On the degree distribution of a growing network model

Linda Farczadi  
University of Waterloo  
lindafarczadi@gmail.com

Nicholas Wormald*  
Monash University  
nick.wormald@monash.edu

Abstract

In this note we make some specific observations on the distribution of the degree of a given vertex in certain models of randomly growing networks. The rule for network growth is the following. Starting with an initial graph of minimum degree at least \( k \), new vertices are added one by one. Each new vertex \( v \) first chooses a random vertex \( w \) to join to, where the probability of choosing \( w \) is proportional to its degree. Then \( k \) edges are added from \( v \) to randomly chosen neighbours of \( w \).

1 Introduction

In this note we make some specific observations on the distribution of the degree of a given vertex in certain models of randomly growing networks. Fix an integer \( k \geq 2 \). We start with a seed graph \( G^1_k \) consisting of one vertex \( v_1 \) with \( k \) loops. For \( t \geq 2 \) given \( G_{t-1}^k \) we obtain \( G_t^k \) as follows:

- we add a new vertex \( v_t \) which connects first to an existing vertex \( v_i \) chosen by preferential attachment and then to \( k-1 \) neighbours of \( v_i \) chosen uniformly at random.

We let \( V_t \) and \( E_t \) denote the vertex set and edge set of \( G_t^k \). Note that \( |V_t| = t \) and \( |E_t| = kt \). We denote by \( d_t(v_i) \) the degree of vertex \( v_i \) in \( G_t^k \). We define \( N_t(v_i) \) to be the set of neighbours of vertex \( v_i \) in \( G_t^k \).

We begin with a simple derivation of the expected value of \( d_t(v_i) \) in Section 2.1, then describe some closely related existing results, and apply them to get a more precise description of the distribution of \( d_t(v_i) \) in Section 2.4.

It is straightforward to modify our results for any given initial seed graph.

2 Degree distribution

2.1 Expected degree of a given vertex

Fix vertex \( v_i \). We want to study \( d_n(v_i) \), the degree of vertex \( i \) at the \( n^{th} \) step in the process. For \( t \geq i \) we have

\[
\Pr(v_{t+1} \text{ connects to } v_i | d_t(v_i)) = \frac{d_t(v_i)}{2|E_t|} + \sum_{v_j \in N_t(v_i)} \frac{d_t(v_j)}{2|E_t|} \frac{k}{d_t(v_j)} = \frac{d_t(v_i)}{2kt} + d_t(v_i) \left( \frac{k-1}{2kt} \right) = \frac{d_t(v_i)^2}{2t}.
\]

Taking expectations of both sides gives

\[
\Pr(v_{t+1} \text{ connects to } v_i) = \frac{E[d_t(v_i)]}{2t}.
\]

*Supported by an ARC Australian Laureate Fellowship. Research supported partly by NSERC
We then have
\[ \mathbf{E}[d_{t+1}(v_i)] = \mathbf{E}[d_t(v_i)] + \mathbf{P}(v_{t+1} \text{ connects to } v_i) \]
\[ = \mathbf{E}[d_t(v_i)] + \frac{\mathbf{E}[d_t(v_i)]}{2t} \]
\[ = \left(1 + \frac{1}{2t}\right) \mathbf{E}[d_t(v_i)]. \]

Since each vertex has degree \(k\) when it joins the graph we have \(\mathbf{E}[d_i(v_i)] = k\). We obtain for \(1 \leq i \leq n\)
\[ \mathbf{E}[d_n(v_i)] = k \prod_{t=i}^{n-1} \left(1 + \frac{1}{2t}\right) \]
\[ = k\Gamma(i)\Gamma(n + 1/2) / \Gamma(n)\Gamma(i + n/2) \]
\[ = k\sqrt{n/i} (1 + O(1/i)). \]

2.2 LCD model of Bollobás and Riordan

The LCD model of Bollobás and Riordan can be described as follows: start with \(G_1^1\) the graph with one vertex and one loop; for \(t \geq 2\) given \(G_{t-1}^1\) obtain \(G_t^1\) by adding one vertex \(v_t\) and one edge connecting \(v_t\) to an existing vertex \(v_i\) chosen randomly with probability given by
\[ \mathbf{P}(v_i = s) = \begin{cases} \frac{d_{t-1}(s)}{2t-1} & \text{if } 1 \leq s \leq t-1 \\ \frac{1}{2t-1} & \text{if } s = t \end{cases} \]

Then for a given parameter \(k > 1\) obtain \(G_k^n\) by first constructing \(G_1^{kn}\) on vertices \(v'_1, v'_2, \ldots, v'_k\) using the process described above. Then identify vertices \(v'_1, \ldots, v'_k\) to form vertex \(v_1\) of \(G_k^n\), vertices \(v'_{k+1}, \ldots, v'_{2k}\) to form vertex \(v_2\), and so on.

We observe that both the \(k\)-neighbour model and the Bollobás-Riordan model satisfy the following condition
\[ \mathbf{P}(v_{t+1} \text{ connects to } v_i | d_t(v_i)) = \frac{d_t(v_i)}{\sum_{j=1}^{k} d_t(v_j)}. \]

This is known as the Barabási-Albert (BA) description. Hence the degree of a given vertex has the same distribution in both models. In particular the following result concerning the degree sequence of Bollobás and Riordan [BRSa01, BR03] applies to the \(k\)-neighbour model as well.

**Theorem 1.** Let \(N_n(d)\) be the number of vertices of degree \(d\) in \(G_k^n\) and define
\[ \alpha(k, d) = \frac{2k(k + 1)}{d(d + 1)(d + 2)}. \]

Then for a fixed \(\epsilon > 0\) and \(0 \leq d \leq n^{1/15}\) the following holds with high probability
\[ (1 - \epsilon)\alpha(k, d) \leq N_n(d) \leq (1 + \epsilon)\alpha(k, d). \]

2.3 General preferential attachment models of Ostroumova et al.

We can obtain results about the \(k\) neighbour model by observing that it belongs to a certain class of general preferential attachment models. Specifically Ostroumova et al. [ORS12] define the PA-class by considering all random graph models \(G_k^n\) that fit the following description:

- \(G_k^n\) is a graph with \(n\) vertices and \(kn\) edges obtained from the following random graph process: start at time \(n_0\) with an arbitrary seed graph \(G_{k}^0\) with \(n_0\) vertices and \(kn_0\) edges; at time \(t\) obtain the graph \(G_t^k\) from \(G_{t-1}^k\) by adding a new vertex and \(k\) edges connecting this vertex to some \(k\) vertices of \(G_{t-1}^k\).
Then $G^t_k$ belongs to the class PA-class if it satisfies the following conditions for some constants $A$ and $B$:

\[
P \left( d_{t+1}(v_i) = d_t(v_i)|G^t_k \right) = 1 - A \frac{d_t(v_i)}{n} - B \frac{1}{n} + O \left( \frac{(d_t(v_i))^2}{n^2} \right)
\]  

(1)

\[
P \left( d_{t+1}(v_i) = d_t(v_i) + 1|G^t_k \right) = A \frac{d_t(v_i)}{n} + B \frac{1}{n} + O \left( \frac{(d_t(v_i))^2}{n^2} \right)
\]  

(2)

\[
P \left( d_{t+1}(v_i) = d_t(v_i) + j|G^t_k \right) = O \left( \frac{(d_t(v_i))^2}{n^2} \right) \quad 2 \leq j \leq k
\]  

(3)

\[
P \left( d_{t+1}(v_i) = k + j|G^t_k \right) = O \left( \frac{1}{n} \right) \quad 1 \leq j \leq k
\]  

(4)

Then we can observe that our model belongs to this PA-class with parameters $A = 1/2$ and $B = 0$. Then the following two results from [ORS12] apply to our model.

**Theorem 2.** Let $N_n(d)$ be the number of vertices of degree $d$ in $G^t_k$ and $\theta(x)$ be an arbitrary function such that $|\theta(x)| < X$. There exists a constant $C > 0$ such that for any $d \geq k$ we have

\[
E \left[ N_n(d) \right] = \alpha(k, d) \left( n + \theta(Cd^4) \right)
\]

where

\[
\alpha(k, d) = \frac{2k(k+1)}{d(d+1)(d+2)} \sim 2k(k+1)d^{-3}.
\]

**Theorem 3.** For any $\delta > 0$ there exists a function $\psi(n) = o(n)$ such that for any $k \leq d \leq n^{\frac{1}{3} - \frac{\delta}{3}}$

\[
\lim_{n \to \infty} P \left( |N_n(d) - E[N_n(d)]| \geq \frac{\psi(n)}{d^3} \right) = 0.
\]

### 2.4 Urn models

We can obtain the distribution of $d_n(v_i)$ by using an urn model. Our urn contains balls of two colours: white and black. White balls represent edge-ends incident with vertex $i$ and black balls represent edge-ends not incident with vertex $i$. Suppose that the urn initially has $a_0$ white balls and $b_0$ black balls where

\[
a_0 = k, \\
b_0 = 2i - k, \\
t_0 = a_0 + b_0 = 2i.
\]

At each step, one ball is drawn randomly from the urn. If the drawn ball is white, replace it and put an additional $\alpha$ white and $\sigma - \alpha$ black. If it is black, replace and put $\sigma$ more black balls.

We now introduce some relevant results about urn models from Flajolet et al. [FDP06]. Consider a triangular urn with replacement matrix

\[
\begin{pmatrix}
\alpha & \sigma - \alpha \\
0 & \sigma
\end{pmatrix}
\]

Let $H_n(a_0, b_0)$ be the number of histories of length $n$ that start in configuration $(a_0, b_0)$ and end in configuration $(a, b)$. Note that $a + b = a_0 + b_0 + \sigma n$. Then the generating function of urn histories is defined as

\[
H(x, 1, z) := \sum_{n,a} H_n(a_0, b_0) x^a z^n n!
\]
and is given by

\[ H(x, 1, z) = x^{a_0} (1 - \sigma z)^{-b_0/\sigma} \left(1 - x^\alpha \left(1 - (1 - \sigma z)^{\alpha/\sigma}\right)\right)^{-a_0/\alpha}. \]

Letting \( \Delta := (1 - \sigma z)^{-1/\sigma} \) we have from [FDP06, Equation (74)]

\[ H(x, 1, z) = x^{a_0} \Delta^{b_0} (1 - x^\alpha \left(1 - \Delta^{-\alpha}\right))^{-a_0/\alpha} \]

(The reader may notice that [FDP06] is not entirely consistent, but in that paper, the ‘balls of the first type’ do always correspond to the first row of the replacement matrix.) Now let \( A_m \) be the number of white balls in the urn after \( m \) trials. The probability that this equals \( a_0 + x \alpha \) for some \( 0 \leq x \leq m \) is given by [FDP06, Equation (75)] as

\[ P(A_m = a_0 + x \alpha) = \left(\frac{x + \frac{a_0}{\alpha} - 1}{x}\right) \sum_{i=0}^{x} (-1)^i \binom{x}{i} \frac{[m + (b_0 - \alpha i)/\sigma - 1]_m}{[m + t_0/\sigma - 1]_m} \]

where \([\cdot]_m\) denotes falling factorial.

We next explain why the urn results apply to the random network process. At any given step, let \( s \) denote the total degrees of the vertices. If a vertex \( v \) has degree \( d \) then the probability it is chosen as the first vertex is \( d/s \). The probability it is chosen as the second vertex via any given one of its \( d \) incident edges is \( (k-1)/s \). Hence, the probability it receives a new edge is \( dk/s \). (This is similar to the derivation in Section 2.1.) So we may use an urn with \( k \) white balls and \( s \) black balls. At each step, \( s \) increases by \( 2k \), whilst \( k \) increases by \( k \) if \( v \) receives a new edge, and by 0 otherwise. Hence, at each step the number of white balls is \( k \) times the degree of \( v \), and the parameters are \( \alpha = k \) and \( \sigma = 2k \). The initial number of white balls, \( a_0 \), is \( k \) times the initial degree of the vertex, and the initial number of black balls is \( 2k \) times the initial number of edges.

Suppose the initial graph is a copy of \( K_j \). Then for any of the \( j \) initial vertices we have \( \alpha = k \), \( a_0 = k(j-1) \), \( t_0 = j(j-1) \) and \( b_0 = t_0 - a_0 = (j-k)(j-1) \). Also \( \sigma = 2k \). For each of those \( j \) vertices \( v_i \) (which initially have degree \( j-1 \)), there are \( m = n-j \) trials. So the probability that the corresponding urn process finishes with \( k(j-1 + x) \) balls is

\[ P(d_n(v_i) = j - 1 + x) = \left(\frac{x + j - 2}{x}\right) \sum_{u=0}^{x} (-1)^u \binom{x}{u} \frac{[n - j + (j-1)^2 - ku - 1]}{[n - j + R - 1]_n} \]

where \( R = j(j-1)/(2k) \). On the other hand, if \( i > j \) then \( t_0 = j(j-1) + 2(i-j)k \), \( a_0 = k^2 \), \( b_0 = t_0 - a_0 \), \( m = n - i \) and so

\[ P(d_n(v_i) = k + x) = \left(\frac{x + k - 1}{x}\right) \sum_{u=0}^{x} (-1)^u \binom{x}{u} \frac{[n - j + R - (u + k) - 1]}{[n - j + R - 1]_n} \]

We can also obtain the moments of \( A_n \) directly from the generating function of urn histories. In particular, as explained in [FDP06], the first and second moments are given by

\[ E[A_n] = \frac{\Gamma(n+1)\Gamma\left(\frac{t_0}{\sigma \alpha} + n\right)}{\sigma^n \Gamma\left(\frac{t_0}{\sigma} + n\right)} [z^n] \left(\frac{\partial H(x, 1, z)}{\partial x}\right)_{x=1} = \frac{\Gamma(n+1)\Gamma\left(\frac{t_0}{\sigma \alpha} + n\right)}{\sigma^n \Gamma\left(\frac{t_0}{\sigma} + n\right)} [z^n] a_0 \Delta^{t_0 + \alpha} \]

\[ E[A_n^2] = \frac{\Gamma(n+1)\Gamma\left(\frac{t_0}{\sigma} + n\right)}{\sigma^n \Gamma\left(\frac{t_0}{\sigma} + n\right)} [z^n] \left(\frac{\partial^2 H(x, 1, z)}{\partial x^2}\right)_{x=1} = \frac{\Gamma(n+1)\Gamma\left(\frac{t_0}{\sigma} + n\right)}{\sigma^n \Gamma\left(\frac{t_0}{\sigma} + n\right)} [z^n] \left(a_0 (a_0 + \alpha) \Delta^{t_0 + 2\alpha} + a_0 (a_0 + 1) \Delta^{t_0 + \alpha}\right). \]
Performing coefficient extraction gives

\[ [z^n] a_0 \Delta^{\alpha+0} = a_0 [z^n] (1 - \sigma z)^{-\frac{t_0 + \alpha}{\sigma}} \]

\[ = a_0 \left( -\frac{t_0 + \alpha}{n} \right)(-\sigma)^n \quad \text{using} \quad [t^n](1 + at)^r = \binom{r}{n} a^n \]

\[ = a_0 \frac{n + \frac{t_0 + \alpha}{\sigma} - 1}{n} \sigma^n \quad \text{using} \quad \binom{r}{n} = \frac{(n + r - 1)}{n}(-1)^n \]

\[ = a_0 \sigma^n \frac{\Gamma \left( n + \frac{t_0 + \alpha}{\sigma} \right)}{\Gamma (n + 1) \Gamma \left( \frac{t_0 + 2 \alpha}{\sigma} \right)} \quad \text{using} \quad \binom{x}{y} = \frac{\Gamma (x + 1)}{\Gamma (y + 1) \Gamma (x - y + 1)} \]

And by linearity:

\[ [z^n] (a_0 (a_0 + \alpha) \Delta^{t_0+2\alpha} - a_0(\alpha + 1)\Delta^{t_0+\alpha}) = a_0 (a_0 + \alpha) [z^n] \Delta^{t_0+2\alpha} - a_0(\alpha + 1) [z^n] \Delta^{t_0+\alpha} \]

\[ = a_0 \left[ (a_0 + \alpha) \sigma^n \frac{\Gamma \left( n + \frac{t_0 + 2\alpha}{\sigma} \right)}{\Gamma (n + 1) \Gamma \left( \frac{t_0 + 2\alpha}{\sigma} \right)} - (\alpha + 1)\sigma^n \frac{\Gamma \left( n + \frac{t_0 + \alpha}{\sigma} \right)}{\Gamma (n + 1) \Gamma \left( \frac{t_0 + \alpha}{\sigma} \right)} \right]. \]

Plugging these coefficients back in the equations for the first and second moment we obtain

\[ E[A_n] = a_0 \gamma \left( \frac{\tau_0}{\alpha} \right) \frac{\Gamma \left( n + \frac{t_0 + 2\alpha}{\sigma} \right)}{\Gamma \left( \frac{t_0 + 2\alpha}{\sigma} \right) \Gamma \left( \frac{\tau_0}{\alpha} + n \right)} \]

\[ E[A_n^2] = a_0 \left[ \frac{(a_0 + \alpha) \gamma \left( \frac{\tau_0}{\alpha} \right) \Gamma \left( n + \frac{t_0 + 2\alpha}{\sigma} \right)}{\Gamma \left( \frac{t_0 + 2\alpha}{\sigma} \right) \Gamma \left( \frac{\tau_0}{\alpha} + n \right)} + \frac{(\alpha + 1) \gamma \left( \frac{\tau_0}{\alpha} \right) \Gamma \left( n + \frac{t_0 + \alpha}{\sigma} \right)}{\Gamma \left( \frac{t_0 + \alpha}{\sigma} \right) \Gamma \left( \frac{\tau_0}{\alpha} + n \right)} \right] \]

Applying these results to our example where \( a_0 = k, b_0 = 2i - k, \alpha = 1, \sigma = 2 \) and the number of trials is \( n - i \) we obtain

\[ E[d_n(i)] = \frac{k \Gamma (i) \Gamma (n + 1/2)}{\Gamma (i + 1/2) \Gamma (n)} \]

\[ E[d_n(i)^2] = k \left[ \frac{(k + 1) \Gamma (i) \Gamma (n + 1)}{\Gamma (i + 1) \Gamma (n)} + \frac{2 \Gamma (i) \Gamma (n + 1/2)}{\Gamma (i + 1/2) \Gamma (n)} \right] \]

\[ = k \frac{(k + 1)n}{i} + 2 E[d_n(i)]. \]

We note that the value for the first moment matches the one obtained previously by solving the simple recursion in Section 2.1.

References

[BR03] B. Bollobás and O.M. Riordan. Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, pages 1–34, 2003.

[BRS+01] B. Bollobás, O. Riordan, J. Spencer, G. Tusnády, et al. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[FDP06] P. Flajolet, P. Dumas, and V. Puyhaubert. Some exactly solvable models of urn process theory. *Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities*, 59–118, Discrete Math. Theor. Comput. Sci. Proc., AG, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2006.

[Gre11] EA Grechnikov. An estimate for the number of edges between vertices of given degrees in random graphs in the bollobás–riordan model. *Moscow Journal of Combinatorics and Number Theory*, 1(2), 2011.
[OG11] L. Ostroumova and E. Grechnikov. The distribution of second degrees in the bollobás–riordan random graph model. Arxiv preprint arXiv:1108.5585, 2011.

[ORS12] L. Ostroumova, A. Ryabchenko, and E. Samosvat. Generalized preferential attachment: tunable power-law degree distribution and clustering coefficient. Arxiv preprint arXiv:1205.3015, 2012.