Depreferential attachment model through preferential attachment

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

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

June 19, 2022

Table of contents

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. Similar research in this spirit can be found in [4]. The authors considered a model similar to the preferential attachment one, albeit the half-edge of the new vertex at time $t+1$ was attached to the smallest degree vertex among $d(d > 0)$ many which were preferentially chosen from the existing graph at time t, identically and with replacement. The asymptotic order of the maximal degree vertex was calculated. Another work [5] by Yuri Malyshkin considers attaching the half-edge of the new vertex with one of d possible neighbours, which are sampled with replacement from the set of the existing vertices with probability proportional to their degrees plus some parameter $\beta > -1$. Finally, the half-edge attaches to the vertex that have maximum degree among the chosen ones. The number of neighbours to choose in each time-step, i.e., the parameter d was taken to be random. It was proved that the for $\mathbb{E}d < 2+\beta$, the maximal degree had sublinear behaviour with a power law; for $\mathbb{E}d > 2+\beta$, it had linear behaviour and for $\mathbb{E}d = 2+\beta$, it was of order $n/\ln n$.

In our model, the novelty is that at each time step, we shall choose a few vertices preferentially, the number of which can possibly depend on the size of the existing graph. These vertices will be made taboo such that the half-edge of the new vertex do not attach to the taboo-ed vertices. Instead, the half-edge would attach to another preferentially chosen vertex, excluding the taboo-ed ones. The suggested model will be called *Depreferential attachment model through preferential attachment.*

1.1 Outline of the report

The report starts with an example of an ecosystem which we tried to model and was motivated to consider the Depreferential attachment model through preferential attachment. We then move on to the description of the model, its parameters and different taboo-ing schemes. Till the mid-term, we had only considered without replacement sampling for *soft-core* taboo-ing; but later introduced with replacement sampling too in order to simplify the mathematical analysis for asymptotics. After that, we describe those features of the model whose asymptotic distributions we are interested in and display some simulations in this regard. The simulations for with replacement sampling in *soft-core* taboo have been included after the mid-term. Next, there is a result, already shown in the mid-term, which gives an exact upper bound for the maximal degree in *hard-core* taboo-ing scheme. Finally, one section, included after the mid-term, is dedicated to the asymptotic distributions of two features of interest, *degree of fixed vertex* and *degree distribution*, considering with replacement sampling and fixed number of taboo-ing. Here, although we could not find the exact asymptotic order of the features, we did propose some upper and lower bound for the same. The report concludes with a few suggestions of future scope of research on this topic.

2 Motivation and description of the model

2.1 A relevant scenario

Consider modelling an ecosystem in which every newly arriving species preys on 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) **Fixed number taboo:** One possible case is that for each $t \geq 1$, we shall choose 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) **Fixed ratio taboo:** 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:** In this case, at time $t \geq 1$, $\lfloor (t+1)p_t \rfloor$ vertices can be chosen with or without replacement, where the probability of choosing the vertex v_j is directly proportional to its degree in $G_{t-1}^{(m,p_t)}$. Clearly, if the choices are made with replacement, the number of taboo vertices is possibly less than or equal to $\lfloor (t+1)p_t \rfloor$, while it would be exactly $\lfloor (t+1)p_t \rfloor$ in the case of without replacement sampling.

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

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

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 [6].

4.1 Degree of fixed vertex

Recall that, for $0 < i \leq t+1$, $D_i(t)$ 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 $t = 2000$.

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 [6].

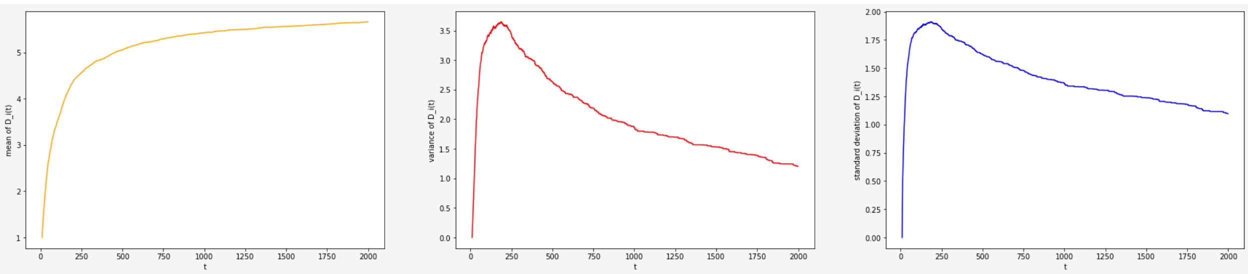


Figure 1: Mean, variance and standard deviation of $D_i(t)$ as function of t

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.

Soft-core taboo-ing

For this part, we have adopted the soft-core taboo-ing scheme. Following are the cases for $\boldsymbol{p = 0.1}$, and for both with and without replacement sampling. Plots for the other values of p are given in the GitHub repository [6].

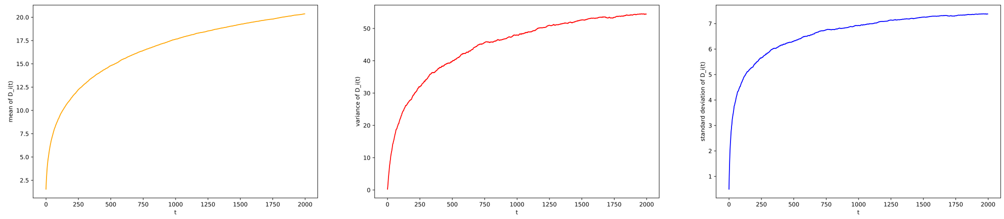


Figure 2: Mean, variance and standard deviation of $D_i(t)$ as function of t (With replacement sampling)

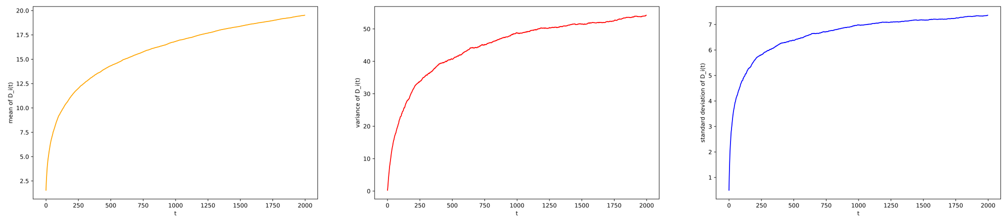


Figure 3: Mean, variance and standard deviation of $D_i(t)$ as function of t (Without replacement sampling)

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 $p \in \{0.01, 0.05, 0.1, 0.25\}$.

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 [6]. In the following plot, the variance is in 10^{-5} scale.

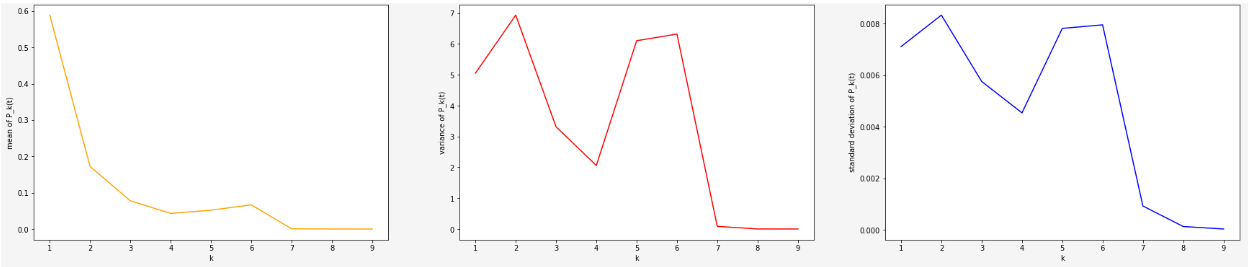


Figure 4: Mean, variance and standard deviation of $P_k(2000)$ as function of k

Soft-core taboo-ing

We have adopted the soft-core taboo-ing scheme for the following simulations. Following are the plots for $\boldsymbol{p = 0.1}$, and for both with and without replacement sampling. Plots for the other values of p are given in the GitHub repository [6]. In the following plot, the variance is in 10^{-5} scale.

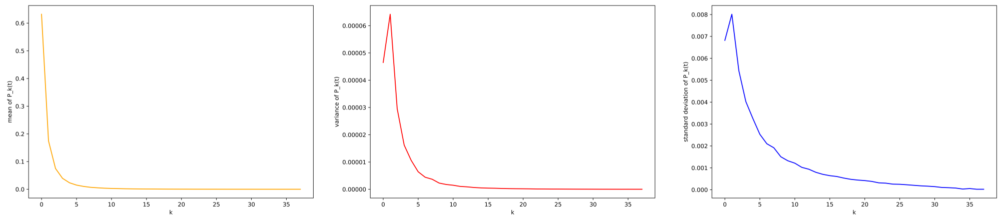


Figure 5: Mean, variance and standard deviation of $P_k(2000)$ as function of k(With replacement sampling)

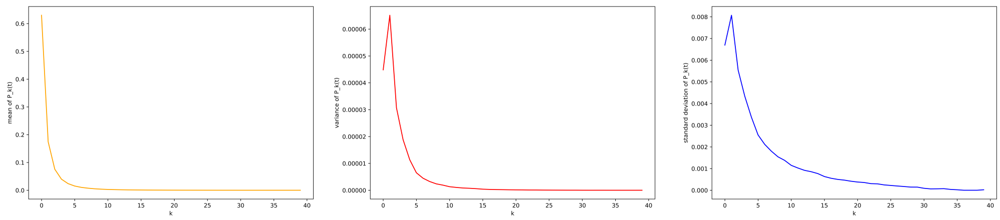


Figure 6: Mean, variance and standard deviation of $P_k(2000)$ as function of k(Without replacement sampling)

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* [7], which may be thought of as the case corresponding to $p = 0$.

4.3 Maximal degree

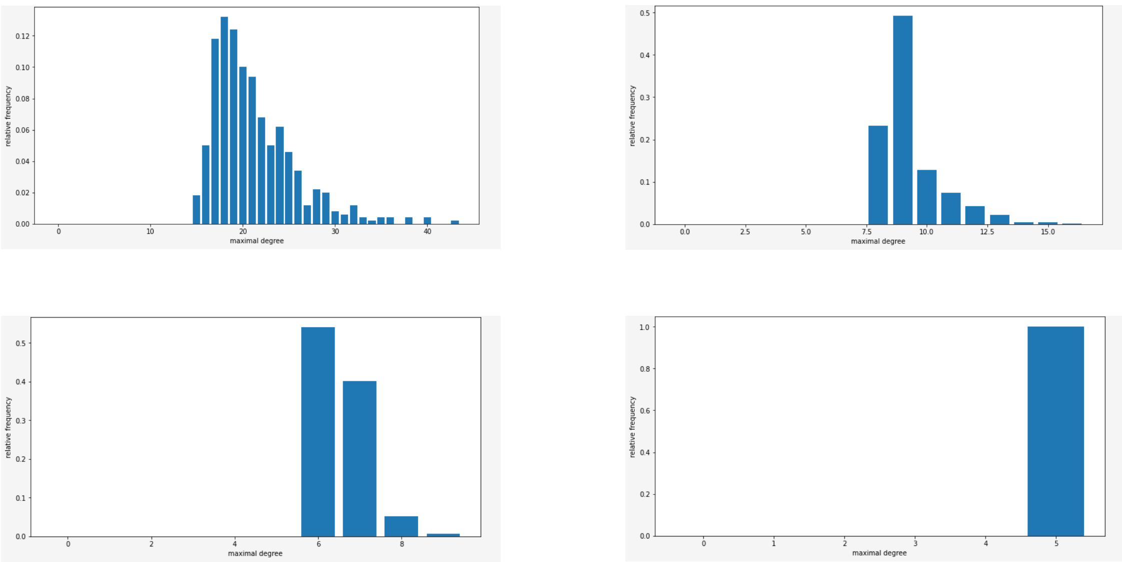
The maximal degree in $G_t^{(1,p)}$ is defined as

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

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 $t \in \{500, 1000, 1500, 2000\}$.

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.

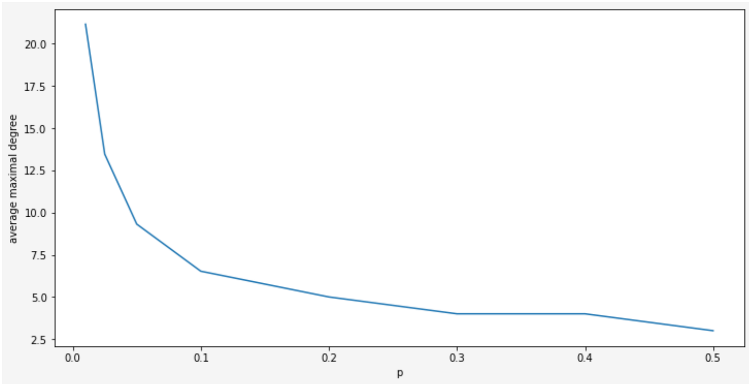


Figure 9: Average maximal degree as a function of p

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.

Soft-core taboo-ing

Like the other taboo-ing scheme, in this part too, we fix $t = 1500$ and vary p in the set $\{0.01, 0.05, 0.1, 0.2\}$. The simulations have been done for both with and without replacement sampling.

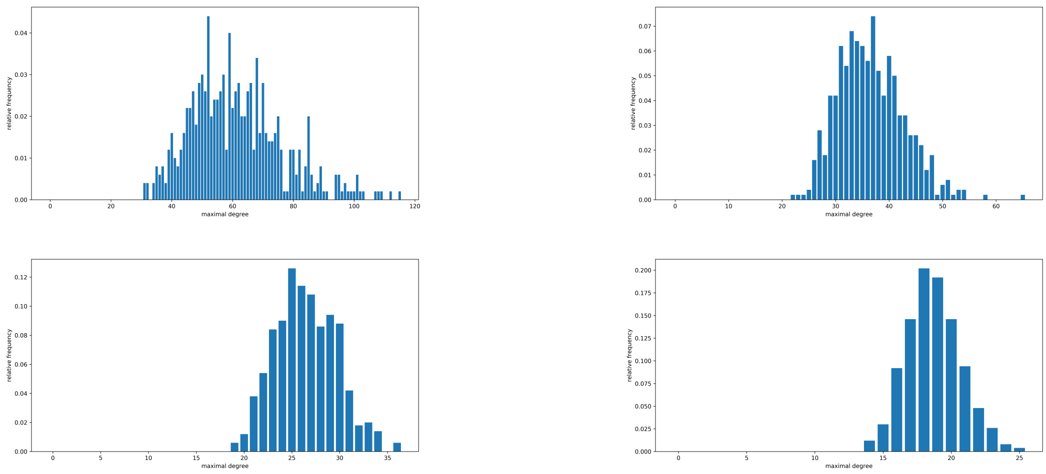


Figure 10: Histograms for $p \in \{0.01, 0.05, 0.1, 0.2\}$ respectively (With replacement sampling)

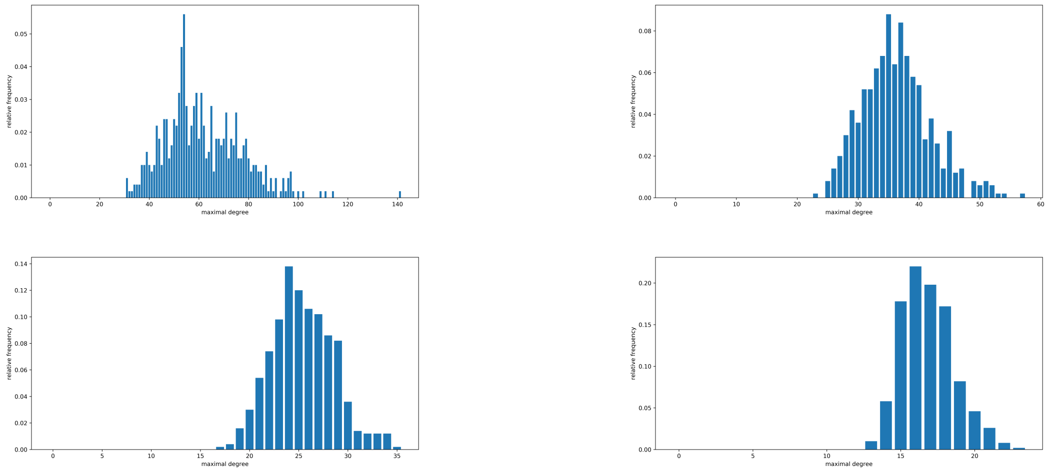


Figure 11: Histograms for $p \in \{0.01, 0.05, 0.1, 0.2\}$ respectively (Without replacement sampling)

The following figure shows how the average maximal degree changes with p, for both with and without sampling.

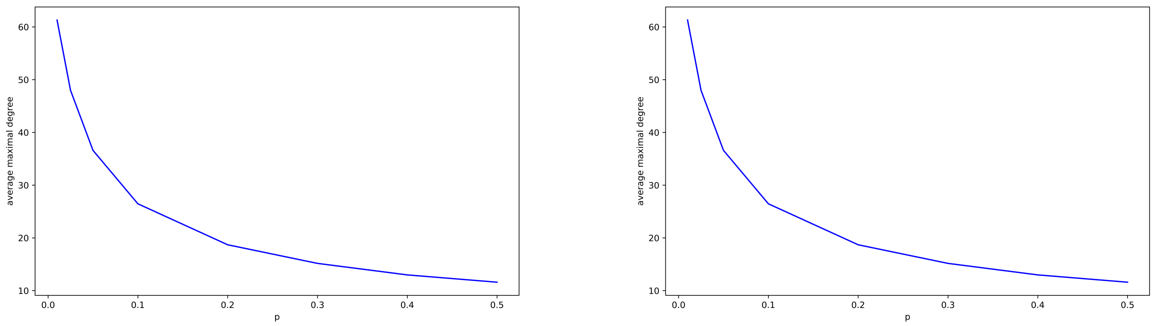


Figure 12: Average maximal degree as function of p for with replacement and without replacement sampling, respectively

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

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

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

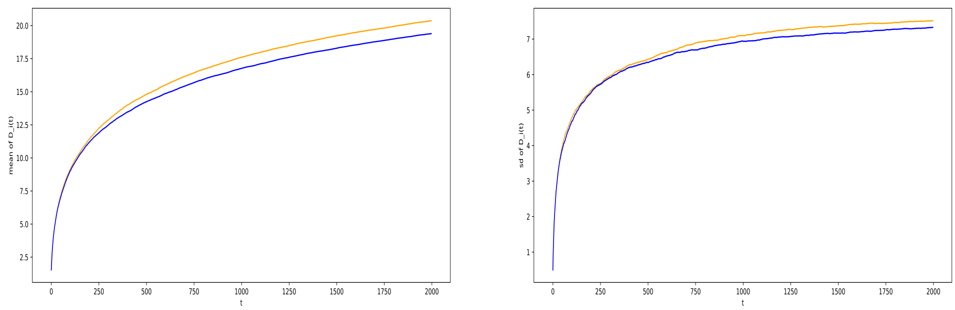
Thus, in $G_t^{(1,p)}$, more than or equal to n vertices are made, i.e., all the d degree vertices are made taboo since they were the highest degree vertices. Hence, no d degree vertex can increase its degree.

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. □

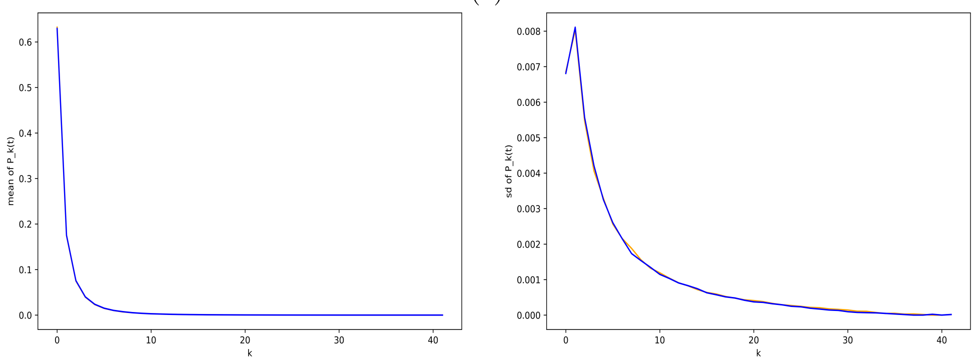
6 Bounds of asymptotic orders of two features of interest in soft-core taboo-ing scheme

We plan to find some upper and lower bounds of the asymptotic orders of degree of fixed 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 $m = 1$.

Also, in addition, we shall restrict to choosing the taboo vertices with replacement only. In fact, we tried the mathematical analysis for without replacement choosing scheme and found it to be analytically too much involved to be pursued in the interest of time. However, it is to be emphasized that both sampling schemes can be easily sampled and we did the sampling only to find that for fixed ratio taboo-ing, both with and without ratio sampling have possibly similar asymptotic behaviour of $D_i(t)$ and $P_k(t)$. The following plots show the mean and standard deviation of degree of fixed vertex $D_i(t)$ as function of t and degree distribution $P_k(t)$ as function of k, for with replacement sampling, in orange and without replacement sampling, in blue. In the latter pair of plots, the blue curves almost overlap the orange ones.



(a)



(b)

Figure 13: Final size of graph is $N = 2000$ in all of the $S = 1000$ samples, and graph parameter is $p = 0.1$. (a): Sample mean and standard deviation of $D_i(t)$ as function of t, (b): Sample mean and standard deviation of $P_k(t)$ in the full graph as function of k

6.1 Degree of fixed vertex

The main assumptions are that we shall choose fixed number d of already existing vertices preferentially and with replacement at each time step to make taboo. We shall first show 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 $(\Omega, \mathcal{F}, \mathbb{P})$.

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

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, for $t + 2 > d$, i.e., for $t > d - 2$,

$$P_i(t) = \sum_{\substack{k_1, k_2, \ldots, \\ k_d \in [t+2] \setminus \{i\}}} \left[\left(\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)} \right) \times \frac{D_i(t)}{2(t+1) - S_t(\{k_1, \ldots, k_d\})} \right], \tag{2}$$

where $S_t(A)$ denotes the sum of the degrees at time t of the unique vertices indexed by the set $A \subset [t+2]$. Observe that, $1 \leq S_t(\{k_1, \ldots, k_d\}) \leq \sum_{j=1}^{d} D_{k_j}(t)$.

On the other hand, for $0 \leq t \leq d-2$, there is no taboo-ing, and hence, $P_i(t) = D_i(t)/2(t+1)$. So, we obtain

$$\mathbb{E}(D_i(t+1) \,|\, \mathcal{F}_t) = D_i(t) + P_i(t) = D_i(t)\{1 + P_i'(t)\}, \tag{3}$$

and since $P_i(t) \geq 0$, $(D_i(t), \mathcal{F}_t)$ is a submartingale. From here, we use the notation $P_i'(k) = P_i(k)/D_i(k)$ and we plan to consider the non-negative, hence convergent martingale $(D_i^*(t), \mathcal{F}_t)$ where

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

We tried to provide some upper and lower bounds of the denominator $\prod_{k=i-1}^{t-1}\{1 + P_i'(k)\}$ and arrived to the following theorem that gives upper and lower bounds to the asymptotic order of $D_i(t)$.

Theorem 6.1. Fix a non-negative integer i and let $(D_i(t))_{t\geq 0}$ be the degree of the i-th vertex in the random graph process $(G_t^{(1,p_t)})_{t\geq 0}$ which admits fixed number $(d \geq 0)$ of with replacement soft-core taboo-ing. Then,

$$t^{2^{-d-1}}\zeta_i' \leq D_i(t) \leq t\zeta_i, \quad \text{for all } t \geq 0,$$

where ζ_i' and ζ_i are two random variables.

6.1.1 Proof of the upper bound of asymptotic order of degree of a fixed vertex

Let us try to find some upper bound on the term $1+P_i'(t)$ for $t > d-2$. Observe that, for $t > d-2$,

$$\begin{aligned}
1 + P_i'(t) &= 1 + \sum_{\substack{k_1,k_2,\ldots,\\ k_d\in[t+2]\setminus\{i\}}} \left[\left(\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)}\right) \times \frac{1}{2(t+1) - S_t(\{k_1,\ldots,k_d\})}\right] \\
&\leq 1 + \sum_{\substack{k_1,k_2,\ldots,\\ k_d\in[t+2]\setminus\{i\}}} \left[\left(\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)}\right) \times \frac{1}{t+2-d}\right] \\
&= 1 + \frac{1}{2^d(t+1)^d} \sum_{\substack{k_1,k_2,\ldots,\\ k_d\in[t+2]\setminus\{i\}}} \left[\left(\prod_{j=1}^{d} D_{k_j}(t)\right) \times \frac{1}{t+2-d}\right],
\end{aligned}$$

by noting that the sum of the degrees of the non-taboo vertices, i.e., $2(t+1)-S_t(\{k_1,\ldots,k_d\})$ can be bounded below by $t+2-d$, which is merely the number of non-taboo vertices. Let us now suppose, $X_1,\ldots,X_d$ are independent and identical samples from the distribution Uniform$(\{1,2,\ldots,t+2\}\setminus\{i\})$. Also, assume the function f is defined such that $f(i) = D_i(t)$. Now, we shall replace the sum in the above equation by an expectation over the i.i.d. random variables $X_1,\ldots,X_d$. Then, we proceed as

$$\begin{aligned}
1 + P_i'(t) &\leq 1 + \frac{(t+1)^d}{2^d(t+1)^d}\mathbb{E}\left[\left(\prod_{j=1}^{d} f(X_j)\right) \times \frac{1}{t+2-d}\right] \\
&= 1 + \frac{1}{2^d(t+2-d)}\left(\mathbb{E}\left[f(X_1)\right]\right)^d.
\end{aligned}$$

Since X_1 is chosen uniformly from the set $[t+2] \setminus \{i\}$, we can write $\mathbb{E}[f(X_1)]$ as

$$\mathbb{E}[f(X_1)] = \frac{2(t+1) - D_i(t)}{t+1} \leq \frac{2(t+1)}{t+1} = 2.$$

Then, we get the upper bound of $1 + P_i'(t)$ to be

$$1 + P_i'(t) \leq 1 + \frac{1}{t+2-d} = \frac{t+3-d}{t+2-d}.$$

Now, observe that from the equation 4, we can write

$$\begin{aligned} D_i(t) &= D_i^*(t) \times \prod_{k=i-1}^{t-1} \{1 + P_i'(k)\} \\ &\leq c_{i,d} D_i^*(t) \times \frac{\Gamma(t+3-d)}{\Gamma(t+2-d)}, \end{aligned}$$

where $c_{i,d}$ is some finite constant dependent on i and d, which too are finite. At this point, we use the result

$$\frac{\Gamma(t+a)}{\Gamma(t)} \to t^a(1 + O(1/t)), \quad \text{as } t \to \infty$$

to obtain that $D_i(t)/t \leq \zeta_i$ for large enough t where ζ_i is a random variable bounded in probability. Since the ratio $D_i(t)/t$ is itself a finite random variable for finite i and t, this concludes the result for all $t \geq 0$.

Remark 6.2. However, since the sum of the degrees of the vertices at time t is $2(t+1)$, it can be trivially seen that in fact, $\zeta_i \leq 2$. So, we shall take resort to an appealing ansatz that states,

$$\frac{S_t(\{k_1, \ldots, k_d\})}{2(t+1)} \xrightarrow{a.s.} 0, \quad \text{as } t \to \infty$$

where $k_1, \ldots, k_d$ are the indices of the taboo vertices at time t. Using this assumption yields, for large enough t,

$$\begin{aligned}
1 + P_i'(t) &= 1 + \sum_{\substack{k_1,k_2,\ldots,\\ k_d \in [t+2]\setminus\{i\}}} \left[\left(\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)} \right) \times \frac{1}{2(t+1) - S_t(\{k_1,\ldots,k_d\})} \right] \\
&= 1 + \frac{1}{2^{d+1}(t+1)^{d+1}} \sum_{\substack{k_1,k_2,\ldots,\\ k_d \in [t+2]\setminus\{i\}}} \left[\left(\prod_{j=1}^{d} D_{k_j}(t) \right) \times \frac{1}{1 - S_t(\{k_1,\ldots,k_d\})/2(t+1)} \right] \\
&\approx 1 + \frac{(t+1)^d}{2^{d+1}(t+1)^{d+1}} \mathbb{E}\left[\prod_{j=1}^{d} f(X_j) \right] \\
&= 1 + \frac{1}{2^{d+1}(t+1)} \left(\mathbb{E}\left[f(X_1) \right] \right)^d \\
&= 1 + \frac{1}{2^{d+1}(t+1)} \left(\frac{2(t+1) - D_i(t)}{t+1} \right)^d \\
&\leq 1 + \frac{1}{2^{d+1}(t+1)} \left(\frac{2(t+1)}{t+1} \right)^d \\
&= 1 + \frac{1}{2t+2} \\
&= \frac{2t+3}{2t+2} = \frac{t+3/2}{t+1}.
\end{aligned}$$

Then, again from equation 4, we obtain that

$$\begin{aligned}
D_i(t) &= D_i^*(t) \times \prod_{k=i-1}^{t-1} \{1 + P_i'(k)\} \\
&\leq c_{i,d}^{'} D_i^*(t) \times \frac{\Gamma(t+3/2)}{\Gamma(t+1)},
\end{aligned}$$

which implies $D_i(t)/\sqrt{t} \leq \zeta_i'$ where ζ_i' is a random variable bounded in probability. This upper bound for $D_i(t)$ is tighter and in fact the same as the asymptotic order of degree of fixed vertex in preferential attachment model [3]. This bound is reliable and is supported by simulations. However, a formal proof would require a proof of the ansatz and why bounding the tail of the product $\prod_{k=i-1}^{t-1}\{1 + P_i'(k)\}$ suffices.

6.1.2 Proof of the lower bound of asymptotic order of degree of a fixed vertex

Let us now try to find some lower bound on the term $1 + P_i'(t)$ for $t > d - 2$. Observe that, for $t > d - 2$,

$$\begin{aligned}
1 + P_i'(t) &= 1 + \sum_{\substack{k_1,k_2,\ldots,\\ k_d \in [t+2]\setminus\{i\}}} \left[\left(\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)} \right) \times \frac{1}{2(t+1) - S_t(\{k_1,\ldots,k_d\})} \right] \\
&\geq 1 + \sum_{\substack{k_1,k_2,\ldots,\\ k_d \in [t+2]\setminus\{i\}}} \left[\left(\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)} \right) \times \frac{1}{2(t+1) - 1} \right],
\end{aligned}$$

by noting that the sum of the degrees of the taboo vertices, i.e., $S_t(\{k_1,\ldots,k_d\})$ can be bounded below by 1. Then, continuing,

$$\begin{aligned}
1 + P_i'(t) &\geq 1 + \frac{(t+1)^d}{2^d(t+1)^d} \mathbb{E}\left[\left(\prod_{j=1}^{d} f(X_j) \right) \times \frac{1}{2t+1} \right] \\
&= 1 + \frac{1}{2^d(2t+1)} \left(\mathbb{E}\left[f(X_1) \right] \right)^d \\
&= 1 + \frac{1}{2^d(2t+1)} \left(\frac{2(t+1) - D_i(t)}{t+1} \right)^d \\
&\geq 1 + \frac{1}{2^d(2t+1)} \\
&= \frac{2^d(2t+1) + 1}{2^d(2t+1)} = \frac{t + (1/2 + 2^{-d-1})}{t + 1/2}.
\end{aligned}$$

Now, observe that from the equation 4, we can write

$$\begin{aligned}
D_i(t) &= D_i^*(t) \times \prod_{k=i-1}^{t-1} \{1 + P_i'(k)\} \\
&\geq b_{i,d} D_i^*(t) \times \frac{\Gamma(t + 1/2 + 2^{-d-1})}{\Gamma(t + 1/2)},
\end{aligned}$$

where $b_{i,d}$ is some finite constant dependent on i and d, which too are finite. Again, we use the result

$$\frac{\Gamma(t+a)}{\Gamma(t)} \to t^a(1 + O(1/t)), \quad \text{as } t \to \infty$$

to obtain that $D_i(t)/t^{2^{-d-1}} \geq \zeta_i'$ for all large enough t where ζ_i' is a random variable bounded in probability. We can again argue that since the ratio $D_i(t)/t^{2^{-d-1}}$ is a finite random variable for finite i and t, this concludes the proof for all $t \geq 0$. It is to be noted that for $d = 0$, the upper bound suggested in remark 6.2 and the lower bound for the asymptotic order of $D_i(t)$ coincides.

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. We 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) [8]. 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

$$\sum_{\substack{1 \leq i \leq t+2: \\ D_i(t)=k-1}} \sum_{\substack{k_1,k_2,\dots, \\ k_d \in [t+2]\setminus\{i\}}} \left\{ \prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)} \right\} \frac{D_i(t)}{2(t+1) - S_t(\{k_1,\dots,k_d\})}$$
$$= \sum_{\substack{k_1,k_2,\dots, \\ k_d \in [t+2]\setminus\{i_0\}}} \left\{ \prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)} \right\} \frac{(k-1)N_{k-1}(t)}{2(t+1) - S_t(\{k_1,\dots,k_d\})} = (k-1)N_{k-1}(t)P_{i_0}(t),$$

where i_0 is a particular $k-1$ degree vertex at time t.

(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

$$\sum_{\substack{1\le i\le t+2:\\ D_i(t)=k}} \sum_{\substack{k_1,k_2,\dots,\\ k_d\in[t+2]\setminus\{i\}}} \left\{\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)}\right\} \frac{D_i(t)}{2(t+1)-S_t(\{k_1,\dots,k_d\})}$$
$$= \sum_{\substack{k_1,k_2,\dots,\\ k_d\in[t+2]\setminus\{j_0\}}} \left\{\prod_{j=1}^{d} \frac{D_{k_j}(t)}{2(t+1)}\right\} \frac{(k)N_k(t)}{2(t+1)-S_t(\{k_1,\dots,k_d\})} = kN_k(t)P_{j_0}(t),$$

where j_0 is a particular k degree vertex at time t.

(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

$$\mathbb{E}(N_k(t+1)-N_k(t)\,\big|\,\mathcal{F}_t) = (k-1)N_{k-1}(t)P'_{i_0}(t) - kN_k(t)P'_{j_0}(t) + \mathbb{1}_{\{k=1\}}. \tag{7}$$

We assume $N_k(t)/t$ converges to $\mathbb{E}N_k(t)/t$ in probability and further assume that $\mathbb{E}(N_k(t)) \approx tp_k$ for large t and hence $\mathbb{E}(N_k(t+1)-N_k(t)) \approx p_k$ for large t. Here, $k\ge 1$ and for $k=0$, we define by convention $N_0(t)=0$. Then, taking repeated expectations, we obtain the following recursion in p_k from the above equation 7.

$$p_k = \mathbb{1}_{\{k=1\}} + (k-1)tp_{k-1}P'_{i_0}(t) - ktp_kP'_{j_0}(t). \tag{8}$$

Now, we know an upper and lower bound for the quantities $P'_{i_0}(t)$ and $P'_{j_0}(t)$ from the calculation done for the asymptotics of degree of fixed vertex. In fact,

$$\frac{1}{2^d(2t+1)} \le P'_{i_0}(t), P'_{j_0}(t) \le \frac{1}{t+2-d}.$$

Substituting these values immediately yield

$$\begin{aligned} &\mathbb{1}_{\{k=1\}} + \frac{(k-1)tp_{k-1}}{2^d(2t+1)} - \frac{ktp_k}{t+2-d} \le p_k \le \mathbb{1}_{\{k=1\}} + \frac{(k-1)tp_{k-1}}{t+2-d} - \frac{ktp_k}{2^d(2t+1)} \\ \implies &\mathbb{1}_{\{k=1\}} + \frac{(k-1)p_{k-1}}{2^{d+1}} - kp_k \le p_k \le \mathbb{1}_{\{k=1\}} + (k-1)p_{k-1} - \frac{kp_k}{2^{d+1}}, \end{aligned} \tag{9}$$

where the second step follows since the inequality being true for all large enough t, we can have t approach ∞ and take limits on all sides. Some careful treatment of the above recursive inequality gives us upper and lower bounds of p_k, the asymptotic proportion of k degree vertices in the graph $(G_t^{(1,p_t)})_{t\geq 0}$.

Theorem 6.3. In the random graph process $(G_t^{(1,p_t)})_{t\geq 0}$, which admits fixed number($d \geq 0$) of with replacement soft-core taboo-ing, let p_k denote the asymptotic proportion of k degree vertices. Then p_k can be bounded on both sides as

$$\frac{\Gamma(k)}{2^{(k-1)(d+1)}\Gamma(k+2)} \leq p_k \leq \frac{2^{k(d+1)}\Gamma(k)\Gamma(2^{d+1}+1)}{\Gamma(2^{d+1}+k+1)}, \quad \text{for all integers } k \geq 1.$$

6.2.1 Proof of the upper bound of asymptotic order of degree distribution

Let us start with $k = 1$ and the inequality $p_k \leq \mathbb{1}_{\{k=1\}} + (k-1)p_{k-1} - \frac{kp_k}{2^{d+1}}$. Then, we shall get by substituting $k = 1$, and recalling $p_0 = 0$,

$$\begin{aligned}
&p_1 \leq 1 - \frac{p_1}{2^{d+1}} \\
\implies &p_1 \cdot \frac{2^{d+1}+1}{2^{d+1}} \leq 1 \\
\implies &p_1 \leq \frac{2^{d+1}}{2^{d+1}+1}.
\end{aligned}$$

Now, from the same inequality, we inductively get for $k \geq 2$,

$$\begin{aligned}
&p_k \leq \mathbb{1}_{\{k=1\}} + (k-1)p_{k-1} - \frac{kp_k}{2^{d+1}} \\
\implies &p_k \cdot \frac{2^{d+1}+k}{2^{d+1}} \leq (k-1)p_{k-1} \\
\implies &p_k \leq \frac{2^{d+1}(k-1)}{2^{d+1}+k}p_{k-1} \\
\implies &p_k \leq \frac{2^{(k-1)(d+1)}(k-1)(k-2)\dots 1}{(2^{d+1}+k)(2^{d+1}+k-1)\dots(2^{d+1}+2)}p_1 \\
\implies &p_k \leq \frac{2^{k(d+1)}\Gamma(k)\Gamma(2^{d+1}+1)}{\Gamma(2^{d+1}+k+1)}.
\end{aligned}$$

6.2.2 Proof of the lower bound of asymptotic order of degree distribution

Let us start with $k = 1$ and the inequality $p_k \geq \mathbb{1}_{\{k=1\}} + \frac{(k-1)p_{k-1}}{2^{d+1}} - kp_k$. Then, we shall get by substituting $k = 1$, and recalling $p_0 = 0$,

$$
\begin{aligned}
& p_1 \geq 1 - p_1 \\
\implies & p_1 \geq 1/2.
\end{aligned}
$$

Now, from the same inequality, we inductively get for $k \geq 2$,

$$
\begin{aligned}
& p_k \geq \frac{(k-1)p_{k-1}}{2^{d+1}} - kp_k \\
\implies & (k+1)p_k \geq \frac{k-1}{2^{d+1}} p_{k-1} \\
\implies & p_k \geq \frac{(k-1)\ldots 1}{2^{(k-1)(d+1)}(k+1)\ldots 3} p_1 \\
\implies & p_k \geq \frac{\Gamma(k)}{2^{(k-1)(d+1)}} \\
\implies & p_k \geq \frac{\Gamma(k)}{2^{(k-1)(d+1)}\Gamma(k+2)}.
\end{aligned}
$$

This concludes the upper and lower bound of p_k, the asymptotic degree distribution.

Remark 6.4. If we take resort to the appealing ansatz referred to at Remark 6.2, we get a tighter bound of $P_i'(t)$ and thus we can find a tighter bound of p_k, given by,

$$
\frac{4\Gamma(k)}{2^{(k-1)d}\Gamma(k+3)} \leq p_k \leq \frac{2^{kd+1}\Gamma(k)\Gamma(2^{d+1}+1)}{\Gamma(2^{d+1}+k+1)}.
$$

7 Future prospects

In this report, we have found upper and lower bounds for asymptotic orders of degree of fixed vertex and degree distribution, particularly for fixed number of taboo-ing with replacement. It may be of interest to know how these features behave asymptotically in case of fixed ratio of taboo-ing, possibly without replacement. In fixed ratio taboo-ing regime, we expect the *rich gets richer* phenomena, as described in the introduction, to get even more curbed as the number of preferentially chosen taboo vertices increase linearly with time. Also, since this report only contains bounds of asymptotic orders of the features, finding the exact asymptotic orders may be another research interest.

We have not delved into the mathematical analysis of maximal degree in case of soft-core taboo-ing in this report. The maximal degree is indeed an important feature of interest as illustrated for different models in [3, 4, 5, 7]. There is scope for research in this regard for our model.

Finally, we conclude by conjecturing that for fixed number of soft-core taboo-ing, either with or without replacement, all the three features of interest, i.e., degree of fixed vertex, degree distribution and maximal degree have same respective asymptotic orders as the preferential attachment model [3]. The remark 6.2, in fact, result into an asymptotic upper bound of $O(\sqrt{t})$ for $D_i(t)$, which happen to be the exact asymptotic order of the same for preferential attachment model [7]. The same remark also results into an lower bound for p_k, which coincides with the exact value of p_k in preferential attachment model [7] when we substitute $d = 0$.

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] Yury Malyshkin and Elliot Paquette. "The power of choice over preferential attachment". In: *ALEA, Lat. Am. J. Probab. Math. Stat.* 12(2) (2015), pp. 903–915.

[5] Yury Malyshkin. *Preferential attachment combined with random number of choices*. 2016. DOI: 10.48550/ARXIV.1612.02229. URL: https://arxiv.org/abs/1612.02229.

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

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

[8] 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.