Network Project
A Growing Network Model

CID: 01881584

17th February 2023

Abstract: You may want to write a concise abstract that briefly puts the work into context, explains what was done, how it was done, the main results, the conclusions you could draw and the implications.

Word Count: ???? words excluding font page, figure captions, table captions, acknowledgement and bibliography.

0 Introduction

Brief paragraph with your Aims and Objectives. Marked under the "Professional Skills" category.

Definition

Give your definition of the BA model as implemented by you in this work. [**3 marks**]
TOTOTO

1 Phase 1: Pure Preferential Attachment Π_{pa}

1.1 Implementation

1.1.1 Numerical Implementation

Describe how you implemented the BA model numerically. [**3 marks**]

1.1.2 Initial Graph

What type of initial graph do you use and why? [**3 marks**]

Tdddd therefore for our system, $p(k < m, t) = n(k < m, t) = 0$ (the *lowest degree condition*).

1.1.3 Type of Graph

What type of graph do you produce and why? [**3 marks**]

1.1.4 Working Code

How do you know that your programme is working correctly? [**2 marks**]

1.1.5 Parameters

Describe the parameters your programme needs, the values you chose, and why. [**2 marks**]

1.2 Preferential Attachment Degree Distribution Theory

1.2.1 Theoretical Derivation

Give your best theoretical derivation for the form of the degree distribution $p(k)$ in the long-time limit for Preferential Attachment (PA) in the BA model. [**4 marks**]

Let t be the number of vertices added to the initial graph (aka time), $N(t) = N_0 + t$ be the number of vertices, $E(t) = E_0 + mt$ be the number of edges, $p(k,t)$ be the probability that a random vertex has degree k, and $n(k,t) = N(t)p(k,t)$ be the expected number of vertices with degree k. Then, in each step, for each degree value k, the expected number of vertices with that degree that has a new edge connected to them is $m\Pi(k,t)n(k,t)$. Also, in each step, one vertex with degree m is being added. Hence we can describe how the *expected* number of vertices of a certain degree changes in each iteration:

$$n(k,t+1) = n(k,t) + m\left[\Pi(k-1,t)n(k-1,t) - \Pi(k,t)n(k,t)\right] + \delta_{k,m} \tag{1}$$

We can express 1 with $p(k,t)$ and $N(t)$:

$$N(t+1)p(k,t+1) \;=\; N(t)p(k,t) + mN(t)\left[\Pi(k-1,t)p(k-1,t) - \Pi(k,t)p(k,t)\right] + \delta_{k,m}$$

Consider the limit

$$\lim_{t\to\infty} p(k,t) \equiv p_\infty(k)$$

In this limit, the probability distribution function of the degrees should stay constant, and is given by the equation

$$p_\infty(k) = \lim_{t\to\infty} m\left[N(t)\Pi(k-1,t)p_\infty(k-1) - N(t)\Pi(k,t)p_\infty(k)\right] + \delta_{k,m} \tag{2}$$

Up until this point, the shape of $\Pi(k,t)$ could be anything. Here, to solve this equation, we apply the Preferential Attachment model, which requires $\Pi(k,t) \propto k$. Writing this as $\Pi(k,t) = c(t)k$, we now require Π to be normalized and hence find $c(t)$:

$$\begin{aligned}
\sum_{i\in\text{vertices}} \Pi(k_i,t) &= 1\\
c(t)\sum_{i\in\text{vertices}} k &= 1\\
c(t)\cdot 2E(t) &= 1\\
\Pi(k,t) &= \frac{k}{2E(t)}
\end{aligned}$$

Now: we want to calculate the following limit:

$$\lim_{t\to\infty}\frac{N(t)}{E(t)} = \lim_{t\to\infty}\frac{N_0+t}{E_0+mt} = \frac{1}{m}$$

Plugging these two results into 2, we obtain

$$p_\infty(k) = \frac{1}{2}\left[(k-1)p_\infty(k-1) - kp_\infty(k)\right] + \delta_{k,m} \tag{3}$$

First, we consider the case $k > m$. To solve this, we rearrange as

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k-1}{k+2}$$

and note that if we make $p_\infty(k)$ proportional to $[k(k+1)(k+2)]^{-1}$, then $p_\infty(k-1)$ will be proportional to this sequence 'shifted' – ergo $p_\infty(k-1) \propto [(k-1)k(k+1)]^{-1}$ – and their ratio is then exactly what's required. Therefore

$$p_\infty(k) = \frac{A}{k(k+1)(k+2)} \quad \text{when} \quad k > m$$

Now we consider the case $k = m$. Applying the lowest degree condition yields the equation

$$p_\infty(m) = -\frac{m}{2}p_\infty(m) + 1$$

Hence

$$p_\infty(m) = \frac{2}{m+2}$$

We now determine the constant A by considering:

$$\begin{aligned} \frac{p_\infty(m+1)}{p_\infty(m)} &= \frac{m}{m+3} \\ \frac{m+2}{2}\frac{A}{(m+1)(m+2)(m+3)} &= \frac{m}{m+3} \\ A &= 2m(m+1) \end{aligned}$$

When we plug this into $p_\infty(k)$, we see that setting $k = m$ yields exactly $\frac{2}{m+2}$, so we can write the result as

$$p_\infty(k) = \begin{cases} \frac{2m(m+1)}{k(k+1)(k+2)} & k \geq m \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

1.2.2 Theoretical Checks

Check your approximate theoretical solution for $p(k)$ has the correct properties. [**4 marks**]

In our derivation of $p_\infty(k)$, we never once used the fact that it must be normalized to 1, since it's a probability distribution function. We have to check if this is true:

$$\begin{aligned} \sum_{k=m}^{\infty} p_\infty(k) &= 2m(m+1)\sum_{k=m}^{\infty}\frac{1}{k(k+1)(k+2)} \\ &= 2m(m+1)\sum_{k=m}^{\infty}\left[\frac{1}{2}\frac{1}{k} - \frac{1}{k+1} + \frac{1}{2}\frac{1}{k+2}\right] \\ &= 2m(m+1)\left(\frac{1}{2}\left(\frac{1}{m} + \frac{1}{m+1}\right) + \sum_{k=m}^{\infty}\left[\frac{1}{k+2} - \frac{1}{k+1}\right]\right) \end{aligned}$$

We recognize that the series on the right side is telescoping, hence we have

$$
\begin{aligned}
&= 2m(m+1)\left(\frac{2m+1}{2m(m+1)} - \frac{1}{m+1} + \lim_{k\to\infty}\frac{1}{k+2}\right) \\
&= 2m(m+1)\left(\frac{1}{2m(m+1)}\right) \\
\sum_{k=m}^{\infty} p_\infty(k) &= 1
\end{aligned}
$$

1.3 Preferential Attachment Degree Distribution Numerics

1.3.1 Fat-Tail

How did you deal with any problems that a fat-tailed distribution causes? [**4 marks**]

1.3.2 Numerical Results

Show how you compared your theoretical result to your numerical data for fixed N but different m. Give a visualisation which shows you whether you have a good or bad fit of your numerical data to your theoretical results. [**4 marks**]

1.3.3 Statistics

Show statistically whether your numerical data fits your theoretical predictions. [**6 marks**]

1.4 Preferential Attachment Largest Degree and Data Collapse

1.4.1 Largest Degree Theory

Give your best theoretical estimate of how the largest expected degree, k_1 (subscript 1 indicating the degree of the vertex ranked first by degree size) depends on the number of vertices N in a finite size system and on m the number of edges added each step. [**4 marks**]

The degree distribution of a large newtork with N vertices will approximately follow the derived distribution in 4, hence

$$n_N(k) \approx N p_\infty(k)$$

The largest degree k_1 is special in the sense that the expected amount of vertices with degree k_1 *or higher* is 1:

$$N \sum_{k=k_1}^{\infty} p_\infty(k) = 1 \tag{5}$$

We already know how to calculate this sum, so we evaluate:

$$
\begin{aligned}
N\sum_{k=k_1}^{\infty} &= 2Nm(m+1)\left[\frac{1}{2}\left(\frac{1}{k_1}+\frac{1}{k_1+1}\right)-\frac{1}{k_1+1}\right] \\
&= Nm(m+1)\frac{1}{k_1(k_1+1)} \\
k_1^2+k_1 &= Nm(m+1) \\
k_1 &= \frac{-1\pm\sqrt{1+4Nm(m+1)}}{2}
\end{aligned}
$$

k_1 cannot be negative, hence we take the positive root:

$$
k_1 = \frac{-1+\sqrt{1+4Nm(m+1)}}{2} \tag{6}
$$

1.4.2 Numerical Results for Largest Degree

Study of the behaviour of k_1 as N is varied for one sensible fixed value of m (justify your choice of parameters). Estimate uncertainties/errors where possible. Compare against your theoretical prediction. [**4 marks**]

1.4.3 Data Collapse

Illustrate the finite size effects by studying a single value of m (justify your choice for m) but varying N and looking for data collapse. You should describe how you tried to use an understanding of the k_1 scale to look for any finite size effects in the tail of the distribution, describing any results found. Further mathematical investigation of finite size effects is not required as it is extremely hard to do. [**4 marks**]

The existence of $k_1(N)$ means that a network of size N will experience a cut-off in its degree distribution around $k = k_1$. We use the following ansatz to describe this finite-size effect:

$$
p_N(k) = p_\infty(k)\mathcal{F}\left(\frac{k}{k_1}\right) \tag{7}
$$

In our data collapse, we separate $\mathcal{F}$ like so:

$$
\mathcal{F}\left(\frac{k}{k_1}\right) = \frac{p_N(k)}{p_\infty(k)} = \frac{k(k+1)(k+2)}{A}p_N(k)
$$

and plot it against k/k_1 (so we divide the x-axis by k_1). $p_N(k)$ here represents the actual measured values for the distribution on a network of a specific size.

2 Phase 2: Pure Random Attachment Π_{rnd}

2.1 Random Attachment Theoretical Derivations

2.1.1 Degree Distribution Theory

Give your best theoretical derivation for the form of the degree distribution in the long-time limit. Check your approximate theoretical solution has the correct properties. [**4 marks**]

For this model, we can use the same equation 2 as for preferential attachment, we'll just use different $\Pi(k,t)$. Namely, in the random attachment model, every vertex has the same selection probability as any other. Hence

$$\Pi_{\mathrm{RA}}(k,t) = \frac{1}{N(t)}$$

This means our governing equation is

$$p_\infty(k) = m\left[p_\infty(k-1) - p_\infty(k)\right] + \delta_{k,m} \tag{8}$$

Hence for $k = m$, using the lowest degree condition:

$$p_\infty(m) = \frac{1}{1+m}$$

And for $k > m$:

$$p_\infty(k) = \frac{m}{m+1} p_\infty(k-1)$$

By induction, we obtain

$$p_\infty(k) = \left(\frac{m}{m+1}\right)^{k-m} p_\infty(m) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m}$$

Hence the degree probability distribution function in this model is

$$p_\infty(k) = \begin{cases} \frac{1}{m+1}\left(\frac{m}{m+1}\right)^{k-m} & k \geq m \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

We need to check if this is normalized:

$$\begin{aligned} \sum_{k=m}^{\infty} p_\infty(k) &= \frac{1}{m+1}\sum_{k=m}^{\infty}\left(\frac{m}{m+1}\right)^{k-m} \\ &= \frac{1}{m+1}\sum_{x=0}^{\infty}\left(\frac{m}{m+1}\right)^{x} \\ &= \frac{1}{m+1}\cdot\frac{1}{1-\frac{m}{m+1}} \\ &= 1 \end{aligned}$$

Therefore $p_\infty(k)$ is normalized. We see that, rather than a power law like in the preferential attachment model, the probability distribution for the random attachement model is governed by exponential decay.

2.1.2 Largest Degree Theory

Give your best theoretical estimate of how the largest expected degree, k_1 depends on the number of vertices N in a finite size system. **[2 marks]**

By substituing 9 into 5 we obtain:

$$
\begin{aligned}
1 &= N \sum_{k=k_1}^{\infty} \frac{1}{m+1} \left(\frac{m}{m+1} \right)^{k-m} \\
&= N \left(\frac{m}{m+1} \right)^{k_1 - m} \sum_{k=m}^{\infty} \frac{1}{m+1} \left(\frac{m}{m+1} \right)^{k-m} \\
&= N \left(\frac{m}{m+1} \right)^{k_1 - m} \\
N &= \left(\frac{m+1}{m} \right)^{k_1 - m} \\
k_1 &= m + \frac{\ln N}{\ln\left(\frac{m+1}{m}\right)}
\end{aligned}
$$

2.2 Random Attachment Numerical Results

2.2.1 Degree Distribution Numerical Results

Show how you compared your theoretical result to your numerical data for different m at one large value of N. Is your theory a good fit to your data? How did you arrive at your conclusion? [**6 marks**]

2.2.2 Largest Degree Numerical Results

Study of the behaviour of k_1 as N is varied for one sensible fixed value of m (justify your choice of parameters) including estimates of uncertainties on any measurements. Compare against your theoretical prediction. Illustrate the finite size effects by looking for possible data collapse. [**4 marks**]

3 Phase 3: Existing Vertices Model

3.1 Existing Vertices Model Theoretical Derivations

Give your best theoretical derivation for the form of the degree distribution in the long-time limit for the specific scenario given, namely where some edges added run between exiting vertices while the rest are between the new vertex and one existing vertex as before. Check your solution. [**8 marks**]

Here, we need to alter the equation 1, since there are more ways the number of vertices with a specific edge can change. Namely, in each step, r edges are added to the system from the new vertex to r existing vertices using PDF $\Pi_1(k,t)$ and $(m-r)$ edges are added between existing vertices using PDF $\Pi_2(k,t)$. The first part can be interpreted the same way as the previous two models, with the number of vertices with changing k being r rather than m; the second part can be as well, except the number of vertices changing k is $2(m-r)$, and there will be no Kronecker delta term, since no new vertex is being added. Hence the new governing equation is

$$n(k,t+1) = n(k,t) + r\left[\Pi_1(k-1,t)n(k-1,t) - \Pi_1(k,t)n(k,t)\right] \\ + 2(m-r)\left[\Pi_2(k-1,t)n(k-1,t) - \Pi_2(k,t)n(k,t)\right] + \delta_{r,k} \quad (10)$$

We can take the limit $t \to \infty$ and express 10 with $p_\infty(k)$:

$$p_\infty(k) = \lim_{t\to\infty} N(t)\left[2(m-r)\Pi_2(k-1,t) + r\Pi_1(k-1,t)\right] p_\infty(k-1) \\ - N(t)\left[2(m-r)\Pi_2(k,t) + r\Pi_1(k,t)\right] p_\infty(k) + \delta_{r,k} \quad (11)$$

Note that since the vertices we're adding begin with degree r rather than m, the lowest degree condition here states that $p(k<r,t) = 0$.

We apply the specific shapes of $\Pi_1(k,t) = \Pi_{\mathrm{ra}}(k,t) = 1/N(t), \Pi_2(k,t) = \Pi_{\mathrm{pa}}(k,t) = k/(2mN(t))$ and rearrange:

$$p_\infty(k) = \left[\frac{m-r}{m}(k-1) + r\right] p_\infty(k-1) - \left[\frac{m-r}{m}k + r\right] p_\infty(k)$$

$$\left[1 + r + \frac{m-r}{m}k\right] p_\infty(k) = \left[r - \frac{m-r}{m} + \frac{m-r}{m}k\right] p_\infty(k-1) + \delta_{r,k}$$

For $k = r$:

$$p_\infty(r) = \frac{1}{1 + r + \frac{m-r}{m}r} = \frac{m}{2mr + m - r^2}$$

For $k > r$:

$$\frac{p_\infty(k)}{p_\infty(k-1)} = \frac{k + \frac{mr+r-m}{m-r}}{k + \frac{mr+m}{m-r}}$$

To solve this equation, we generalize our approach for the PA model and note that for a function $f(x) = A\Gamma(x+1+a)/\Gamma(x+1+b)$, we have

$$\frac{f(x)}{f(x-1)} = \frac{\Gamma(x+1+a)\Gamma(x+b)}{\Gamma(x+a)\Gamma(x+1+b)} = \frac{(x+a)\Gamma(x+a)}{\Gamma(x+a)}\frac{\Gamma(x+b)}{(x+b)\Gamma(x+b)} = \frac{x+a}{x+b}$$

Identifying $x \equiv k, a \equiv \frac{mr+r-m}{m-r}, b \equiv \frac{mr+m}{m-r}$ we obtain

$$p_\infty(k>r) = A\frac{\Gamma(k+1+\frac{mr+r-m}{m-r})}{\Gamma(k+1+\frac{mr+m}{m-r})} = A\frac{\Gamma(k+z_1)}{\Gamma(k+z_2)} \text{ where } z_1 = \frac{mr}{m-r}, z_2 = \frac{mr+2m-r}{m-r}$$

Now:

$$\frac{p_\infty(r+1)}{p_\infty(r)} = \frac{(m-r)(r+1) + mr - m + r}{(m-r)(r+1) + mr + m} = A\frac{\Gamma(r+1+z_1)}{\Gamma(r+1+z_2)}\frac{2mr+m-r^2}{m}$$

$$A = \frac{mr}{(r+1)(2mr+m-r^2)}\frac{\Gamma(r+1+z_2)}{\Gamma(r+1+z_1)}$$

We can check whether this is normalized: we want all the probabilities to sum up to 1. By induction, we can find the partial sum:

$$\sum_{k=k_{\min}>r}^{k_{\max}} p_\infty(k) = A\left[\frac{\Gamma(z_1+k_{\min})}{\Gamma(z_2+k_{\min})}\left(1 + r + k_{\min}\left(1 - \frac{r}{m}\right)\right) + \right. \\ \left. \frac{\Gamma(z_1+1+k_{\max})}{\Gamma(z_2+1+k_{\max})}\left(\frac{r}{m} - r - 2 + k_{\max}\left(\frac{r}{m} - 1\right)\right)\right] \quad (12)$$

The gamma fraction in the second term goes like $(k_{\max}+1)^{z_1-z_2} = (k_{\max}+1)^{-\frac{2m-r}{m-r}}$, so the sum converges if $-\frac{2m-r}{m-r} < -1 \Leftrightarrow m > 0, m > r$, which we can automatically assume to be satisfied. Hence:

$$\begin{aligned}
\sum_{k=r}^{\infty} p_\infty(k) &= \frac{m}{2mr+m-r^2} + A\frac{\Gamma(z_1+r+1)}{\Gamma(z_2+r+1)}\frac{(r+1)(2m-r)}{m} \\
&= \frac{m}{2mr+m-r^2} + \frac{(2m-r)r}{2mr+m-r^2} = 1
\end{aligned}$$

So $p_\infty(k)$ is normalized and looks like so:

$$p_\infty(k) = \begin{cases} \frac{m}{2mr+m-r^2} & k = r \\ \frac{mr}{(r+1)(2mr+m-r^2)}\frac{\Gamma(r+1+z_2)}{\Gamma(r+1+z_1)}\frac{\Gamma(k+z_1)}{\Gamma(k+z_2)} & k > r \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

3.2 Existing Vertices Model Numerical Results

Show how you compared your theoretical result to your numerical data for different m at one large value of N. Is your theory a good fit to your data? How did you arrive at your conclusion? Analyse the finite size effects by looking for possible data collapse. [**8 marks**]

4 Conclusions

Just one or two sentences to round off the report. Marked under the "Professional Skills" category.

Delete text in this box from your final version.

Note: *The Acknowledgement and Bibliography do not count towards the 2500 word limit nor the 16 page limit. An acknowledgement section is not required. For a report of this type and size, references may well not be needed but if you do reference other work, such as a useful book [1] or a useful result [2], a formal bibliography is one way to do this. No formal marks for either of these sections but they could be considered as part of the overall Organisation, Presentation or English marks.*

Acknowledgements

Optional. Not required but you might want to thank A.Demonstrator or A.Friend for help.

References

[1] K.Christensen and N.Maloney, *Complexity and Criticality*, Imperial College Press, London, 2005.

[2] T.S. Evans, *Astounding Paper*, Journal of Amazing Results, **9** (2023) 3134.

Delete text in this box from your final version.

Professional Skills
Do not include this section in your report, it is for your information only
Your report should be

- written in clear comprehensible English [**6 marks**]
- be well presented [**6 marks**]
- well organised and follows brief [**6 marks**]

This is highly subjective but generally most reports score well on these aspects. Many of the problems come from plots which might have: small unreadable fonts used for labels, labels used for variables are different from those in the text (they reflect the naming scheme used in the programme), unexplained equations of some fit with unrealistic accuracy in unexplained coefficients.

[**Total 100 marks**]