Depreferential attachment model through preferential attachment

M.Stat. Dissertation midterm report by: Somak Laha
Indian Statistical Institute, Kolkata

Supervisor: Antar Bandyopadhyay
Theoretical Statistics and Mathematics Division,
Indian Statistical Institute.

February 21, 2022

1 Introduction

Erdős-Rényi [1, 2] random graph is an example of a *static model* of random graphs. By *static model*, it is meant that the vertex set of such graph do not evolve with time. Instead, the edge set evolve with time according to certain probability distributions. Unlike *static models*, most networks grow with time, i.e., have a sequentially increasing vertex set. There is a substantial amount of work investigating such models, also termed as *dynamic models*. Such random graph models come in use for the study of complex networks, often in the context of World Wide Web and also citation or biological networks. One of the most discussed model in this context is the *preferential attachment model*, first suggested by Barabási and Albert [3]. The Barabási-Albert [3] model is an algorithm for generating random graphs through a *preferential attachment* mechanism. This model has succeeded to provide a possible explanation for the occurrence of power-law degree sequences, often observed in real life networks.

In a preferential attachment model proposed by Albert and Barabási, vertices having a fixed number of half-edges are sequentially added to the graph. Given the graph at time t, the half-edges of the vertex labelled $t + 1$ choose some of the older vertices to be incident on according to a probability distribution. This probability distribution is usually taken to be an increasing, possibly affine, linear function of the degrees of the older vertices. Consequently, vertices those already have a high degree attract half-edges of later vertices. This phenomenon is often called the *rich gets richer* phenomenon.

Our aim is to suggest a random graph model where the *rich gets richer* phenomenon is penalized to some extent. The suggested model will be called *Depreferential attachment model through preferential attachment.*

2 Motivation and description of the model

2.1 A relevant scenario

Consider modelling an ecosystem in which every newly arriving species predates older species. It is reasonable to believe that a new predator would want to choose its prey from such species that are not eaten by many. So, it would not eat the first few most popular preys, and look for the popular one from the remaining options. As a result, a popular prey will face restrictions in getting increasingly popular, which in turn will help it survive.

Thus, if having a lot of predators is thought to be equivalent to being rich, the *rich getting richer* phenomena will be penalized to some extent in the above setup. One may be interested in modelling such an ecosystem and now we shall move on to describe the model which is also expected to restrict the *rich getting richer* phenomena.

2.2 Description of the model

Our model will have two parameters, denoted by p_t and m. For each $t \geq 1$, p_t is a real number in the interval $(0,1)$ and m is a positive integer. We shall start with two vertices and an edge between them. At each discrete time point $t \geq 1$, a new vertex appears with m half-edges. The other end of each of these half-edges is to be attached to some of the already existing vertices. Let us denote the process by $(G_t^{(m,p_t)})_{t\geq 0}$. So at time point $t = 1$, the graph is $G_1^{(m,p_t)}$ and at time point $t = 2$, the graph is $G_2^{(m,p_t)}$ and so on. And, according to the initial condition, $G_0^{(m,p_t)}$ consists of only two vertices and an edge between them.

At time $t \geq 1$, there were total $(t+1)$ already existing vertices excluding the newly arrived one. Among the already existing ones, we shall choose $\lfloor (t+1)p_t \rfloor$ many in such a way that the probability of choosing the vertex v_k is an increasing function of its degree in $G_{t-1}^{(m,p_t)}$. These preferentially chosen vertices are made taboo, i.e., any of the m half-edges of the newly arrived vertex would not attach to any of these taboo vertices.

In the next step, each of the m half-edges are attached to non-taboo vertices in such a way that the probability of attaching an edge to v_j is proportional to its degree in $G_{t-1}^{(m,p_t)}$. Note that, in this way, the model is not allowed to have self-loops. This model differs from the Albert-Barabasi model in the way that there was no concept of creating taboo vertices in the latter. Also, by forcing the higher degree vertices to not attach to newer vertices, the *rich gets richer* phenomena is expected to get suppressed to some degree.

2.3 Taboo-ing schemes

Consider the parameter p_t first. There are two possible scenarios that we shall take into consideration.

(i) One possible case is that for each $t \geq 1$, we shall taboo a fixed number of vertices, say k, given that the size of the graph is larger than k. In such cases, p_t should be taken as

$$p_t = \frac{k}{t+1} \cdot \mathbb{1}_{\{k<t+1\}}.$$

(ii) Another possible scenario is that the number of vertices to be made taboo at time t is proportional to the size of the graph at that time. Then, we shall take p_t to be fixed $p\,(0 < p < 1)$. And in such cases, the model will be denoted simply by $(G_t^{(m,p)})_{t\geq 0}$.

Now that we have decided how many vertices to be made taboo, it is important to discuss the algorithms which are used to preferentially choose the taboo vertices.

(i) **Hard-core taboo:** At time $t \geq 1$, $\lfloor (t+1)p_t \rfloor$ vertices with the highest degrees in $G_{t-1}^{(m,p_t)}$ are made taboo.

(ii) **Soft-core taboo:** At time $t \geq 1$, $\lfloor (t+1)p_t \rfloor$ vertices are made taboo where the probability of choosing the vertex v_j is directly proportional to its degree in $G_{t-1}^{(m,p_t)}$.

3 Features of interest

The following statistics of the random graph $(G_t^{(m,p_t)})_{t\geq 0}$ are of our interest for this study.

(i) **Degree of fixed vertex**: For $i \geq 1$, let $D_i(t)$ be the degree of the i-th vertex at time t. We shall investigate the nature of $D_i(t)$ as $t \to \infty$ fixing i.

(ii) **Degree distribution**: Since at time t, there are $t+2$ vertices present in the graph, including the newly arrived one, the proportion of vertices with degree k at time t is denoted as

$$P_k(t) = \frac{1}{t+2} \sum_{i=1}^{t+2} 1_{\{D_i(t)=k\}}.$$

We shall investigate the asymptotic behaviour of $P_k(t)$ as $t \to \infty$.

(iii) **Maximal degree**: The maximal degree of a vertex in $G_t^{(m,p_t)}$ is defined as

$$M_t = \max_{1 \le i \le t+2} D_i(t).$$

We want to study the behaviour of M_t as $t \to \infty$.

4 Simulations

To start with, we keep our interests limited to the case $m = 1$. The simulations of $G_t^{(1,p)}$ has been performed to observe the asymptotic behaviour of the different features of interest. The python codes for the simulations are available in a public repository [4].

4.1 Degree of fixed vertex

Recall that, for $0 < i \le t+1$, $D_i(t)$ is the degree of the i-th vertex at time t. We have taken $i = 10$ and $p \in \{0.01, 0.1, 0.25\}$ and for each case, performed 500 simulation to plot the mean, variance and standard deviation of $D_i(t)$ as a function of t. The simulations have been run up to $t = 2000$.

4.1.1 Hard-core taboo-ing

We have adopted the hard-core taboo-ing scheme for the following simulations. Following is the plot for the case $\boldsymbol{p = 0.1}$. The plots for the other values of p is given in the GitHub repository [4].

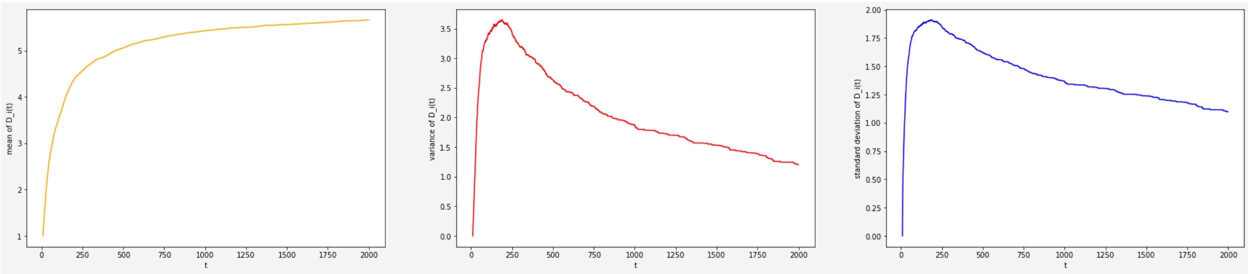


Figure 1: Mean, variance and standard deviation of $D_i(t)$ as function of t

Observing the above plot, it seems like $D_i(t)$ gets bounded above as t approaches $+\infty$ for hard-core taboo-ing. The bound is lower for higher values of p. This conjecture is supported by the value of the standard deviation of $D_i(t)$ which approaches 0 with increasing t.

4.1.2 Soft-core taboo-ing

For this part, we have adopted the soft-core taboo-ing scheme. Following is the case for $\boldsymbol{p = 0.1}$. Plots for the other values of p are given in the GitHub repository [4].

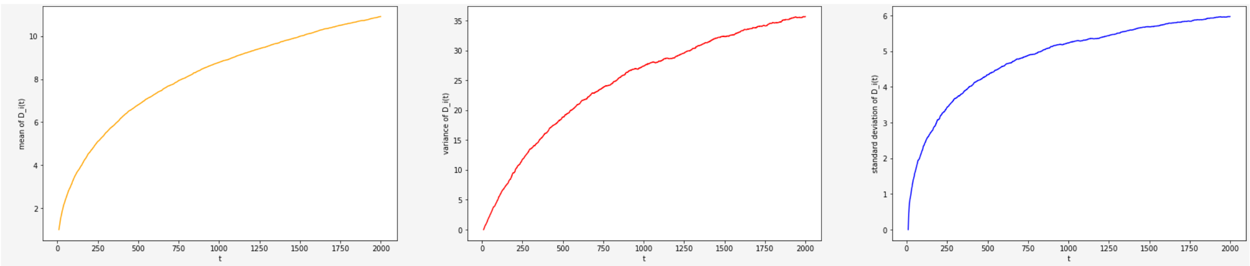


Figure 2: Mean, variance and standard deviation of $D_i(t)$ as function of t

Unlike hard-core taboo-ing, there is no apparent upper bound of degree of fixed vertices in soft-core taboo-ing.

4.2 Degree distribution

The proportion of vertices with degree k in the graph at time t is denoted by

$$P_k(t) = \frac{1}{t+2} \sum_{i=1}^{t+2} \mathbb{1}_{\{D_i(t)=k\}}.$$

In order to study asymptotic behaviour of $P_k(t)$, we have plotted the mean, variance and standard deviation of $P_k(2000)$ based on 500 simulations for different values of $p \in \{0.01, 0.05, 0.1, 0.25\}$.

4.2.1 Hard-core taboo-ing

We have adopted the hard-core taboo-ing scheme for the following simulations. Following is the plot for $\boldsymbol{p = 0.1}$. Plots for the other values of p is attached in the GitHub repository [4]. In the following plot, the variance is in 10^{-5} scale.

Figure 3: Mean, variance and standard deviation of $P_k(2000)$ as function of k

4.2.2 Soft-core taboo-ing

We have adopted the soft-core taboo-ing scheme for the following simulations. Following is the plot for $\boldsymbol{p = 0.1}$. Plots for the other values of p are given in the GitHub repository [4]. In the following plot, the variance is in 10^{-5} scale.

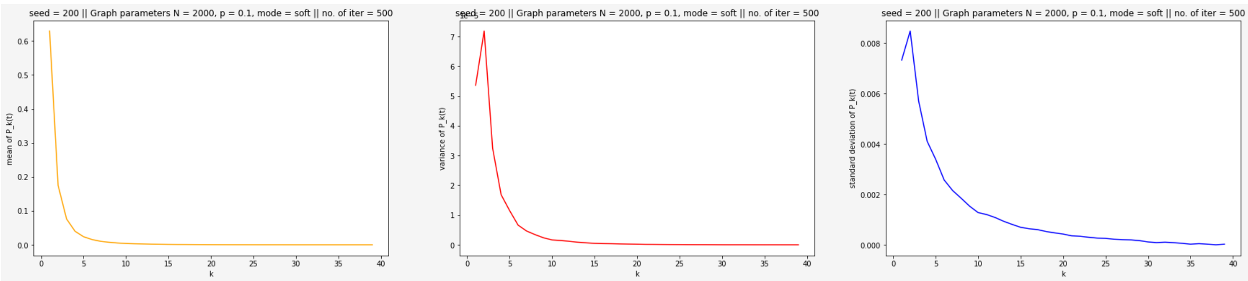


Figure 4: Mean, variance and standard deviation of $P_k(2000)$ as function of k

It may be observed that these plots have similar structure as in preferential attachment model. For smaller values of p, the plots are closer to the same for *preferential attachment model* [5], which may be thought of as the case corresponding to $p = 0$.

4.3 Maximal degree

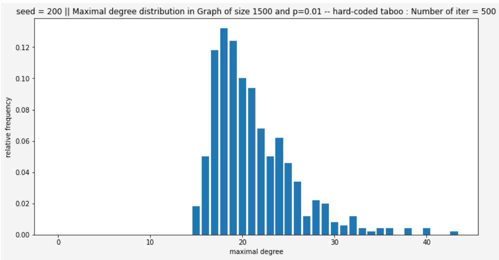
The maximal degree in $G_t^{(1,p)}$ is defined as

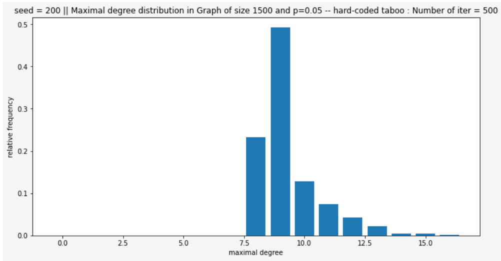
$$M_t = \max_{1 \le i \le t+2} D_i(t).$$

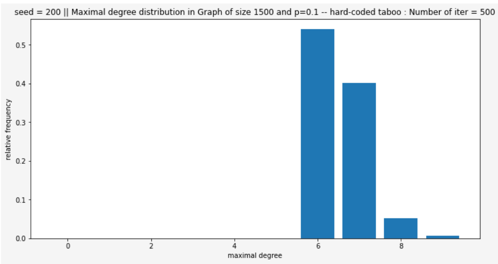
Like all the other features of interest, we have performed simulations to study the asymptotic behaviour of M_t. For each taboo-ing scheme, we have plotted the histogram of the distribution of M_t by varying $p \in \{0.01, 0.05, 0.1, 0.2\}$ and $t \in \{500, 1000, 1500, 2000\}$.

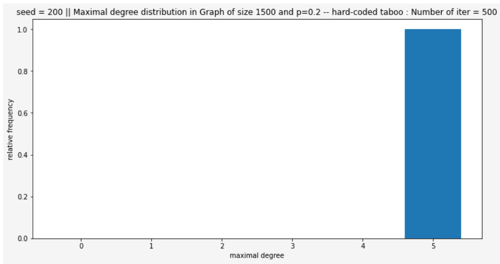
4.3.1 Hard-core taboo-ing

In this part, we fix $t = 1500$ and vary p to obtain the following histograms. The plots correspond to the values $0.01, 0.05, 0.1, 0.2$ respectively of p.









The following figure illustrates how the average maximal degree changes with p. Not surprisingly, as p increases, i.e., more is the number of taboo vertices, lesser is the expected maximal degree.

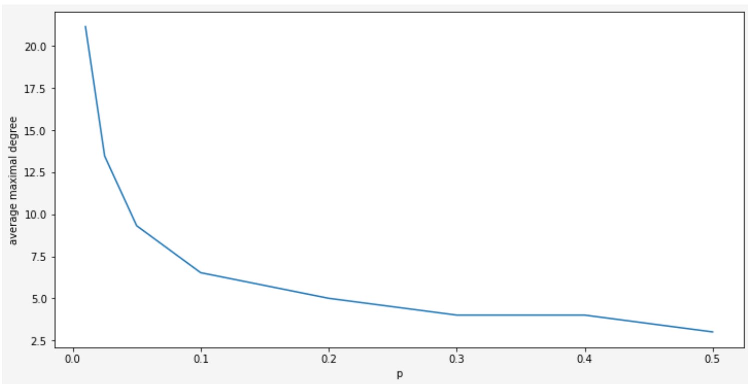
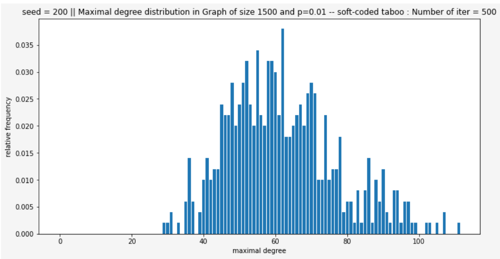


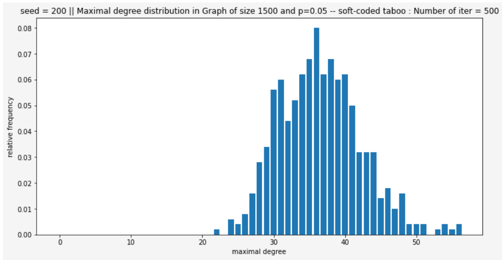
Figure 7: Average maximal degree as a function of p

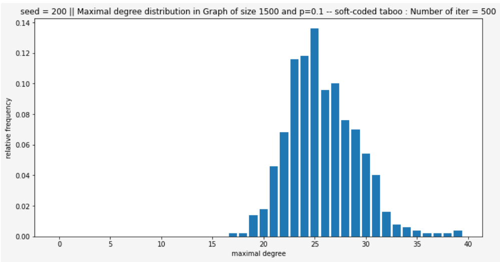
On the other hand, if we fix p, and vary $t \in \{500, 1000, 1500, 2000\}$, the distributions of the maximal degree from all the cases are very close.

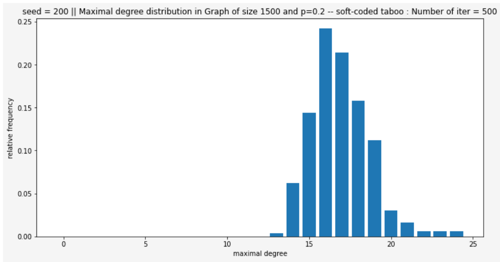
4.3.2 Soft-core taboo-ing

Like the other taboo-ing scheme, in this part too, we first fix $t = 1500$ and vary p in the set $\{0.01, 0.05, 0.1, 0.2\}$.









The following figure shows how the average maximal degree changes with p.

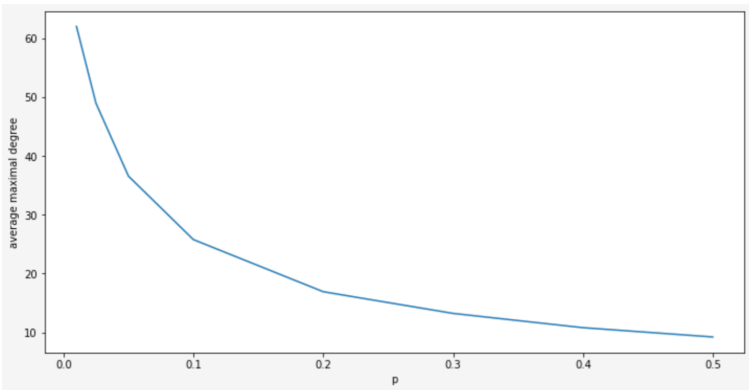
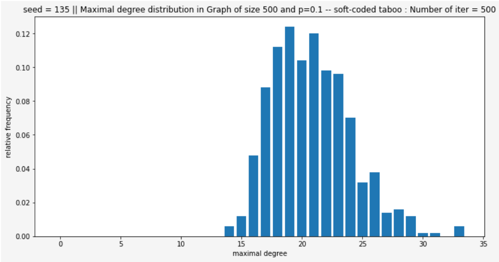
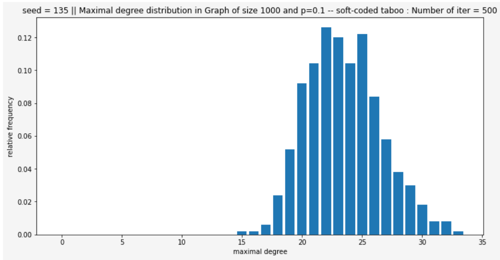
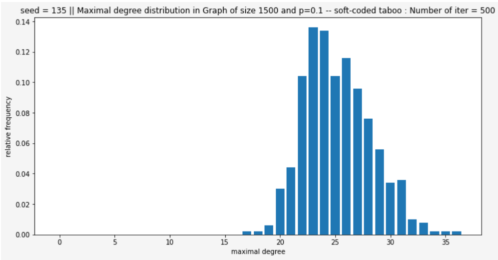


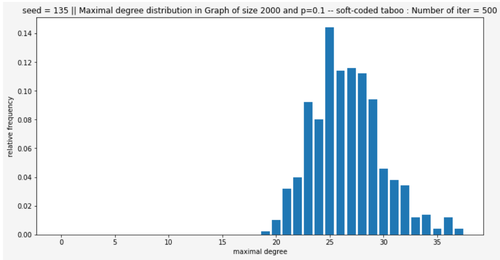
Figure 10: Average maximal degree as a function of p

In this case too, the average maximal degree decreases as p increases. For the other part, we fix $p = 0.1$ and vary $t \in \{500, 1000, 1500, 2000\}$. Following are the plots of the respective histograms.









Finally, observe how the average maximal degree changes with the increasing value of t. As expected, the average maximal degree increases with increasing t.

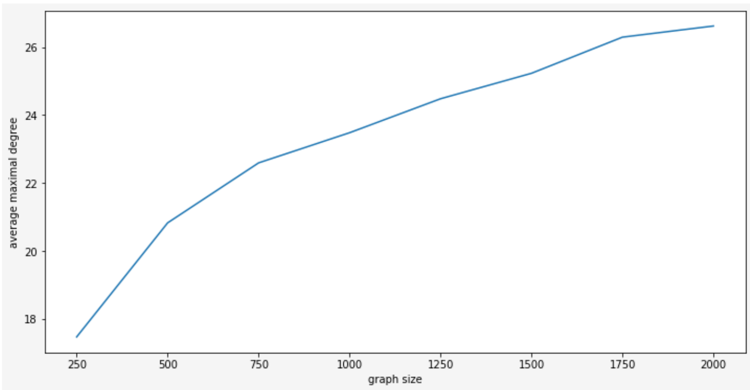


Figure 13: Average maximal degree as a function of t

5 A result on bound of maximal degree in hard-core taboo

As observed from the simulations for studying the distribution of maximal degree, in the random graph $(G_t^{(1,p)})_{t\geq 0}$ with hard-core taboo, the maximal degree is bounded above. We arrive at the following result after some investigation. In this context, the parameter p lies in the interval $(0,1)$.

Theorem 5.1. In hard-core taboo-ing scheme, the maximal degree in the random graph $(G_t^{(1,p)})_{t\geq 0}$ is bounded by $\lceil 1/p \rceil$ for all $t \geq 0$.

Proof. Observe that, since $m = 1$, the degree of a vertex can increase by at most 1 in unit time gap. Also note that, there exists some $t \geq 1$ such that $G_{t-1}^{(1,p)}$ has n vertices having the maximum degree $d = \lceil 1/p \rceil$ for some $n \in \mathbb{N}$. Then, only these n vertices can have maximum degree more than $\lceil 1/p \rceil$ in $G_t^{(1,p)}$.

Given the above information, we try to give a lower bound to the number of vertices in $G_{t-1}^{(1,p)}$. The first observation is that the graph is necessarily acyclic since $m = 1$. There are exactly n many d degree vertices in the graph. Each of these d degree vertices has d neighbours. Next, observe that all these n vertices can share at most $n-1$ of all these nd many neighbours because otherwise the graph would be cyclic. So, the number of vertices in the graph is bounded below by

$$n + nd - (n-1) = nd + 1.$$

Hence, the number of taboo vertices in $G_t^{(1,p)}$ would be bounded below by

$$\begin{aligned}
\lfloor p(nd+1) \rfloor &= \lfloor np\lceil 1/p \rceil + p \rfloor \\
&\geq \lfloor np\lceil 1/p \rceil \rfloor \\
&\geq \lfloor n \rfloor \\
&= n.
\end{aligned}$$

Thus, in $G_t^{(1,p)}$, more than or equal to n vertices are made, i.e., all the d degree vertices are made taboo since they were the highest degree vertices. Hence, no d degree vertex can increase its degree.

By similar argument, at any time point $t-1$, if the maximal degree in the graph is d, it cannot increase any further as all the maximal degree vertices will get taboo-ed. This completes the required proof. □

6 Future plans

We plan to study the asymptotic behaviours of fixed degree vertices $D_i(t)$ and the degree distribution $P_k(t)$ as t approaches $+\infty$. In this context, we shall restrict ourselves to soft-core taboo-ing and the special case $m = 1$.

6.1 Degree of fixed vertex

We aim to illustrate that $(D_i(t))_{t\geq 0}$ is a submartingale with respect to its natural filtration. Let us formally define the filtration $(\mathcal{F}_t)_{t\geq 0} \subset \mathcal{F}$ as below. Here, the relevant probability space is $(\Omega, \mathcal{F}, \mathbb{P})$.

$$\begin{aligned} \mathcal{F}_0 &= \sigma\left(\{D_i(0) \,:\, i = 1, 2\}\right), \\ \text{and, } \mathcal{F}_t &= \sigma\left(\{D_i(s) \,:\, 0 \leq s \leq t, 1 \leq i \leq s+2\}\right) \quad \forall t \geq 1. \end{aligned}$$

Then, for fixed $i \leq t+2$, the stochastic process $(D_i(t))_{t\geq 0}$ is $\mathcal{F}_t$ adapted. Let us now find the conditional expectation $\mathbb{E}\left(D_i(t+1) \,\middle|\, \mathcal{F}_t\right)$ for $i \leq t+2$.

$$\begin{aligned} \mathbb{E}(D_i(t+1) \,|\, \mathcal{F}_t) &= \mathbb{E}(D_i(t) + D_i(t+1) - D_i(t) \,|\, \mathcal{F}_t) \\ &= D_i(t) + \mathbb{E}(D_i(t+1) - D_i(t) \,|\, \mathcal{F}_t). \end{aligned} \tag{1}$$

Note that, since $m = 1$, $D_i(t)$ can increase by at most 1 or remain the same at time $t+1$. Hence, the conditional expectation $\mathbb{E}(D_i(t+1) - D_i(t) \,|\, \mathcal{F}_t)$ is the conditional probability that the newly arriving vertex at time $t+1$ attaches to the vertex i, given $\mathcal{F}_t$. Let us denote this conditional probability by $P_i(t)$. Observe that this probability can be seen as the probability that the i-th vertex is not taboo-ed times the conditional probability that the new vertex attaches to the i-th vertex. Here, for $n \in \mathbb{N}$, we shall use $[n]$ to denote the set $\{1, 2, \ldots, n\}$ for ease of notation. Thus,

$$\begin{aligned} P_i(t) = \sum_{\substack{k_1, k_2, \ldots, \\ k_{\lfloor (t+1)p \rfloor} \in [t+1] \setminus \{i\}}} \Bigg[& \frac{D_{k_1}(t) \ldots D_{k_{\lfloor (t+1)p \rfloor}}(t)}{\sum_{\substack{j_1, j_2, \ldots, \\ j_{\lfloor (t+1)p \rfloor} \in [t+1]}} D_{j_1}(t) \ldots D_{j_{\lfloor (t+1)p \rfloor}}(t)} \\ & \times \frac{D_i(t)}{\sum_{j=1}^{t+1} D_j(t) - \left\{ D_{k_1}(t) \ldots D_{k_{\lfloor (t+1)p \rfloor}}(t) \right\}} \Bigg]. \end{aligned} \tag{2}$$

Thus,

$$\mathbb{E}(D_i(t+1) \,|\, \mathcal{F}_t) = D_i(t) + P_i(t), \tag{3}$$

and since $P_i(t) \geq 0$, $(D_i(t), \mathcal{F}_t)$ is a submartingale. From here, we plan to consider the non-negative, hence convergent martingale $(D_i^*(t), \mathcal{F}_t)$ where

$$D_i^*(t) = \frac{D_i(t)}{\prod_{k=0}^{t-1} \{1 + P_i'(k)\}}. \tag{4}$$

6.2 Degree distribution

In this section, we fix $m = 1$ and try to write a recursion relation for the expected degree distribution. Instead of working with $P_k(t)$, which denotes the proportion of k degree vertices at time t, we define $N_k(t)$ which would denote the total number of k degree vertices at time t. In $G_t^{(1,p)}$, define

$$N_k(t) = \sum_{i=1}^{t+2} \mathbb{1}_{\{D_i(t)=k\}} = (t+2)P_k(t). \tag{5}$$

We want to investigate the asymptotic behaviour of $N_k(t)$ for fixed $k \geq 1$ as t approaches ∞. For that, we start by finding the conditional expectation $\mathbb{E}(N_k(t+1) \,|\, \mathcal{F}_t)$. For that, we shall adapt an argument similar to one given by Bollobás *et al* (2003) [6]. First observe that,

$$\mathbb{E}(N_k(t+1) \,|\, \mathcal{F}_t) = N_k(t) + \mathbb{E}(N_k(t+1) - N_k(t) \,|\, \mathcal{F}_t). \tag{6}$$

Now, conditional on $\mathcal{F}_t$, there are three ways $N_k(t+1) - N_k(t)$ can be non-zero. We list them as below.

(I) The half-edge of the new vertex attaches to a particular $k-1$ degree vertex. Then, $N_k(t)$ increases by 1 to $N_k(t+1)$. The probability of this event is $\frac{k-1}{\sum_{j=1}^{t+1} D_j(t)}$; and there are $N_{k-1}(t)$ many $k-1$ degree vertices to choose from.

(II) The half-edge of the new vertex attaches to a particular k degree vertex. Then, $N_k(t)$ decreases by 1 to $N_k(t+1)$. The probability of this event is $\frac{k}{\sum_{j=1}^{t+1} D_j(t)}$; and there are $N_k(t)$ many $k-1$ degree vertices to choose from.

(III) Another contribution to $k = 1$ arises from the newly arriving vertex itself. The newly arriving vertex at time $t+1$ has degree exactly 1, since the half edge attached to it is not allowed to form a self-loop.

Combining all of the above observations, we obtain

$$\mathbb{E}(N_k(t+1) - N_k(t) \,|\, \mathcal{F}_t) = \frac{(k-1)N_{k-1}(t)}{\sum_{j=1}^{t+1} D_j(t)} - \frac{kN_k(t)}{\sum_{j=1}^{t+1} D_j(t)} + \mathbb{1}_{\{k=1\}}. \tag{7}$$

Here, $k \geq 1$ and for $k = 0$, we define by convention $N_0(t) = 0$. Then, from equation 6 and by further taking expectation in the above equation 7 we get the following recursion involving

$\mathbb{E}(N_{k-1}(t)), \mathbb{E}(N_k(t))$ and $\mathbb{E}(N_k(t+1))$.

$$
\begin{aligned}
& \mathbb{E}(N_k(t+1)) = \mathbb{E}(N_k(t)) + \mathbb{E}(N_k(t+1) - N_k(t)) \\
\implies & \mathbb{E}(N_k(t+1)) = \mathbb{E}(N_k(t)) + \mathbb{E}(\mathbb{E}(N_k(t+1) - N_k(t) \,|\, \mathcal{F}_t)) \\
\implies & \mathbb{E}(N_k(t+1)) = \mathbb{E}(N_k(t)) + \frac{k-1}{\sum_{j=1}^{t+1} D_j(t)} \mathbb{E}(N_{k-1}(t)) - \frac{k}{\sum_{j=1}^{t+1} D_j(t)} \mathbb{E}(N_k(t)) + \mathbb{1}_{\{k=1\}}.
\end{aligned} \tag{8}
$$

The simple observation that $\sum_{j=1}^{t+1} D_j(t) = 2t$ yields some further simplification in notations.

$$
\mathbb{E}(N_k(t+1)) - \mathbb{E}(N_k(t)) = \frac{k\left[\mathbb{E}(N_{k-1}(t)) - \mathbb{E}(N_k(t))\right]}{2t} - \frac{\mathbb{E}(N_{k-1}(t))}{2t} + \mathbb{1}_{\{k=1\}}. \tag{9}
$$

We plan to further explore this recurrence relation 9 in order to study the asymptotic behaviour of $N_k(t)$ and $P_k(t)$.

References

[1] P. Erdős and A. Rényi. “On Random Graphs. I”. In: *Publicationes Mathematicae* 6 (1959), pp. 290–297.

[2] Béla Bollobás. *Random Graphs*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001. DOI: 10.1017/CBO9780511814068.

[3] Albert-László Barabási and Réka Albert. “Emergence of Scaling in Random Networks”. In: *Science* 286.5439 (1999), pp. 509–512. DOI: 10.1126/science.286.5439.509. eprint: https://www.science.org/doi/pdf/10.1126/science.286.5439.509. URL: https://www.science.org/doi/abs/10.1126/science.286.5439.509.

[4] Somak Laha. https://github.com/somak135/diss_rg/tree/main/mid-report. 2022.

[5] Remco van der Hofstad. *Random Graphs and Complex Networks*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016. DOI: 10.1017/9781316779422.

[6] Béla Bollobás et al. “Directed Scale-Free Graphs”. In: *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA ’03. Baltimore, Maryland: Society for Industrial and Applied Mathematics, 2003, pp. 132–139. ISBN: 0898715385.