CHAPTER 3

A Generative Model — the Preferential Attachment Scheme

The preferential attachment scheme is often attributed to Herbert Simon. In his paper [**111**] of 1955, he gave a model for word distribution using the preferential attachment scheme and derived *Zipf's law*. Namely, the probability of a word having occurred exactly i times is proportional to $1/i$.

The basic setup for the preferential attachment scheme is a simple *local* growth rule which leads to a *global* consequence — a power law distribution. Since this local growth rule gives preferences to vertices with large degrees, the scheme is often described by "*the rich get richer*".

In this chapter, we shall give a clean and rigorous treatment of the preferential attachment scheme. Of interest is to determine the exponent of the power law from the parameters of the local growth rule.

3.1. Basic steps of the preferential attachment scheme

There are two parameters for the preferential attachment model:

- A probability p, where $0 \leq p \leq 1$.
- An initial graph G_0, that we have at time 0.

Usually, G_0 is taken to be the graph formed by one vertex having one loop. (We consider the degree of this vertex to be 1, and in general a loop adds 1 to the degree of a vertex.) Note, in this model multiple edges and loops are allowed.

We also have two operations we can do on a graph:

- *Vertex-step* — Add a new vertex v, and add an edge $\{u, v\}$ from v by randomly and independently choosing u in proportion to the degree of u in the current graph.
- *Edge-step* — Add a new edge $\{r, s\}$ by independently choosing vertices r and s with probability proportional to their degrees.

Note that for the edge-step, r and s could be the same vertex. Thus loops could be created. However, as the graph gets large, the probability of adding a loop can be well bounded and is quite small.

The random graph model $G(p, G_0)$ is defined as follows:

Begin with the initial graph G_0.
For $t > 0$, at time t, the graph G_t is formed by modifying G_{t-1} as follows:
with probability p, take a vertex-step,
otherwise, take an edge-step.

When G_0 is the graph consisting of a single loop, we will simplify the notation and write $G(p) = G(p, G_0)$.

3.2. Analyzing the preferential attachment model

To analyze the graph generated by the preferential attachment model $G(p)$, we let n_t denote the number of vertices of $G(p)$ at time t and let e_t denote the number of edges of $G(p)$ at time t. We have

$$e_t = t + 1.$$

The number of vertices n_t, however, is a sum of t random indicator variables,

$$n_t = 1 + \sum_{i=1}^{t} s_t$$

where

$$\begin{aligned} \Pr(s_j = 1) &= p, \\ \Pr(s_j = 0) &= 1 - p. \end{aligned}$$

It follows that the expected value $\mathrm{E}(n_t)$ satisfies

$$\mathrm{E}(n_t) = 1 + pt.$$

To get a handle on the actual value of n_t, we use the binomial concentration inequality as described in Theorem 2.4. Namely,

$$\Pr(|n_t - \mathrm{E}(n_t)| > a) \leq e^{-a^2/(2pt+2a/3)}.$$

Thus, n_t is exponentially concentrated around $\mathrm{E}(n_t)$.

We are interested in the degree distribution of a graph generated by $G(p)$. Let $m_{k,t}$ denote the number of vertices of degree k at time t. First we note that

$$m_{1,0} = 1, \text{ and } m_{0,k} = 0.$$

We wish to derive the recurrence for the expected value $\mathrm{E}(m_{k,t})$. Note that a vertex of degree k at time t could have come from two cases, either it was a vertex of degree k at time $t-1$ and had no edge added to it, or it was a vertex of degree $k-1$ at time $t-1$ and the new edge was put in incident to it. Let $\mathcal{F}_t$ be the σ-algebra associated with the probability space at time t. Thus, for $t > 0$ and $k > 1$, we have

$$\begin{aligned} \mathrm{E}(m_{k,t}|\mathcal{F}_{t-1}) &= m_{k,t-1}\big(1 - \frac{kp}{2t} - \frac{(1-p)2k}{2t}\big) \\ &\quad + m_{k-1,t-1}\big(\frac{(k-1)p}{2t} + \frac{(1-p)2(k-1)}{2t}\big) \\ &= m_{k,t-1}\big(1 - \frac{(2-p)k}{2t}\big) + m_{k-1,t-1}\big(\frac{(2-p)(k-1)}{2t}\big). \qquad (3.1) \end{aligned}$$

If we take the expectation on both sides, we get the following recurrence formula.

$$\mathrm{E}(m_{k,t}) = \mathrm{E}(m_{k,t-1})\big(1 - \frac{(2-p)k}{2t}\big) + \mathrm{E}(m_{k-1,t-1})\big(\frac{(2-p)(k-1)}{2t}\big).$$

For $t > 0$ and $k = 1$, we have

$$\mathrm{E}(m_{1,t}|\mathcal{F}_{t-1}) = m_{1,t-1}\big(1 - \frac{(2-p)}{2t}\big) + p. \qquad (3.2)$$

Thus,

$$\mathrm{E}(m_{1,t}) = \mathrm{E}(m_{1,t-1})\big(1 - \frac{(2-p)}{2t}\big) + p.$$

To solve this recurrence, some existing papers made the (unjustified) assumption $\mathrm{E}(m_{k,t}) \approx a_k t$ where a_k is independent of t. The peril of such innocent-looking assumptions will be discussed later in this chapter.

Here we will give a rigorous proof that the expected values $\mathrm{E}(m_{k,t})$ follow a power law when t goes to infinity. To do so, we invoke Lemma 3.1 (to be proved in the next section) which asserts that for a sequence $\{a_t\}$ satisfying the recursive relation $a_{t+1} = (1 - \frac{b_t}{t})a_t + c_t$, the limit $\lim_{t\to\infty} \frac{a_t}{t}$ exists and

$$\lim_{t\to\infty} \frac{a_t}{t} = \frac{c}{1+b}$$

provided that $\lim_{t\to\infty} b_t = b > 0$ and $\lim_{t\to\infty} c_t = c$.

We proceed by induction on k to show that $\lim_{t\to\infty} \mathrm{E}(m_{k,t})/t$ has a limit M_k for each k.

The first case is $k = 1$. In this case, we apply Lemma 3.1 with $b_t = b = (2-p)/2$ and $c_t = c = p$ to deduce that $\lim_{t\to\infty} \mathrm{E}(m_{1,t})/t$ exists and

$$M_1 = \lim_{t\to\infty} \frac{\mathrm{E}(m_{1,t})}{t} = \frac{2p}{4-p}.$$

Now we assume that $\lim_{t\to\infty} \mathrm{E}(m_{k-1,t})/t$ exists and we apply the lemma again with $b_t = b = k(2-p)/2$ and $c_t = \mathrm{E}(m_{k-1,t-1})(2-p)(k-1)/(2t)$, so in this case $c = M_{k-1}(2-p)(k-1)/2$. Lemma 3.1 implies that the limit $\lim_{t\to\infty} \mathrm{E}(m_{k,t})/t$ exists and is equal to

$$M_k = M_{k-1}\frac{(2-p)(k-1)}{2+k(2-p)} = M_{k-1}\frac{k-1}{k+\frac{2}{2-p}}. \qquad (3.3)$$

Thus we can write

$$M_k = \frac{2p}{4-p}\prod_{j=2}^{k}\frac{j-1}{j+\frac{2}{2-p}} = \frac{2p}{4-p}\frac{\Gamma(k)\Gamma(2+\frac{2}{2-p})}{\Gamma(k+1+\frac{2}{2-p})} \tag{3.4}$$

where $\Gamma(k)$ is the Gamma function.

We wish to show that the graph G generated by $G(p)$ is a power law graph with $M_k \propto k^{-\beta}$ (where $\propto$ means "is proportional to") for large k. If $M_k \propto k^{-\beta}$, then

$$\frac{M_k}{M_{k-1}} = \frac{k^{-\beta}}{(k-1)^{-\beta}} = (1-\frac{1}{k})^{\beta} = 1 - \frac{\beta}{k} + O(\frac{1}{k^2}).$$

From (3.3) we have

$$\frac{M_k}{M_{k-1}} = \frac{k-1}{k+\frac{2}{2-p}} = 1 - \frac{1+\frac{2}{2-p}}{k+\frac{2}{2-p}} = 1 - \frac{1+\frac{2}{2-p}}{k} + O(\frac{1}{k^2})$$

Thus the exponent β of the power-law graph satisfies

$$\beta = 1 + \frac{2}{2-p} = 2 + \frac{p}{2-p}.$$

Since p is between 0 and 1, the range for β is $2 \le \beta \le 3$ as illustrated in Figure 1.

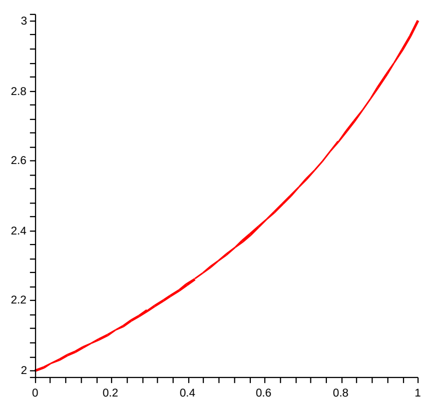

FIGURE 1. The exponent $\beta = 2 + \frac{p}{2-p}$ falls into the range $[2, 3]$.

The equation for M_k in (3.4) can be expressed by using the Beta function:

$$\begin{aligned} B(a,b) &= \int_0^1 x^{a-1}(1-x)^{b-1}dx \\ &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \end{aligned}$$

Therefore M_k satisfies

$$\begin{aligned} M_k &= \frac{2p}{4-p}\frac{\Gamma(k)\Gamma(2+\frac{2}{2-p})}{\Gamma(k+1+\frac{2}{2-p})} \\ &= \frac{p(\beta-1)}{\beta}\frac{\Gamma(k)\Gamma(1+\beta)}{\Gamma(k+\beta)} \\ &= p(\beta-1)\frac{\Gamma(k)\Gamma(\beta)}{\Gamma(k+\beta)} \\ &= p(\beta-1)\int_0^1 x^{k-1}(1-x)^{\beta-1}dx \\ &= p(\beta-1)B(k,\beta). \end{aligned}$$

Another consequence of the above derivation for M_k is the following equality:

$$\sum_{k=1}^{\infty}\frac{\Gamma(k)}{\Gamma(k+\beta)} = \frac{1}{\Gamma(\beta)(\beta-1)}. \tag{3.5}$$

One way to prove (3.5) is to use the fact that the expected number of vertices is $1+pt$. Since $\sum_{k=1}^{\infty} M_k = p$, the equation (3.5) immediately follows.

An alternative way to directly prove (3.5) is the following:

$$\begin{aligned} \sum_{k=1}^{\infty}\frac{\Gamma(k)}{\Gamma(k+\beta)} &= \frac{1}{\Gamma(\beta)}\sum_{k=1}^{\infty}\frac{\Gamma(k)\Gamma(\beta)}{\Gamma(k+\beta)} \\ &= \frac{1}{\Gamma(\beta)}\sum_{k=1}^{\infty}B(k,\beta) \\ &= \frac{1}{\Gamma(\beta)}\sum_{k=1}^{\infty}\int_0^1 x^{k-1}(1-x)^{\beta-1}dx \\ &= \frac{1}{\Gamma(\beta)}\int_0^1\sum_{k=1}^{\infty}x^{k-1}(1-x)^{\beta-1}dx \\ &= \frac{1}{\Gamma(\beta)}\int_0^1(1-x)^{\beta-2}dx \\ &= \frac{1}{\Gamma(\beta)(\beta-1)}. \end{aligned}$$

Equation 3.5 is proved.

3.3. A useful lemma for rigorous proofs

LEMMA 3.1. *Suppose that a sequence* $\{a_t\}$ *satisfies the recurrence relation*

$$a_{t+1} = (1-\frac{b_t}{t+t_1})a_t + c_t \text{ for } t \geq t_0.$$

Furthermore, suppose $\lim_{t\to\infty} b_t = b > 0$ *and* $\lim_{t\to\infty} c_t = c$. *Then* $\lim_{t\to\infty} \frac{a_t}{t}$ *exists and*

$$\lim_{t\to\infty} \frac{a_t}{t} = \frac{c}{1+b}.$$

Proof. Without loss of generality, we can assume $t_1 = 0$ after shifting t by t_1.

By rearranging the recurrence relation, we have

$$\begin{aligned}
\frac{a_{t+1}}{t+1} - \frac{c}{1+b} &= \frac{(1-\frac{b_t}{t})a_t + c_t}{t+1} - \frac{c}{1+b} \\
&= (\frac{a_t}{t} - \frac{c}{1+b})(\frac{t}{t+1})(1-\frac{b_t}{t}) + \frac{t}{t+1}(1-\frac{b_t}{t})(\frac{c}{1+b}) \\
&\quad + \frac{c_t}{t+1} - \frac{c}{1+b} \\
&= (\frac{a_t}{t} - \frac{c}{1+b})(1-\frac{1+b_t}{t+1}) + \frac{c_t}{t+1} - \frac{(1+b_t)c}{(t+1)(1+b)} \\
&= (\frac{a_t}{t} - \frac{c}{1+b})(1-\frac{1+b_t}{t+1}) + \frac{(1+b)c_t - (1+b_t)c}{(1+b)(1+t)}.
\end{aligned}$$

Letting $s_t = |\frac{a_t}{t} - \frac{c}{1+b}|$, the triangle inequality now gives :

$$s_{t+1} \le s_t|1 - \frac{1+b_t}{t+1}| + |\frac{(1+b)c_t - (1+b_t)c}{(1+b)(1+t)}|.$$

Using the fact that $\lim_{t\to\infty} b_t = b$ and $\lim_{t\to\infty} c_t = c$, we have

$$|(1+b)c_t - (1+b_t)c| < \epsilon$$

for any fixed $\epsilon > 0$ provided t is sufficiently large. So, for some T, we have $b_t > b/2$ if $t \ge T$. Thus,

$$s_{t+1} - \epsilon < (s_t - \epsilon)(1 - \frac{1+b/2}{t}).$$

Since $b > 0$, it is not difficult to show that $\prod(1-(1+b/2)/t)$ goes to 0 as $t \to \infty$. Repeated application of the above inequality gives $s_t < 2\epsilon$ for large t. Since ϵ can be chosen arbitrarily, we have $s_t \to 0$ as t goes to infinity, as desired. Therefore we have proved that

$$\lim_{t\to\infty} \frac{a_t}{t} = \frac{c}{1+b}.$$

□

3.4. The peril of heuristics via an example of balls-and-bins

Here we give an example of an incorrect deduction of the power law. This example of a balls-and-bins problem is a generalized version of Polya's urn problem and is quite interesting in its own.

The classical problem of Polya's urns has the following setup:

Start with a fixed number of bins each with one ball. At each tick of the clock, a new ball is placed in one of the bins with the probability of choosing the ith bin proportional to the number of balls in the ith bin.

Here we consider the balls-and-bins process when the number of bins is not fixed. We have two parameters, p, a probability between 0 and 1 and a real number r. We call this model Polya(p, r).

Imagine we have a stream of balls arriving one at a time.
At the very beginning, we place the first ball in a bin.
At time t, with probability p, we place the new ball in a new bin.
Otherwise, we place the new ball in an existing bin, where we choose a bin with probability proportional to the rth power of the number of the balls in that bin.

We can modify Polya(p, r) into the following model, denoted by Polya$^*(p, r)$:

We have a stream of balls arriving two at a time.
At the very beginning, we place the first set of two balls in a bin.
At time t, with probability p, we place one new ball in a new bin and the other ball in an existing bin with probability proportional to the to the rth power of the number of balls in that bin.
Otherwise, we place each of the two new balls in an existing bin with probability proportional to the rth power of the number of balls in that bin.

For the case of $r = 1$, the model Polya$^*(p, 1)$ is the preferential attachment model in Section 3.1 if we view the bins as vertices and edges connect the bins the two balls that arrive at the same time go into. The model Polya$^*(p, r)$ is regarded as a preferential attachment with *feedback*. When $r > 1$, it is preferential attachment with *positive* feedback. When $r < 1$, it is preferential attachment with *negative* feedback. This general form of preferential attachment has been examined in a number of papers [**32, 48, 49, 87, 107**]. For example, it was shown that for $r > 1$, a single bin dominates. In fact, for any $k > r/(r-1)$, with high probability only finitely many bins ever reach size k.

In the remainder of this section, we will give a "proof" that for $r > 1$ in Polya(p, r), the bin sizes have a power law distribution. The exercise here is to find what is wrong in this "proof"!

Let $n_k(t)$ be the number of bins at time t with k balls. Note that

$$\mathrm{E}(n_k(t+1)) = \mathrm{E}(n_k(t)(p + (1-p)(1 - \frac{k^r}{w_t}))) + (1-p)\mathrm{E}(\frac{n_{k-1}(t)(k-1)^r}{w_t})$$

where w_t denotes $\sum_i n_i(t) i^r$. Let us assume that as t gets large $\mathrm{E}(n_k(t))$ converges to a fixed fractions of the total number of balls. In other words, $n_k(t) \approx a_k t$. *(A very dangerous assumption indeed!)* Furthermore, assume w_t converges to wt for some constant w. By plugging those assumptions in the above equation, we get

$$a_k(t+1) = a_k t(p + (1-p)(1 - \frac{k^r}{wt})) + (1-p)a_{k-1}t\frac{(k-1)^r}{wt}.$$

This implies

$$\begin{aligned} \frac{a_k}{a_{k-1}} &= \frac{(1-p)(k-1)^r}{w + (1-p)k^r} \\ &= \frac{(k-1)^r}{\frac{w}{1-p} + k^r} \\ &\approx \left(\frac{k-1}{k}\right)^r \end{aligned}$$

for k large. Thus, one might be inclined to conclude that the bin size distribution is a power law distribution with exponent r if $r > 1$!

However, the truth (see [**32**]) is that all but one of the a_i's are zero. A quick simulation will show that almost all balls go into one bin. In fact, it can be shown with high probability that all balls go into one bin with the exceptions of the balls in bins of size 1 and finitely many other balls. This model gives an explanation for the forming of a *monopoly*.

What went wrong in the above "proof"? The power law distribution is a consequence of an unfortunate ratio 0/0. That is exactly why rigorous mathematics is needed here.

3.5. Scale-free networks

Quite a few recent papers use the term "scale-free networks" to mean graphs with a power law degree distribution. However, power law and scale-free are very different concepts. In fact, the term "scale-free" has rarely been properly defined.

Here we intend to clarify the distinction of the two. To discuss "scale-free", first we have to answer the question concerning "scale". What is the appropriate scale or scales? How should "scale-free" be defined in a natural way?

Two types of scale come to mind — *space* and *time*. In fact, scales of space and time can coexist simultaneously. For example, the call graphs have very similar shape (the same exponent in the power law distribution) while sampling at different geographical locations and at different sampling intervals. To simplify the issues, we separately discuss "scale-free in space" and "scale-free in time".

3.5.1. Scale-free in space. "Self-similarity" is one of the visible traits that exist in numerous networks. By comparing the web crawls of [**11, 13**] and [**24, 88**] we see that the same power law appears to govern various subgraphs of the web as well as the whole. However, while some subgraphs obey the same power law and appear to be self-similar (i.e., similar to the entire graph), there exist subgraphs of the web which would not obey the power law (e.g., the subgraph defined by all nodes with outdegree 50). So, for what kind of subgraphs can "self-similarity" be considered or even formally defined?

For the family of recursive trees [**94**] as rooted trees, the definition comes naturally. The special subtrees consisting of all descendants of a vertex are similar to the whole tree.

For a general graph, additional information will be needed to help define the special subgraphs for which self-similarity will hold. One direction is to consider a geometric embedding of the graph into some specified metric space. Then we use the metric to define the special subgraphs. Another direction is to take the graph as given but to extract a so-called "local graph" from it. The graph metric of the local graph provides the geometry of the graph. In Chapter 12, we will define the local graphs and discuss this idea further.

3.5.2. Scale-free in time. It is easier to define scale-free in terms of time than space perhaps because time is one-dimensional but space is multi-dimensional. The generative model is a process of growing graphs by adding nodes and edges one at a time. One way is to divide the time into almost equal units and combine all nodes born in the same unit time into one super-node. The bigger time unit one chooses, the fewer nodes the resulting graph has. We say a model is *scale-free* if it generates power law graphs with the same exponent regardless of the choice of time scale. In other words, a generative model is invariant with respect to time in the sense that if we change the time scale by any given factor, then the original graph and the scaled graph should satisfy the power law with the same exponent for the degrees.

We can modify the previous model by adding an additional integer parameter m. Here are the two generalized steps:

- *Vertex-m-step* — Add a new vertex v, and m new edges $\{u_i, v\}$, $i = 1, \ldots, m$, by choosing u_i with probability proportional to the degree of u in the current graph.
- *Edge-m-step* — Add m new edges $\{r_i, s_i\}$, $i = 1, \ldots, m$, by choosing vertices r_i with probability proportional to the degree of r_i, and by choosing vertices s_i with probability proportional to the degree of s_i.

Now we define a graph $G(p, m, G_0)$:

> Begin with the initial graph G_0.
> For $t > 0$, at time t,
> with probability p, take a vertex-m-step,
> otherwise, take an edge-m-step.

If G_0 is taken to be the graph consisting of a vertex with m loops, we write $G(p, m) = G(p, m, G_0)$.

In this model every vertex has degree at least m. Let $m_{k,t}$ be the number of vertices with degree k at time t. At time t, G_t has exactly $e_0 + mt$ edges. We will denote this by e_t. Let $\mathcal{F}_t$ be the σ-algebra generated by the probability space at time t. Thus, for $t > 0$ and $k > m$, we have

$$\begin{aligned} \mathrm{E}(m_{k,t}|\mathcal{F}_{t-1}) &= m_{k,t-1}\big(1 - \frac{kmp}{2e_{t-1}} - \frac{m(1-p)2k}{2e_{t-1}}\big) \\ &\quad + m_{k-1,t-1}\big(\frac{(k-1)mp}{2e_{t-1}} + \frac{(1-p)2m(k-1)}{2e_{t-1}}\big) + O(\frac{1}{t^2}) \\ &= m_{k,t-1}\big(1 - \frac{(2-p)mk}{2e_{t-1}}\big) + m_{k-1,t-1}\big(\frac{(2-p)m(k-1)}{2e_{t-1}}\big) + O(\frac{1}{t^2}). \qquad (3.6) \end{aligned}$$

Note that the $O(1/t^2)$ term above makes it possible to absorb the error terms caused by loops or multiple edges. Now by taking the expectation on both sides, we get the following recurrence formula.

$$\mathrm{E}(m_{k,t}) = \mathrm{E}(m_{k,t-1})(1 - \frac{(2-p)mk}{2e_{t-1}}) + \mathrm{E}(m_{k-1,t-1})(\frac{(2-p)m(k-1)}{2e_{t-1}}) + O(\frac{1}{t^2}).$$

In the random graph model $G(p,m)$, we have $e_t = m(t+1)$. If we substitute e_t in the above inequality, all appearances of m are cancelled out. Indeed, we get exactly the same recurrence formula as we previously had for $G(p)$ in (3.1). Therefore, graphs generated by $G(p,m)$ have the same power law distribution as graphs generated by $G(p)$. So we see the exponent β is independent of the scale unit m.

If we compare the figures of the degree distributions of $G(p)$ and $G(p,m)$ in their logarithmic representation, the figures are almost identical in the sense that the shape of the curves are straight lines of the same slope. The only difference is that the line associated with $G(p,m)$ is a slight linear translation to the right. Mainly, the density of $G(p,m)$ differs from that of $G(p)$ by a factor of m. In the logarithmic representation, the difference is an additive term of $\log m$, which is rather small in comparison with n, the number of nodes. Nevertheless, the main characteristic of the power law is the exponent of the power law as seen from the same slope in both figures.

3.6. The sharp concentration of preferential attachment scheme

In Section 3.2 we considered the expected degrees for graphs generated by the preferential attachment scheme and we derived the power law distribution for the expected degree sequence. However, the expected degrees can be quite different from the actual degrees of a random graph in hand. Can we give a (probabilistic) estimate of the difference? The goal of the section is to answer this question.

Since the preferential attachment scheme is an on-line model, the concentration bound that we intend to give involves nontrivial arguments and is somewhat lengthy.

We will prove the following theorem.

THEOREM 3.2. *For the preferential attachment model $G(p)$, almost surely the number of vertices with degree k at time t is*

$$M_k t + O(2\sqrt{k^3 t \ln(t)}).$$

Recall $M_1 = \frac{2p}{4-p}$ *and* $M_k = \frac{2p}{4-p}\frac{\Gamma(k)\Gamma(1+\frac{2}{2-p})}{\Gamma(k+1+\frac{2}{2-p})} = O(k^{-(2+\frac{p}{2-p})})$, *for* $k \geq 2$. *In other words, almost surely the graphs generated by* $G(p)$ *have the power law degree distribution with the exponent* $\beta = 2 + \frac{p}{2-p}$.

Proof. We have shown that

$$\lim_{t\to\infty} \frac{\mathrm{E}(m_{k,t})}{t} = M_k,$$

where M_k is defined recursively in (3.3). It is sufficient to show $m_{k,t}$ is concentrated on the expected value.

We shall prove the following claim.

Claim: For any fixed $k \geq 1$ and $c > 0$, with probability at least $1 - 2(t+1)^{k-1}e^{-c^2}$, we have

$$|m_{k,t} - M_k(t+1)| \leq 2kc\sqrt{t}.$$

To see that the claim implies Theorem 3.2, we choose $c = \sqrt{k \ln t}$. Note that

$$2(t+1)^{k-1}e^{-c^2} = 2(t+1)^{k-1}t^{-k} = o(1).$$

So, from the claim, with probability $1 - o(1)$, we have

$$|m_{k,t} - M_k(t+1)| \leq 2\sqrt{k^3 t \ln t},$$

as desired.

It remains to prove the claim.

Proof of claim: We shall prove it by induction on k.

The base case of $k = 1$*:*

For $k = 1$, from equation (3.2), we have

$$\begin{aligned}
\mathrm{E}(m_{1,t} - M_1(t+1)|\mathcal{F}_{t-1}) &= \mathrm{E}(m_{1,t}|\mathcal{F}_{t-1}) - M_1(t+1) \\
&= m_{1,t-1}(1 - \frac{2-p}{2t}) + p - M_1 t - M_1 \\
&= (m_{1,t-1} - M_1 t)(1 - \frac{2-p}{2t}) + p - M_1\frac{2-p}{2} - M_1 \\
&= (m_{1,t-1} - M_1 t)(1 - \frac{2-p}{2t})
\end{aligned}$$

since $M_1 = \frac{2p}{4-p}$ and $p - M_1\frac{2-p}{2} - M_1 = 0$.

Let $X_{1,t} = \frac{m_{1,t} - M_1(t+1)}{\prod_{j=1}^{t}(1 - \frac{2-p}{2j})}$. We consider the martingale $1 = X_{1,0}, X_{1,1}, \ldots, X_{1,t}$.

We have

$$\begin{aligned}
X_{1,t} - X_{1,t-1} &= \frac{m_{1,t} - M_1(t+1)}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})} - \frac{m_{1,t-1} - M_1 t}{\prod_{j=1}^{t-1}(1-\frac{2-p}{2j})} \\
&= \frac{1}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})}[(m_{1,t} - M_1(t+1)) - (m_{1,t-1} - M_1 t)(1 - \frac{2-p}{2t})] \\
&= \frac{1}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})}[(m_{1,t} - m_{1,t-1}) + \frac{2-p}{2t}(m_{1,t-1} - M_1 t) - M_1].
\end{aligned}$$

We note that $|m_{1,t} - m_{1,t-1}| \leq 2$, $m_{1,t-1} \leq t$, and $M_1 = \frac{2p}{4-p} < 1$. We have

$$|X_{1,t} - X_{1,t-1}| \leq \frac{4}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})}. \tag{3.7}$$

Since $|m_{1,t} - m_{1,t-1}| \leq 2$, we have

$$\begin{aligned}
\mathrm{Var}(m_{1,t}|\mathcal{F}_{t-1}) &\leq \mathrm{E}((m_{1,t} - m_{1,t-1})^2|\mathcal{F}_{t-1}) \\
&\leq 4.
\end{aligned}$$

Therefore, we have the following upper bound for $\mathrm{Var}(X_{1,t}|\mathcal{F}_{t-1})$.

$$\begin{aligned}
\mathrm{Var}(X_{1,t}|\mathcal{F}_{t-1}) &= \mathrm{Var}\big((m_{1,t} - M_1(t+1))\frac{1}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})}\big|\mathcal{F}_{t-1}\big) \\
&= \frac{1}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})^2}\mathrm{Var}(m_{1,t} - M_1(t+1)|\mathcal{F}_{t-1}) \\
&= \frac{1}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})^2}\mathrm{Var}(m_{1,t}|\mathcal{F}_{t-1}) \\
&\leq \frac{4}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})^2}. \qquad (3.8)
\end{aligned}$$

We apply Theorem 2.22 to the martingale $\{X_{1,t}\}$ with $\sigma_i^2 = \frac{4}{\prod_{j=1}^{i}(1-\frac{2-p}{2j})^2}$, $M = \frac{4}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})}$ and $a_i = 0$. We have

$$\Pr(X_{1,t} \geq \mathrm{E}(X_{1,t}) + \lambda) \leq e^{-\frac{\lambda^2}{2(\sum_{i=1}^{t}\sigma_i^2 + M\lambda/3)}}.$$

Here $\mathrm{E}(X_{1,t}) = X_{1,0} = 1$. We will use the following approximation.

$$\begin{aligned}
\prod_{j=1}^{i}(1-\frac{2-p}{2j}) &= \frac{\Gamma(i+\frac{p}{2})}{\Gamma(i+1)\Gamma(\frac{p}{2})} \\
&= \big(\frac{1}{\Gamma(\frac{p}{2})} + O\big(\frac{1}{i}\big)\big)i^{-1+p/2}.
\end{aligned}$$

For any $c > 0$, we choose $\lambda = \frac{2c\sqrt{t}}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})} \approx 2\Gamma(\frac{p}{2})ct^{(3-p)/2}$. We have

$$\begin{aligned}\sum_{i=1}^{t}\sigma_i^2 &= \sum_{i=1}^{t}\frac{4}{\prod_{j=1}^{i}(1-\frac{2-p}{2j})^2}\\ &\approx \sum_{i=1}^{t}4\Gamma^2(\frac{p}{2})i^{2-p}\\ &\approx \frac{4\Gamma^2(\frac{p}{2})}{3-p}t^{3-p}\\ &< 2\Gamma^2(\frac{p}{2})t^{3-p}.\end{aligned}$$

We note that

$$M\lambda/3 \approx \frac{4}{3}\Gamma^2(\frac{p}{2})ct^{5/2-p} < 2\Gamma^2(\frac{p}{2})t^{3-p}$$

provided $c < \sqrt{t}$. We have

$$\begin{aligned}\Pr(X_{1,t} \geq 1+\lambda) &\leq e^{-\frac{\lambda^2}{2(\sum_{i=1}^{t}\sigma_i^2+M\lambda/3)}}\\ &< e^{-\frac{4\Gamma^2(\frac{p}{2})c^2t^{3-p}}{(4+o(1))\Gamma^2(\frac{p}{2})t^{3-p}}}\\ &\approx e^{-c^2}.\end{aligned}$$

Since 1 is much smaller than λ, we can replace $1+\lambda$ by λ without loss of generality. Thus, with probability at least $1-e^{-c^2}$, we have

$$X_{1,t} \leq \lambda.$$

Similarly, with probability at least $1-e^{-c^2}$, we have

$$m_{1,t} - M_1(t+1) \leq 2c\sqrt{t}. \tag{3.9}$$

We remark that the inequality (3.9) holds for any $c > 0$. In fact, it is trivial when $c > \sqrt{t}$ since $|m_{1,t} - M_1(t+1)| \leq 2t$ always holds.

Similarly, by applying Theorem 2.26 to the martingale, the following lower bound

$$m_{1,t} - M_1(t+1) \geq -2c\sqrt{t}$$

holds with probability at least $1-e^{-c^2}$.

We have proved the claim for $k = 1$.

The inductive step:

Suppose the claim holds for $k-1$. For k, we define

$$X_{k,t} = \frac{m_{k,t} - M_k(t+1) - 2(k-1)c\sqrt{t}}{\prod_{j=1}^{t}(1-\frac{(2-p)k}{2j})}.$$

we have

$$\begin{aligned}
&\mathrm{E}(m_{k,t} - M_k(t+1) - 2(k-1)c\sqrt{t}|\mathcal{F}_{t-1}) \\
&\quad = \mathrm{E}(m_{k,t}|\mathcal{F}_{t-1}) - M_k(t+1) - 2(k-1)c\sqrt{t} \\
&\quad = m_{k,t-1}\big(1 - \frac{(2-p)k}{2t}\big) + m_{k-1,t-1}\big(\frac{(2-p)(k-1)}{2t}\big) \\
&\qquad -M_k(t+1) - 2(k-1)c\sqrt{t}.
\end{aligned}$$

By the induction hypothesis, with probability at least $1 - 2t^{k-2}e^{-c^2}$, we have

$$|m_{k-1,t-1} - M_{k-1}t| \le 2(k-1)c\sqrt{t-1}.$$

By using this estimate, with probability at least $1 - 2t^{k-2}e^{-c^2}$, we have

$$\mathrm{E}(m_{k,t}-M_k(t+1)-2(k-1)c\sqrt{t}|\mathcal{F}_{t-1}) \le (1-\frac{(2-p)k}{2t})(m_{k,t-1}-M_k t-2(k-1)c\sqrt{t-1})$$

by using the fact that $M_k \le M_{k-1}$ as seen in (3.3).

Therefore, $0 = X_{k,0}, X_{k,1}, \ldots, X_{k,t}$ forms a submartingale with failure probability at most $2t^{k-2}e^{-c^2}$.

Similar to inequalities (3.7) and (3.8), it can be easily shown that

$$|X_{k,t} - X_{k,t-1}| \le \frac{4}{\prod_{j=1}^{t}(1 - \frac{(2-p)k}{2j})}$$

and

$$\mathrm{Var}(X_{k,t}|\mathcal{F}_{t-1}) \quad \le \quad \frac{4}{\prod_{j=1}^{t}(1 - \frac{(2-p)k}{2j})^2}.$$

We apply Theorem 2.39 on the submartingale with $\sigma_i^2 = \frac{4}{\prod_{j=1}^{i}(1-\frac{(2-p)k}{2j})^2}$, $M = \frac{4}{\prod_{j=1}^{t}(1-\frac{2-p}{2j})}$ and $a_i = 0$. We have

$$\Pr(X_{k,t} \ge \mathrm{E}(X_{k,t}) + \lambda) \le e^{-\frac{\lambda^2}{2(\sum_{i=1}^{t}\sigma_i^2 + M\lambda/3)}} + \Pr(B),$$

where $\Pr(B) \le t^{k-1}e^{-c^2}$ by the induction hypothesis.

Here $\mathrm{E}(X_{k,t}) = X_{k,0} = 0$. We will use the following approximation.

$$\begin{aligned}
\prod_{j=1}^{i}(1 - \frac{(2-p)k}{2j}) &= \frac{\Gamma(i+1-\frac{(2-p)k}{2})}{\Gamma(i+1)\Gamma(1-\frac{(2-p)k}{2})} \\
&= \big(\frac{1}{\Gamma(1-\frac{(2-p)k}{2})} + O\big(\frac{1}{i}\big)\big)i^{-k(2-p)/2}.
\end{aligned}$$

For any $c > 0$, we choose $\lambda = \frac{2c\sqrt{t}}{\prod_{j=1}^{t}(1-\frac{(2-p)k}{2j})} \approx 2\Gamma(1-\frac{(2-p)k}{2})ct^{1/2+k(2-p)/2}$. We have

$$\begin{aligned}\sum_{i=1}^{t}\sigma_i^2 &\leq \sum_{i=1}^{t}\frac{4}{\prod_{j=1}^{i}(1-\frac{(2-p)k}{2j})^2}\\ &\approx \sum_{i=1}^{t}4\Gamma^2(1-\frac{(2-p)k}{2})i^{k(2-p)}\\ &\approx \frac{4\Gamma^2(1-\frac{(2-p)k}{2})}{1+(2-p)k}t^{1+k(2-p)}\\ &< 2\Gamma^2(1-\frac{(2-p)k}{2})t^{1+k(2-p)}.\end{aligned}$$

We note that

$$M\lambda/3 \approx \frac{4}{3}\Gamma^2(1-\frac{(2-p)k}{2})ct^{\frac{1}{2}+(2-p)k} < 2\Gamma^2(1-\frac{(2-p)k}{2})t^{1+(2-p)k}$$

as long as $c < \sqrt{t}$. We have

$$\begin{aligned}\Pr(X_{k,t}\geq\lambda) &\leq e^{-\frac{\lambda^2}{2(\sum_{i=1}^{t}\sigma_i^2+M\lambda/3)}}+\Pr(B)\\ &< e^{-\frac{4\Gamma^2(1-\frac{(2-p)k}{2})c^2t^{1+(2-p)k}}{(4+o(1))\Gamma^2(1-\frac{(2-p)k}{2})t^{1+(2-p)k}}}+\Pr(B)\\ &< e^{-c^2}+t^{k-1}e^{-c^2}\\ &\leq (t+1)^{k-1}e^{-c^2}.\end{aligned}$$

With probability at least $1-(t+1)^{k-1}e^{-c^2}$, we have

$$X_{k,t}\leq\lambda.$$

Equivalently, with probability at least $1-(t+1)^{k-1}e^{-c^2}$, we have

$$m_{k,t}-M_k(t+1)\leq 2kc\sqrt{t}. \tag{3.10}$$

We remark that the inequality (3.10) holds for any $c > 0$. In fact, it is trivial when $c > \sqrt{t}$ since $|m_{k,t}-M_k(t+1)| \leq 2kt$ always holds.

To obtain the lower bound, we consider

$$X'_{k,t} = \frac{m_{k,t}-M_k(t+1)+2(k-1)c\sqrt{t}}{\prod_{j=1}^{t}(1-\frac{(2-p)k}{2j})}.$$

It can be easily shown that $X'_{k,t}$ is nearly a supermartingale. Similarly, by applying Theorem 2.42 to $X'_{k,t}$, the following lower bound

$$m_{k,t}-M_k(t+1)\geq -2kc\sqrt{t}$$

holds with probability at least $1-(t+1)^{k-1}e^{-c^2}$.

This completes the induction. The proof of Theorem 3.2 is complete. □

For completeness, we here state the corresponding theorem for $G(p,m,G_0)$.

THEOREM 3.3. *For the preferential attachment model $G(p, m, G_0)$, almost surely the number of vertices with degree k at time t is*

$$M_k t + m_{k,0} + O(2m\sqrt{(k+m-1)^3 t \ln(t)}).$$

Recall $M_m = \frac{2p}{4-p}$ *and* $M_k = \frac{2p}{4-p}\frac{\Gamma(k)\Gamma(1+\frac{2}{2-p})}{\Gamma(k+1+\frac{2}{2-p})} = O(k^{-(2+\frac{p}{2-p})})$, *for* $k \geq m+1$. *In other words, almost surely the graphs generated by* $G(p, m, G_0)$ *have the power law degree distribution with the exponent* $\beta = 2 + \frac{p}{2-p}$.

3.7. Models for directed graphs

Many real-world graphs are directed graphs. For example, the WWW-graph has edges each of which represents a link from one webpage to another. There are vertices with large indegrees but relatively small outdegrees such as Yahoo, CNN or USA Today. Such vertices are often called *authorities* [**84**]. There are also vertices, called *hubs*, with large outdegrees but relatively small indegrees. For directed graphs, we can have quite different distributions for indegrees and outdegrees. For example, the indegree sequence of the WWW graph follows the power law distribution with the exponent β about 2.1 while the outdegree sequence follows a different power law with exponent β about 2.7.

In this section, we will consider a preferential attachment model that can generate a directed graph with power law indegree distributon and power law outdegree distribution. Furthermore, the exponents for the power law distributions can be specified to be different values.

To generate such a directed graph, we have three parameters for the preferential attachment model:

- Two given probabilities p_1, p_2, satisfying $0 \leq p_1, p_2 \leq p_1 + p_2 \leq 1$.
- An initial graph G_0 at time 0.

We also have three operations:

- *Source-vertex-step* — Add a new vertex v, and add a directed edge (v, u) from v by randomly and independently choosing u in proportion to the indegree of u in the current graph.
- *Sink-vertex-step* — Add a new vertex v, and add a directed edge (u, v) to v by randomly and independently choosing u in proportion to the outdegree of u in the current graph.
- *Edge-step* — Add a new edge (r, s) by independently choosing vertices r and s with probability proportional to its outdegree (respectively indegree).

The random graph model $D_0(p_1, p_2, G_0)$ is defined as follows:

Begin with the initial graph G_0.
For $t > 0$, at time t, the graph G_t is formed by modifying G_{t-1} as follows:
with probability p_1, take a source-vertex-step,

with probability p_2, take a sink-vertex-step,
otherwise, take an edge-step.

This simple model generates a power law graph with different exponents (as functions of p_1 and p_2) for indegree and outdegree distributions. We remark that the vertices with indegree zero (i.e., source vertices) will always have zero indegree. Conversely, the vertices with outdegree zero (i.e., sink vertices) will always have outdegree zero. Except for the vertices in G_0, the rest of the vertices are partitioned into two groups — source vertices and sink vertices. This model might not be feasible for modeling most realistic networks.

We here consider a modified preferential attachment scheme with an additional parameter $\alpha \geq 0$, defined as follows:

α-preferential attachment scheme (or α-scheme, in short):
A vertex u is chosen for the tail (or head) of a new edge with probability proportional to its in-weight (or out-weight). The in-weight of u is defined to be the sum of the indegree of u and α, while the out-weight of u is the sum of the outdegree of u and α.

The random graph model $D(p_1, p_2, \alpha, G_0)$ is defined as follows:

Begin with the initial graph G_0.
For $t > 0$, the graph G_t is formed by modifying G_{t-1} as follows:
with probability p_1, take a source-vertex-step using the α-scheme,
with probability p_2, take a sink-vertex-step using the α-scheme,
otherwise, take an edge-step using the α-scheme.

We note that an alternative model is to add loops to a new vertex in each step. It is not hard to see that adding a loop is equivalent to the 1-preferential attachment scheme. In fact, the α-preferential attachment scheme can be viewed as adding α loops. When G_0 is the graph consisting of a single vertex, we simplify the notation and write $G(p_1, p_2, \alpha) = G(p_1, p_2, \alpha, G_0)$.

The number of edges of $G(p_1, p_2, \alpha)$ at time t is exactly t, while the total weight at time t is $t + \alpha n_t$. The number of vertices n_t at time t follows the binomial distribution. The expected value $\mathrm{E}(n_t)$ satisfies

$$\mathrm{E}(n_t) = 1 + (p_1 + p_2)t.$$

To deal with the actual value n_t, we use the binomial concentration inequality as described in Theorem 2.4. Namely,

$$\Pr(|n_t - \mathrm{E}(n_t)| > a) \leq e^{-a^2/(2pt+2a/3)}.$$

Thus, n_t is exponentially concentrated around $\mathrm{E}(n_t)$.

Let $m_{k,t}^{in}$ denote the number of vertices of in-degree k at time t. We note that

$$m_{0,k}^{in} = 0.$$

We wish to derive a recurrence formula for the expected value $\mathrm{E}(m^{in}_{k,t})$. A vertex of indegree k at time t could have come from two cases, either it was a vertex of degree k at time $t-1$ and had no edge added directed to it, or it was a vertex of indegree $k-1$ at time $t-1$ and the new edge was directed to it.

Let $\mathcal{F}_t$ denote the σ-algebra generated by the probability space at time t. For $t > 0$ and $k > 1$, we have

$$\begin{aligned} \mathrm{E}(m^{in}_{k,t}|\mathcal{F}_{t-1}) &= m^{in}_{k,t-1}(1 - \frac{(k+\alpha)p_1}{t-1+\alpha n_t} - \frac{(1-p_1-p_2)(k+\alpha)}{t-1+\alpha n_t}) \\ &\quad + m^{in}_{k-1,t-1}(\frac{(k-1+\alpha)p_1}{t-1+\alpha n_t} + \frac{(1-p_1-p_2)(k-1+\alpha)}{t-1+\alpha n_t}) \\ &= m^{in}_{k,t-1}(1 - \frac{(1-p_2)(k+\alpha)}{t-1+\alpha n_t}) + m^{in}_{k-1,t-1}(\frac{(1-p_2)(k-1+\alpha)}{t-1+\alpha n_t}). \end{aligned}$$

If we take the expectation on both sides and apply the estimation $n_t \approx (p_1 + p_2)t$, we obtain the following recurrence formula.

$$\mathrm{E}(m^{in}_{k,t}) \approx \mathrm{E}(m^{in}_{k,t-1})(1 - \frac{(1-p_2)(k+\alpha)}{t(1+(p_1+p_2)\alpha)}) + \mathrm{E}(m^{in}_{k-1,t-1})(\frac{(1-p_2)(k-1+\alpha)}{t(1+(p_1+p_2)\alpha}).$$

For $t > 0$ and $k = 0, 1$, we have

$$\begin{aligned} \mathrm{E}(m^{in}_{1,t}|\mathcal{F}_{t-1}) &= m^{in}_{1,t-1}(1 - \frac{(1-p_2)(1+\alpha)}{t-1+\alpha n_t}) + m^{in}_{0,t-1}(\frac{(1-p_2)\alpha}{t-1+\alpha n_t}) + p_2, \\ \mathrm{E}(m^{in}_{0,t}|\mathcal{F}_{t-1}) &= m^{in}_{0,t-1}(1 - \frac{(1-p_2)\alpha}{t-1+\alpha n_t}) + p_1. \end{aligned}$$

Thus,

$$\begin{aligned} \mathrm{E}(m^{in}_{1,t}) &\approx \mathrm{E}(m^{in}_{1,t-1})(1 - \frac{(1-p_2)(1+\alpha)}{t(1+(p_1+p_2)\alpha)}) + \mathrm{E}(m^{in}_{0,t-1})\frac{(1-p_2)\alpha}{t(1+(p_1+p_2)\alpha)} + p_2, \\ \mathrm{E}(m^{in}_{0,t}) &\approx \mathrm{E}(m^{in}_{0,t-1})(1 - \frac{(1-p_2)\alpha}{t(1+(p_1+p_2)\alpha)}) + p_1. \end{aligned}$$

Here these asymptotic equalities come from the fact that $n_t \approx (p_1 + p_2)t$.

We proceed by induction on k to show that $\lim_{t\to\infty} \mathrm{E}(m^{in}_{k,t})/t$ has a limit M^{in}_k for each k.

The first case is $k = 0$. In this case, we apply Lemma 3.1 with $b_t = b = (1-p_2)\alpha/(1+(p_1+p_2)\alpha)$ and $c_t = c = p_2$ to deduce that $\lim_{t\to\infty} \mathrm{E}(m^{in}_{0,t})/t = M^{in}_0$ exists. We have

$$\begin{aligned} M^{in}_0 &= \frac{c}{1+b} \\ &= \frac{p_2}{1 + \frac{(1-p_2)\alpha}{(1+(p_1+p_2)\alpha)}} \\ &= \frac{p_2(1+(p_1+p_2)\alpha)}{1+(1+p_1)\alpha}. \end{aligned} \tag{3.11}$$

For $k = 1$, we use Lemma 3.1 with $b_t = b = (1 - p_2)(1 + \alpha)/(1 + (p_1 + p_2)\alpha))$ and $c_t = \mathrm{E}(m_{0,t-1}^{in})\frac{(1-p_2)\alpha}{t(1+(p_1+p_2)\alpha)} + p_1$. We have

$$c = \lim_{t\to\infty} c_t = M_0^{in}\frac{(1 - p_2)\alpha}{1 + (p_1 + p_2)\alpha} + p_1.$$

It implies that $\lim_{t\to\infty} \mathrm{E}(m_{0,t}^{in})/t = M_1^{in}$ exists. We have

$$\begin{aligned} M_1^{in} &= \frac{c}{1+b} \\ &= \frac{M_0^{in}\frac{(1-p_2)\alpha}{1+(1+p_1)\alpha} + p_1}{1 + \frac{(1-p_2)(1+\alpha)}{(1+(p_1+p_2)\alpha)}} \\ &= \frac{p_1 + (p_1 + p_2 + p_1^2 - p_2^2)\alpha}{2 - p_2 + (1 + p_1)\alpha}. \end{aligned} \tag{3.12}$$

For $k > 1$, we assume that $\lim_{t\to\infty} \mathrm{E}(m_{k-1,t}^{in})/t = M_{k-1}^{in}$ exists and we apply the lemma again with $b_t = b = \frac{(1-p_2)(k+\alpha)}{(1+(p_1+p_2)\alpha)}$ and $c_t = \mathrm{E}(m_{k-1,t-1}^{in})\frac{(1-p_2)(k-1+\alpha)}{t(1+(p_1+p_2)\alpha)}$, so $c = M_{k-1}^{in}\frac{(1-p_2)(k-1+\alpha)}{(1+(p_1+p_2)\alpha)}$. Lemma 3.1 implies that the limit $\lim_{t\to\infty} \mathrm{E}(m_{k,t}^{in})/t = M_k^{in}$ exists and is equal to

$$\begin{aligned} M_k^{in} &= \frac{c}{1+b} \\ &= M_{k-1}^{in}\frac{\frac{(1-p_2)(k-1+\alpha)}{1+(1+p_1)\alpha}}{1 + \frac{(1-p_2)(k+\alpha)}{(1+(p_1+p_2)\alpha)}} \\ &= M_{k-1}^{in}\frac{k - 1 + \alpha}{k + \alpha + \frac{1+(p_1+p_2)\alpha}{1-p_2}}. \end{aligned} \tag{3.13}$$

Thus we can write

$$\begin{aligned} M_k^{in} &= M_1^{in}\prod_{j=2}^{k}\frac{j - 1 + \alpha}{j + \alpha + \frac{1+(p_1+p_2)\alpha}{1-p_2}} \\ &= M_1^{in}\frac{\Gamma(k + \alpha)\Gamma(2 + \alpha + \frac{1+(p_1+p_2)\alpha}{1-p_2})}{\Gamma(1 + \alpha)\Gamma(k + 1 + \alpha + \frac{1+(p_1+p_2)\alpha}{1-p_2})} \\ &\approx M_1^{in}\frac{\Gamma(2 + \alpha + \frac{1+(p_1+p_2)\alpha}{1-p_2})}{\Gamma(1 + \alpha)}k^{1+\frac{1+(p_1+p_2)\alpha}{1-p_2}} \end{aligned}$$

where $\Gamma(k)$ is the Gamma function.

Thus we have a power law graph for the indegree sequence with

$$\beta^{in} = 1 + \frac{1 + (p_1 + p_2)\alpha}{1 - p_2} = 2 + \frac{p_2 + (p_1 + p_2)\alpha}{1 - p_2}.$$

Let $m^{out}(t,k)$ be the number of vertices with outdegree k at time t. Similarly we can show $\lim_{t\to\infty} \frac{\mathrm{E}(m^{out}_{t,k})}{t}$ exists. We denote it by M^{out}_k. We have

$$M_0^{out} = \frac{p_2(1+(p_1+p_2)\alpha)}{1+(1+p_2)\alpha}, \tag{3.14}$$

$$M_1^{out} = \frac{p_2+(p_1+p_2+p_2^2-p_1^2)\alpha}{2-p_1+(1+p_2)\alpha}. \tag{3.15}$$

For $k > 1$, we have

$$\begin{aligned} M_k^{out} &= M_1^{out}\frac{\Gamma(k+\alpha)\Gamma(2+\alpha+\frac{1+(p_1+p_2)\alpha}{1-p_1})}{\Gamma(1+\alpha)\Gamma(k+1+\alpha+\frac{1+(p_1+p_2)\alpha}{1-p_1})} \\ &\approx M_1^{out}\frac{\Gamma(2+\alpha+\frac{1+(p_1+p_2)\alpha}{1-p_1})}{\Gamma(1+\alpha)}k^{1+\frac{1+(p_1+p_2)\alpha}{1-p_1}}. \end{aligned} \tag{3.16}$$

The exponent β^{out} for the outdegree distribution is

$$\beta^{out} = 1 + \frac{1+(p_1+p_2)\alpha}{1-p_1} = 2 + \frac{p_1+(p_1+p_2)\alpha}{1-p_1}.$$

Similar to Section 3.6, we can prove a sharp concentration result for the indegree and outdegree distributions. For completeness, we state the following theorem for the directed preferential attachment model.

THEOREM 3.4. *For the preferential attachment model $G(p_1, p_2, \alpha)$, we have*

(1) *Almost surely the number of vertices with indegree k at time t is*

$$M_k^{in}t + O(2\sqrt{k^3 t \ln(t)}),$$

where M_k^{in} is defined in equation (3.11), (3.12), and (3.16).

(2) *Almost surely the number of vertices with outdegree k at time t is*

$$M_k^{out}t + O(2\sqrt{k^3 t \ln(t)}),$$

where M_k^{out} is defined in equation (3.14), (3.15), and (3.16).

(3) *Almost surely it is a power law directed graph with the exponent $\beta^{in} = 2 + \frac{p_2+(p_1+p_2)\alpha}{1-p_2}$ for the indegree distribution and the exponent $\beta^{out} = 2 + \frac{p_1+(p_1+p_2)\alpha}{1-p_1}$ for the outdegree distribution.*

The exponents β^{in} and β^{out} have special meanings. It is not difficult to see that both values are greater than 2. It can be observed that $p_2 + (p_1+p_2)\alpha$ is the expected increment for the indegree of the new vertex while $1 - p_2$ is the expected increment for the indegrees of the current graphs. Hence, $\beta^{in} - 2$ is the ratio of the increment of edges to the new vertex and the increment of edges to the current graph. There is a similar interpretation for $\beta^{out} - 2$ as well.

In this chapter, we focused on the rigorous analysis of the preferential attachment schemes. Further analysis on the strengthened model allowing deletion will be given in Chapter 10. Another local growth model with emphasis on duplication will be examined in Chapter 4.