma in the previous subsection are fulfilled (check them all carefully), and $y_t\to\frac{2}{m+2}$, therefore,

$$x_{m,t}\to\frac{2}{m+2}.$$

Finally consider case $k>m$. Here we have $\delta_{mk}=0$ and the expression for y_t will change to

$$y_t=\frac{k-1}{k+c_t/m}x_{k-1,t}.$$

Now we can proceed by induction. Note that we obtain

$$\lim_{t\to\infty}x_{k,t}=\frac{2}{m+2}\prod_{l=m+1}^{k}\frac{l-1}{l+2}=\frac{2m(m+1)}{k(k+1)(k+2)},$$

since the product is telescoping. This finishes the proof. ■

Concentration result for the degree distribution

Theorem 4.6. *In preferential attachment model $\mathcal{G}(m)$,*

$$\operatorname{Var}\left(\frac{N_{k,t}}{t}\right) \to 0, \quad \text{as } t \to \infty.$$

Proof. See, e.g., A. Bonato, A Course on the Web Graph, AMS, 2008 ■

Generalizations

The analysis above shows that for $\mathcal{G}(m)$ we obtain the power law distribution with the exponent $\alpha = 3$. It would be desirable to modify the model to accommodate a range of exponents. It can be done by using the following modification. We chose a fixed $q \in (-m, \infty)$ and redefine the probability that the new vertex is connected by one of m edges to the vertex w as

$$\frac{q + \deg w}{\sum_{u \in V(G_t)} (q + \deg u)}.$$

Call this model now $\mathcal{G}(q, m)$.

Theorem 4.7. *In $\mathcal{G}(q, m)$ model for $t \to \infty$*

1.
$$\mathsf{E}\left(\frac{N_{k,t}}{t}\right) \to \frac{(2 + q/m)\Gamma(3 + q/m + m + q)}{\Gamma(2 + q/m + q + m)\Gamma(m + q)} \frac{\Gamma(k + q)}{\Gamma(3 + q/m + k + q)},$$

2.
$$\operatorname{Var}\left(\frac{N_{k,t}}{t}\right) \to 0\,.$$

In the statement of the theorem $\Gamma(x)$ is the gamma function.

Problem 4.25. Two special functions often appear in analysis of power law distributions: the gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \mathrm{d}t\,,$$

and Euler's beta function

$$\mathrm{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}\,.$$

- Show that $\Gamma(x + 1) = x\Gamma(x)$, hence showing that the gamma function generalizes the definition of factorial to non integer values.
- Show that (using the Stirling's formula)

$$\frac{\Gamma(x + \alpha)}{\Gamma(x)} = x^\alpha (1 + o(x))$$

- From the previous point find that

$$\mathrm{B}(x, y) = x^{-y} \Gamma(y)(1 + o(x)).$$

Aiello, Chung, and Lu model

Consider a sequence of random graphs that can be generated using the following algorithm[6]. Let G_0 be a graph consisting of 1 vertex and one self-loop. Define a vertex step as adding a new vertex v and adding a new edge uv so that u is chosen from the existing vertices by the preferential attachment. An edge-step consists of adding an edge rs, where r and s are randomly chosen from the existing vertices by means of preferential attachment (i.e., r and s can coincide). To form G_{t+1} with probability p (this is the only parameter of the model) take a vertex step, and with probability $1-p$ take an edge-step. We obtain that the number of edges in G_t is t and the number of vertices is a random variable with the average $1+pt$. Denote this model as $\mathcal{G}(p)$.

Theorem 4.8. *In $\mathcal{G}(p)$ model*

-
$$\mathsf{E}\left(\frac{N_{1,t}}{t}\right) \to \frac{2p}{4-p},$$

- *For $k>1$*
$$\mathsf{E}\left(\frac{N_{k,t}}{t}\right) \to \frac{2p}{4-p}\frac{\Gamma(k)\Gamma(2+2/(2-p))}{\Gamma(k+1+2/(2-p))}.$$

-
$$N_{k,t} = L_k t + \mathcal{O}(2\sqrt{k^3 t\log t}) \quad a.a.s.,$$

  where
$$L_k = \frac{2p}{4-p}\frac{\Gamma(k)\Gamma(2+2/(2-p))}{\Gamma(k+1+2/(2-p))}.$$

This theorem shows that in $\mathcal{G}(p)$ model we observe that the sequence of degree distribution concentrates around the power law with the exponent $\alpha = 1+2/(2-p)$, hence $\alpha\in[2,3]$.

Bollobás–Riordan model

Here I briefly describe one of the first preferential attachment model that has been rigorously analyzed[7]. First consider the dynamical picture.

Let G_1^1 be the initial graph that consists of unique vertex and one self-loop. Now to build G_t^1 from G_{t-1}^1 we add a vertex t and one edge connecting vertices i and t with probability

$$\mathsf{P}(t\sim i) = \begin{cases} \dfrac{\deg i}{2t-1}, & i\neq t,\\ \dfrac{1}{2t-1}, & i=t. \end{cases} \tag{4.2}$$

The first line describes the preferential attachment principle, and the probabilities are chosen in a way to guarantee a proper normalization:

$$\sum_{i=1}^{t-1}\frac{\deg i}{2t-1}+\frac{1}{2t-1} = \frac{2t-t}{2t-1}+\frac{1}{2t-1}=1.$$

At step t our random graph G_t^1 has exactly t vertices and t edges. How to obtain a graph G_t^m such that at each step m edges are added? Consider G_{tm}^1 and partition its vertices into t sets with m elements each:

$$\{1,\dots,m\},\ \{m+1,\dots,2m\},\dots,\{m(t-1),\dots,mt\}.$$

[6] Aiello, William, Fan Chung, and Linyuan Lu. "Random evolution in massive graphs." Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on. IEEE, 2001.

[7] Bollobá s, Béla, and Oliver Riordan. Mathematical results on scale-free random graphs. Handbook of graphs and networks 1 (2003): 34.

Now each of these sets is defined to be a vertex of new graph, which has exactly t vertices and tm edges, which all kept intact. Hence as a result we obtain a multigraph with possible self loops whose evolution is governed by the preferential attachment principle.

Now to present the static incarnation of the same problem, consider a combinatorial object called *linearized chord diagram.* Assume that we have $2t$ points:

$$1, 2, 3, \dots, 2t-1, 2t.$$

Consider all the possible pairing of these points, the total number is

$$(2t-1)!! = \frac{(2t)!}{2^t t!}\,.$$

Now to choose one random realization, pick uniformly one of these $(2t-1)!!$ pairings. To build a graph, imagine these $2t$ points on a straight line, and those that are paired connect with a chord. Now move in the direction from 1 to $2t$ (from left to right) until we find the right end of one of the chords, these points that we passed form the first vertex, then we again move to the right until we find next right end of a cord, and so on. Hence we get exactly t vertices (since there are t right ends), and each chord denotes an edge in this graph. The key observation here is that random graph, obtained in such a way has probabilities for existence of edges exactly as in (4.2).

Problem 4.26. Give a formal proof of the last statement.

Analyzing the linearized chord diagram, it is possible to obtain some important results. In particular,

Theorem 4.9. *For any $m \geq 2$ and any $\varepsilon > 0$*

$$\mathsf{P}\left((1-\varepsilon)\frac{\log t}{\log\log t} \leq \operatorname{diam} G_t^m \leq (1+\varepsilon)\frac{\log t}{\log\log t}\right) \to 1.$$

Theorem 4.10. *For any $m \geq 1$ and any $k \leq t^{1/15}$*

$$(1-\varepsilon)\alpha_{m,k} \leq \frac{N_{k,t}}{t} \leq (1+\varepsilon)\alpha_{m,k},$$

where

$$\alpha_{m,k} = \frac{2m(m+1)}{(k+m)(k+m+1)(k+m+2)} \sim Ck^{-3}.$$

Proof. ■

The copying model

For full details see the paper[8].

One of significant disadvantages of the preferential attachment principle is that it requires the global knowledge of the current network structure to decide to which vertices a new vertex will be connected. This is of course not realistic.

Let $\alpha \in (0,1)$ and fix $d \geq 2$, $d \in \mathbf{N}$. For the initial graph in our random graph process consider any d-regular graph. Let us have $G_t = (V_t, E_t)$, where the set of vertices is $V_t = \{v_1, \dots, v_s\}$. To construct G_{t+1} we add one vertex and d edges. The exact rules as follows: We pick a uniformly random vertex, let us call it w. w has at least d neighbors. After this with probability α we connect v_{s+1} with a random vertex from G_t and with probability $1-\alpha$ with one of the neighbors of w. Hence we actually need only local structure of w. Here is the main result.

[8]Kumar, Ravi, et al. "Stochastic models for the web graph." Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on. IEEE, 2000.

Theorem 4.11. *If $k > 0$, then*

$$\frac{N_{k,t}}{t} = \Theta\left(k^{-\frac{2-\alpha}{1-\alpha}}\right)$$

a.a.s.

Hence we obtain the power law distribution with the exponent $\frac{2-\alpha}{1-\alpha}$. It is important to understand that despite the fact that this model also produces a power law graph, its structure is very different from the models produced with the preferential attachment principle.

4.4 Small world model