3.2 Non uniform and growth models of Random Graphs

3.2.1 Non uniform models

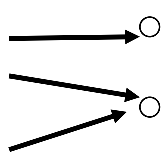
So far we have considered random graphs in which all vertices have the same expected degree. However, large graphs occurring in the real world tend to have power law degree distributions. This raises the question of when phase transitions occur in random graphs with arbitrary degree distributions. In this section, we consider when a random graph with a non uniform degree distribution has a giant component.

3.2.2 Molloy/Reed

Molloy/Reed address the issue of when a random graph with a non uniform degree distribution has a giant component. Let λ_i be the fraction of vertices of degree *i*. There will be a giant component if

$$\sum_{i=0}^{\infty} i(i-2)\lambda_i > 0.$$

To intuitively see that this is the correct formula, consider exploring a component starting from a given seed vertex. Degree zero vertices do not occur except in the case where the vertex is the seed. If we encounter a degree one vertex, then that terminates the expansion along the edge into the vertex. Thus, we do not want to encounter too many degree one vertices. A degree two vertex is neutral in that we come in one edge and go out the other. We get no net increase in the frontier. Vertices of degree *i* greater than two increase the frontier by *i-2* edges. We enter the vertex by one of its edges and thus have *i-1* new edges in the frontier for a net gain of *i-2*. The $i\lambda_i$ in $i(i-2)\lambda_i$ is proportional to the probability that of reaching a degree *i* vertex and i-2 accounts for the increase or decrease in size of the frontier when a degree *i* vertex is reached.



Consider a graph in which half of the vertices are degree one and half are degree two. If a vertex is selected at random it is equally likely to be degree one or degree two. However, if we select an edge at random and walk to its endpoint, it is twice as likely to be degree two as degree one. In many graph algorithms, a vertex is reached by randomly selecting an edge and traversing the edge to reach an endpoint. In this case, the probability of reaching a degree i vertex is proportional to $i\lambda_i$ where λ_i is the fraction of vertices that are degree i.

Example 3. XXX: Consider applying the Molloy Reed conditions to the $G(n,p)$ model. The summation gives value zero precisely when $p=\frac{1}{n}$, the point at which the phase transition occurs. At $p=\frac{1}{n}$, the average degree of each vertex is one and there are n/2 edges. However, the actual degree distribution of the vertices is binomial where the probability that a vertex is of degree i is given by $p_i=\binom{n}{i}p^i(1-p)^{n-i}$. We now show that $\lim_{n\to\infty}\sum_{i=0}^{n} i(i-2)p_i=0$ for $p_i=\binom{n}{i}p^i(1-p)^{n-i}$ when p=1/n.

$$\lim_{n\to\infty}\sum_{i=0}^{n} i(i-2)\binom{n}{i}\left(\frac{1}{n}\right)^i(1-\frac{1}{n})^{n-i}=\lim_{n\to\infty}\sum_{i=0}^{n} i(i-2)\frac{n(n-1)\cdots(n-i+1)}{i!n^i}\left(1-\frac{1}{n}\right)^{n-i}$$

$$=\frac{1}{e}\lim_{n\to\infty}\sum_{i=0}^{n} i(i-2)\frac{n(n-1)\cdots(n-i+1)}{i!n^i}\left(\frac{n}{n-1}\right)^i\le\sum_{i=0}^{\infty}\frac{i(i-2)}{i!}$$

To see that $\sum_{i=0}^{\infty}\frac{i(i-2)}{i!}=0$ note that

$$\sum_{i=0}^{\infty}\frac{i}{i!}=\sum_{i=1}^{\infty}\frac{i}{i!}=\sum_{i=1}^{\infty}\frac{1}{(i-1)!}=\sum_{i=0}^{\infty}\frac{1}{i!}$$

and

$$\sum_{i=0}^{\infty}\frac{i^2}{i!}=\sum_{i=1}^{\infty}\frac{i^2}{i!}=\sum_{i=1}^{\infty}\frac{i}{(i-1)!}=\sum_{i=0}^{\infty}\frac{i+1}{i!}=\sum_{i=0}^{\infty}\frac{i}{i!}+\sum_{i=0}^{\infty}\frac{1}{i!}=2\sum_{i=0}^{\infty}\frac{1}{i!}.$$

Thus $\sum_{i=0}^{\infty}\frac{i(i-2)}{i!}=\sum_{i=0}^{\infty}\frac{i^2}{i!}-2\sum_{i=0}^{\infty}\frac{i}{i!}=0$.

■

In the limit as n goes to infinity the distribution $p_i=\binom{n}{i}p^i(1-p)^i$ becomes Poisson and $p_i=e^m\frac{m^i}{i!}$ where when $p=\frac{1}{n}$, the mean $m=np=1$. Values for p_i are given in the following table.

Degree	0	1	2	3	4	5	6
Probability	0.3679	0.3679	0.1840	0.0613	0.0153	0.0031	0.0005

What is purpose of above table. Do we want to give example of exponential degree distribution?

3.4 Growth models

3.4.1 Growth model without preferential attachment

Many graphs that arise in the outside world started as small graphs that grew over time. In considering a model for such graphs there are two ways in which to select the vertices for attaching a new edge. The first is to select two vertices uniformly at random from the set of existing vertices. The second is to select two vertices with probability proportional to their degree. This latter method is referred to as preferential attachment.

We now consider a growth model for a random graph without preferential attachment. Start with zero vertices at time zero. At each unit of time a new vertex is created and with probability δ, two vertices chosen at random are joined by an edge.

The degree distribution for this growth model is calculated as follows. Let $D_k(t)$ be the number of vertices of degree k at time t and let $d_k(t)$ be the expectation of $D_k(t)$. The number of isolated vertices, $D_0(t)$, increases by one at each unit of time and decreases by the number $b(t)$ of vertices which are picked to be end points of the new edge (if any added) at time t. [$b(t)$ can take on values 0,1 or 2.] Taking expectations, we get

Comment [k1]: John: I made a slight change here with D_k and d_k which makes this part rigorous.

$$d_0(t+1) = d_0(t) + 1 - E(b(t)).$$

$b(t)$ is the sum of two random variables: $b_1(t)$, which is the number of degree 0 vertices picked to be the first end point of the new edge and $b_2(t)$ which is the number of vertices picked to be the second end point of the edge. These are dependent, but we certainly have $Eb_1(t) = Eb_2(t) = \delta \frac{d_0(t)}{t}$ and $Eb(t) = Eb_1(t) + Eb_2(t)$. Thus,

$$d_0(t+1) = d_0(t) + 1 - 2\delta \frac{d_0(t)}{t}$$

The number of vertices of degree k , increases whenever a new edge is added to a degree k-1 vertex and decreases when a new edge is added to a degree k vertex. Thus, reasoning as above, we get

$$d_k(t+1) = d_k(t) + 2\delta \frac{d_{k-1}(t)}{t} - 2\delta \frac{d_k(t)}{t}$$

Consider a solution of the form $d_k(t) = p_k t$. Then

$$(t+1)\, p_0 = p_0 t + 1 - 2\delta \frac{p_0 t}{t}$$

$$p_0 = 1 - 2\delta p_0$$

$$p_0 = \frac{1}{1+2\delta}$$

and

$$(t+1)\, p_k = p_k t + 2\delta \frac{p_{k-1} t}{t} - 2\delta \frac{p_k t}{t}$$

$$p_k = 2\delta p_{k-1} - 2\delta p_k$$

$$p_k = \frac{2\delta}{1+2\delta}\, p_{k-1}$$

$$p_k = \left(\frac{2\delta}{1+2\delta}\right)^k p_0$$

$$p_k = \frac{1}{1+2\delta}\left(\frac{2\delta}{1+2\delta}\right)^k$$

Thus, the model gives rise to a graph with a degree distribution that falls off exponentially with degree.

Generating function for component size **CHANGE g TO h_0 ?**

Let $N_k(t)$ be the expected number of components of size k at time t. Another way of expressing this is that $N_k(t)$ is proportional to the probability that a randomly picked component is of size k. Note this is not the same as the component containing a randomly picked vertex – see the picture. Indeed the probability that the size of the component containing a randomly picked vertex is k is proportional to $kN_k(t)$. We will also use the quantities $a_k(t) = \frac{N_k(t)}{t}$. We assume that the numbers $a_k(t)$ are independent of t and denote them by just a_k. We will justify this by formulating recursive equations for $N_k(t)$ and showing that a solution of them has the property that $\frac{N_k(t)}{t}$ are independent of t. After showing this, we focus on the generating function g(x) of the numbers $ka_k(t)$ and use g(x) to show the threshold for giant components.

Figure: In selecting a component at random, each of the two components is equally likely to be selected. In selecting the component containing a random vertex, the larger component is twice as likely to be selected.

Consider $N_1(t)$, the expected number of isolated vertices at time t. At each time unit, an isolated vertex is added to the graph and an expected $\frac{2\delta N_1(t)}{t}$ many isolated vertices will be chosen for attachment and thereby leave the set of isolated vertices. Thus

$$N_1(t+1) = N_1(t) + 1 - 2\delta \frac{N_1(t)}{t}$$

For k>1, $N_k(t)$ increases when two smaller components whose sizes sum to k are joined by an edge and decreases when a vertex in a component of size k is chosen for attachment. The probability that a vertex selected at random will be in a size k component is $\frac{kN_k(t)}{t}$. Thus

$$N_k(t+1) = N_k(t) + \delta \sum_{j=1}^{k-1} \frac{jN_j(t)}{t} \frac{(k-j)N_{k-j}(t)}{t} - 2\delta \frac{kN_k(t)}{t}$$

This is not precise: we really ought to consider the actual number of components of various sizes, rather than expected numbers. Also, if we chose both vertices between which the new edge will be put in to be in the same k-vertex component, then $N_k(t)$ does not really go down as claimed. But we ignore these small inaccuracies here.

Comment [k2]: We may want to add this.

Consider solutions of the form $N_k(t) = a_k t$. Solving for a_k just as we solved for p_k yields

$a_1 = \frac{1}{1+2\delta}$ and $a_k = \frac{\delta}{1+2k\delta} \sum_{j=1}^{k-1} j(k-j)a_j a_{k-j}$.

The quantity a_k is proportional to the probability that a component selected at random is of size k but is not actually a probability since $\sum_{k=0}^{\infty} a_k \neq 1$. The quantity ka_k is, however, the probability that a randomly chosen vertex is in a component of size k since $\sum_{k=0}^{\infty} ka_k = \sum_{k=0}^{\infty} \frac{kN_k}{t} = 1$.

Consider the generating function g(x) for the distribution of component sizes where the coefficient of x^k is the probability that a vertex chosen at random is in a component of size k

$$g(x) = \sum_{k=1}^{\infty} ka_k x^k$$

From the formula for the a_i's, we will derive the differential equation:

$$g = -2\delta xg' + 2\delta xgg' + x$$

and then use the equation for g to determine the value of δ at which the phase transition for the appearance of the giant component occurs.

Derivation of g(x)

From

$$a_1 = \frac{1}{1+2\delta}$$

and

$$a_k = \frac{\delta}{1+2k\delta} \sum_{j=1}^{k-1} j(k-j) a_j a_{k-j}$$

we derive the equations

$$a_1 \left(1+2\delta\right) - 1 = 0$$

and

$$a_k \left(1+2k\delta\right) = \delta \sum_{j=1}^{k-1} j(k-j) a_j a_{k-j}$$

The generating function is formed by multiplying the k[th] equation by kx^k and summing over all k. This gives

$$-x+\sum_{k=1}^{\infty}ka_kx^k+2\delta x\sum_{k=1}^{\infty}a_kk^2x^{k-1}=\delta\sum_{k=1}^{\infty}kx^k\sum_{j=1}^{k-1}j(k-j)a_ja_{k-j}$$

Note that $g(x)=\sum_{k=1}^{\infty}ka_kx^k$, and $g'(x)=\sum_{k=1}^{\infty}a_kk^2x^{k-1}$.

Thus

$$-x+g(x)+2\delta xg'(x)=\delta\sum_{k=1}^{\infty}kx^k\sum_{j=1}^{k-1}j(k-j)a_ja_{k-j}$$

Working with the right hand side

$$\delta\sum_{k=1}^{\infty}kx^k\sum_{j=1}^{k-1}j(k-j)a_ja_{k-j}=\delta x\sum_{k=1}^{\infty}\sum_{j=1}^{k-1}j(k-j)(j+k-j)x^{k-1}a_ja_{k-j}$$

Now breaking the j+k-j into two sums gives

$$\delta x\sum_{k=1}^{\infty}\sum_{j=1}^{k-1}j^2a_jx^{j-1}(k-j)a_{k-j}x^{k-j}+\delta x\sum_{k=1}^{\infty}\sum_{j=1}^{k-1}ja_jx^j(k-j)^2a_{k-j}x^{k-j-1}$$

Notice that the second sum is obtained from the first by substituting k-j for j and that both terms are $\delta xg'g$. Thus

$$-x+g(x)+2\delta xg'(x)=2\delta xg'(x)g(x)$$

Hence

$$g'=\frac{1}{2\delta}\frac{1-\frac{g}{x}}{1-g}.$$

Phase transition for giant component

The generating function g(x) contains information about the finite components of the graph. A finite component is a component of size 1, 2, … which does not depend on n. Observe that $g(1)=\sum_{k=0}^{\infty}ka_k$ and hence g(1) is the probability that a randomly chosen vertex will belong to a component of finite size. Thus, if g(1)=1 there is no giant component. When $g(1)\neq 1$, then 1-g(1) is the expected fraction of the vertices that are in the giant component.

John: I am not clear on how we get this last statement. What I see is that 1-g(1) is the expected fraction of vertices which are in components of sizes going to infinity. Why should they all be in one giant component ??

We now calculate the value of δ at which the giant component phase transition occurs. Recall that the generating function for $g(x)$ satisfies

$$g'(x)=\frac{1}{2\delta}\frac{1-\frac{g(x)}{x}}{1-g(x)}$$

If δ is greater than some $\delta_{critical}$, then a giant component exists and $g(1)\neq 1$. In this case the above formula simplifies. The average size of the finite components, is $g'(1)=\sum_{k=1}^{\infty}k^2 a_k$, since ka_k is the probability that a randomly chosen vertex is in a a component of size k. Now, $g'(1)$ is given by

(A1) $$g'(1)=\frac{1}{2\delta} \text{ for all } \delta>\delta_{critical}.$$

If δ is less than $\delta_{critical}$, then the giant component does not exist. In this case g(1)=1 and both the numerator and the denominator approach zero. Appling L'Hopital's rule

$$\lim_{x\to 1} g'(x)=\frac{1}{2\delta}\left.\frac{\frac{xg'(x)-g(x)}{x^2}}{g'(x)}\right|_{x=1} \quad \text{or} \quad \left(g'(1)\right)^2=\frac{1}{2\delta}\left(g'(1)-g(1)\right).$$

The quadratic $\left(g'(1)\right)^2-\frac{1}{2\delta}g'(1)+\frac{1}{2\delta}g(1)=0$ has solutions

(A2) $$g'(1)=\frac{\frac{1}{2\delta}\pm\sqrt{\frac{1}{4\delta^2}-\frac{4}{2\delta}}}{2}=\frac{1\pm\sqrt{1-8\delta}}{4\delta}.$$

The two solutions given by (A2) become complex for $\delta>\frac{1}{8}$ and thus can be only valid for $0\le\delta\le\frac{1}{8}$. For $\delta>\frac{1}{8}$ the only solution is $g'(1)=\frac{1}{2\delta}$ and a giant component exists. As δ is decreased, at $\delta=\frac{1}{8}$ there is a singular point where for $\delta<\frac{1}{8}$ there are three possible solutions, one from (A1) which implies a giant component and two from (A2) which imply no giant component. To determine which one of the three solutions is valid, consider the limit as $\delta\to 0$. In the limit all components are of size one since there are no edges. Only (A2) with the minus sign gives the correct solution

$$g'(1) = \frac{1-\sqrt{1-8\delta}}{4\delta} = \frac{1-\left(1-\frac{1}{2}8\delta-\frac{1}{4}64\delta^2+\cdots\right)}{4\delta} = 1+4\delta+\cdots = 1 .$$

In the absence of any non analytic behavior in the equation for $g'(x)$ in the region $0 \le \delta < \frac{1}{8}$, we conclude that (A2) with the minus sign is the correct solution for $0 \le \delta < \frac{1}{8}$ and hence the critical value of δ for the phase transition is $\frac{1}{8}$. As we shall see this is different from the static case.

As the value of δ is increased the average size of the finite components increase from one to

$$\left.\frac{1-\sqrt{1-8\delta}}{4\delta}\right|_{\delta=\frac{1}{8}} = 2$$

when δ reaches the critical value of $\frac{1}{8}$. At $\delta = \frac{1}{8}$, the average size of the finite components jumps to $\left.\frac{1}{2\delta}\right|_{\delta=\frac{1}{8}} = 4$ and then decreases as $\frac{1}{2\delta}$ as the giant component swallows up the finite components starting with the larger components.

Comparison to static random graph

Consider a static random graph with the same degree distribution as the grown graph. Again let p_k be the probability of a vertex being of degree k. We use

$$p_k = \frac{(2\delta)^k}{(1+2\delta)^{k+1}} .$$

Recall the Molloy Reed analysis of random graphs with given degree distributions which asserts that there is a phase transition at $\sum_{i=0}^{\infty} i(i-2)p_i = 0$.

Using this, it will be easy to see that for $\delta = \frac{1}{4}$ we get a phase transition.

For $\delta = \frac{1}{4}$, $p_k = \frac{(2\delta)^k}{(1+2\delta)^{k+1}} = \frac{\left(\frac{1}{2}\right)^k}{\left(1+\frac{1}{2}\right)^{k+1}} = \frac{\left(\frac{1}{2}\right)^k}{\frac{3}{2}\left(\frac{3}{2}\right)^k} = \frac{2}{3}\left(\frac{1}{3}\right)^k$ and

$$\sum_{i=0}^{\infty} i(i-2)\tfrac{2}{3}\left(\tfrac{1}{3}\right)^i = \tfrac{2}{3}\sum_{i=0}^{\infty} i^2\left(\tfrac{1}{3}\right)^i - \tfrac{4}{3}\sum_{i=0}^{\infty} i\left(\tfrac{1}{3}\right)^i = \tfrac{2}{3}\tfrac{3}{2} - \tfrac{4}{3}\tfrac{3}{4} = 0 .$$

Footnote: Recall that $1+a+a^2+\cdots = \frac{1}{1-a}$, $a+2a^2+3a^3\cdots = \frac{a}{(1-a)^2}$, and $a+4a^2+9a^3\cdots = \frac{a(1+a)}{(1-a)^3}$

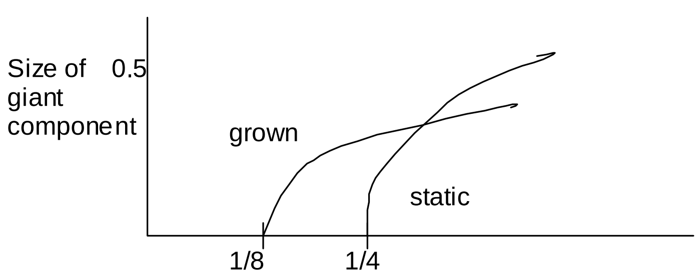


Figure obtained by integrating g'. [How does one integrate g' for x=1].

See references at end of section for calculating the size of the giant component in the static graph. The result is

$$S_{static} = \begin{cases} 0 & \delta \le \frac{1}{4} \\ 1 - \dfrac{1}{\delta + \sqrt{\delta^2 + 2\delta}} & \delta > \frac{1}{4} \end{cases}$$

3.4.2 A Growth Model with Preferential Attachment

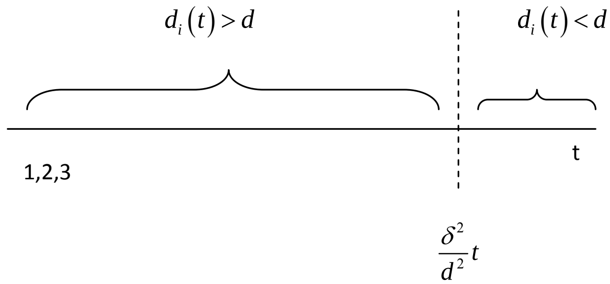
Consider a growth model with preferential attachment. At each time unit a vertex is added to the graph. Then with probability δ an edge is attached to the new vertex and to a vertex selected at random with probability proportional to its degree.

Let $d_i(t)$ be the degree of the i^{th} vertex at time *t*. The sum of the degrees of all vertices at time *t* is $2\delta t$ and thus the probability that an edge is connected to vertex *i* at time *t* is $\frac{d_i(t)}{2\delta t}$. The degree of vertex *i* is governed by the equation

$$\begin{aligned} \frac{d}{dt} d_i(t) &= \delta \frac{d_i(t)}{2\delta t} \\ &= \frac{d_i(t)}{2t} \end{aligned}.$$

where δ is the probability that an edge is added at time *t* and $\frac{d_i(t)}{2\delta t}$ is the probability that the vertex *i* is selected for the end point of the edge.

The two in the denominator governs the solution which is of the form $at^{\frac{1}{2}}$. The value of a is determined by the initial condition $d_i(t) = \delta$ at $t = i$. Thus $\delta = ai^{\frac{1}{2}}$ or $a = \delta i^{-\frac{1}{2}}$. Hence $d_i(t) = \delta\sqrt{\frac{t}{i}}$.



If a vertex is created early enough its expected degree will exceed d. This occurs for $i \le \frac{\delta^2}{d^2}t$.

Next we want to determine the probability distribution of vertex degrees. Now $d_i(t)$ is less than d provided $i > \frac{\delta^2 t}{d^2}$. The fraction of the t vertices at time t for which $i > \frac{\delta^2}{d^2}t$ and thus that the degree is less than d is $1-\frac{\delta^2}{d^2}$. Thus, the probability that a vertex has degree less than d is $1-\frac{\delta^2}{d^2}$. The probability density is obtained from the derivative of $\text{Prob}(\text{degree} < d)$.

$$P(d) = \frac{d}{dd}\left(1-\frac{\delta^2}{d^2}\right) = 2\frac{\delta^2}{d^3},$$

a power law distribution.

Exercises

Exercise: For p asymptotically greater than $\frac{1}{n}$ show that $\sum_{i=0}^{\infty} i(i-2)\lambda_i > 0$.

■

reference

Barabasi – Albert 99 gives proof of power law degree for grown graph with preferential attachment.
Ballobas et al 01 gives rigorous proof.

M. Molloy and B. Reed, Combinatorics, Probab, Comput 7, 295 (1998)