PROTEAN GRAPHS WITH A VARIETY OF RANKING SCHEMES

JEANNETTE JANSSEN AND PAWEŁ PRAŁAT

Abstract. We introduce a new class of random graph models for complex real-world networks, based on the protean graph model by Łuczak and Prałat. Our generalized protean graph models have two distinguishing features. First, they are not growth models, but instead are based on the assumption that a "steady state" of large but finite size has been reached. Second, the models assume that the vertices are ranked according to a given ranking scheme, and the rank of a vertex determines the probability that that vertex receives a link in a given time step. Precisely, the link probability is proportional to the rank raised to the power $-\alpha$, where the attachment strength α is a tunable parameter. We show that the model leads to a power law degree distribution with exponent $1 + 1/\alpha$ for ranking schemes based on a given prestige label, or on the degree of a vertex. We also study a scheme where each vertex receives an initial rank chosen randomly according to a biased distribution. In this case, the degree distribution depends on the distribution of the initial rank. For one particular choice of parameters we obtain a power law with an exponent that depends both on α and on a parameter determining the initial rank distribution.

1. Introduction

There is considerable interest in using random graphs to model complex real-world networks in order to gain insight into their properties (see for example [1] or [3]). Most prevalent models are based on the principle of *preferential attachment*: new vertices link to existing vertices with a probability that is proportional to the degree of the existing vertex. The preferential attachment principle has been successful in explaining the power law degree distribution that has been observed in many real-life networks. On the other hand, it is hard to adapt the principle to incorporate more diverse criteria that make a vertex attractive to receive a link, such as innate popularity or initial advantage (see [2]). Moreover, most models are *growth models*, where the graphs grow larger over time. In many real-world networks, such as protein-protein interaction networks, social networks, and even the World Wide Web, a more realistic assumption seems to be that the network will eventually reach a "steady state" where the size stays approximately constant, but vertices keep appearing and disappearing over time.

In [7], Łuczak and Prałat introduced a random graph model called the protean graph $\mathcal{P}_n(d, \alpha)$, where the model is controlled by two parameters: initial degree

1991 *Mathematics Subject Classification.* Primary: 05C80. Secondary: 05C07.

Key words and phrases. random graphs, web graphs, protean graphs, degree distribution, differential equations method, power law graphs, scale-free networks.

The authors are supported by MITACS and NSERC. This work is part of the MITACS project Modelling and Mining of Networked Information Spaces (MoMiNIS).

$d \in \mathbb{N}$ and attachment strength $0 < \alpha < 1$). The major feature of this model is that the *link probability*, that is, the probability that a vertex receives a link in a given time step, is based on a ranking of the vertices. More precisely, each vertex v has a rank $r(v,t)$ at time t, and its link probability is proportional to $r(v,t)^{-\alpha}$. In the original protean graph model, vertices were ranked according to age, leading to a model where the old get richer. In the paper it was proven that the degrees in $\mathcal{P}_n(d,\alpha)$ are distributed according to a power law with exponent $1+1/\alpha$. This implies that the attachment strength α can be used to tune the exponent of the power law: In order to establish the right attachment strength to model a given real-life network where the number of vertices of degree k is approximately $k^{-\gamma}$, the attachment strength should be chosen to be $\alpha = 1/(\gamma - 1)$.

In [7], the behaviour of the protean graph model near the connectivity threshold is studied, where $d = d(n)$ is allowed to grow with n. The second author of this paper showed also in [9] that the protean graph $\mathcal{P}_n(d,\alpha)$ asymptotically almost surely (*aas*) has one giant component, containing a positive fraction of all vertices. The diameter of this component is equal to $\Theta(\log n)$. In [10], the *recovery time* of certain connectivity properties was studied. Suppose that the protean graph *aas* possesses property P, but at some particular time instance, the graph looses property P. Then the following natural question can be asked: how much time does it take for the protean graph to regain its typical property P? Since P holds *aas* and after $O(n \log n)$ steps each vertex from the original graph is deleted *aas*, the recovery time is *aas* bounded from above by $O(n \log n)$. It has been shown, however, that the recovery time for connectivity is smaller than the above universal bound implied by the coupon collector problem.

In this paper, we extend the idea of the protean graph to include variations where vertices are ranked according to other criteria than age. The general approach of using a link probability based on a ranking of the vertices according to degree or an externally determined prestige label was first proposed by Fortunato, Flammini and Menczer in [4]. The specific model used by Fortunato *et al.* was a growth model, where one new vertex is added in every time step, and no vertices disappear. The occurrence of a power law with exponent $1+1/\alpha$ was postulated based on simulations. The authors of this paper provided rigorous proofs and extended their results in [6]. A growth model based on the original protean graph, where vertices are ranked by age, was proposed in [12]. Here new vertices are added at a faster rate than old vertices are deleted. A prelimary version of this paper appeared in [11].

As we will show, the protean graph model leads to power law graphs for a variety of different ranking schemes. The ranking scheme of the original protean graph from [7, 9] ranks vertices by age (*the old get richer*); as we already mentioned, this leads to a power law with exponent $1+1/\alpha$. As a contrast, here we also consider the case where vertices are ranked inversely according to age (*the young get richer*). As suspected, the young are not young long enough to accumulate a lot of wealth, and we find that in the resulting graph, *aas* all vertices have degree bounded by $\log^2 n$.

We also study a ranking scheme where each vertex receives an independently chosen prestige label, and vertices are ranked according to their prestige label.

Here we do obtain a power law, with the same exponent. In order to allow for a non-uniform distribution of "prestige" over the vertices, we considered also a scheme based on *random initial rank*. Here, each vertex is assigned an initial rank according to a given distribution. We consider distributions of the following form. Let R_i be the initial rank of a vertex born at time i. Then $\mathbb{P}(R_i \leq k) = (k/n)^s$. Thus, when $s = 1$ the initial rank is chosen according to the uniform distribution. In this case, the behaviour is very similar to that of ranking by by prestige label: vertices with initial rank R_i exhibit behaviour as if they had received fitness R_i/n. When $s > 1$, so the initial rank of new vertices is biased towards the lower ranks (highly ranked vertices tend to be older). In this case, the behaviour is similar to that of ranking by age, and we obtain a power law degree distribution with exponent $1 + 1/\alpha$. If $0 < s < 1$, so the initial rank is biased towards the higher ranks (young vertices tend to be ranked high), the behaviour is more complex, and depends on both s and α. If $0 < s \leq 1 - \alpha$, the behaviour is like ranking by inverse age, and we have a "flat" degree distribution with maximum degree bounded by $\log^2 n$. If $1 - \alpha < s < 1$, then we do obtain a power law, but with exponent $1 + s/(s + \alpha - 1)$.

These results suggest a broader explanation for the power law degree distributions often observed in real-life networks such as the web graph, protein interaction networks, and social networks, even when they have reached a stage where their size does not grow significantly. The distribution of new links in such networks can be seen as governed by a rank-based attachment scheme, based on a ranking scheme that can be derived from a number of different factors such as age, degree, or fitness. The exponent of the power law is independent of these factors, but is rather a consequence of the attachment strength. In addition, rank-based attachment accentuates the difference between higher ranked vertices: the difference in link probability between the vertices ranked 1 and 2 is much larger than that between the vertices ranked 100 and 101. This again corresponds to our intuition of what constitutes a credible mechanism for link attachment.

2. Definitions

In this section, we formally define the graph generation model based on rank-based attachment which will lead to the limiting protean graph. The model produces a sequence $\{G_t\}_{t=0}^{\infty} = \{(V_t, E_t)\}_{t=0}^{\infty}$ of undirected graphs on n vertices, where t denotes time. Our model has two fixed parameters: initial degree $d \in \mathbb{N}$ and attachment strength $\alpha \in (0, 1)$. At each time t, each vertex $v \in V_t$ has rank $r(v, t) \in [n]$ (we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$). In order to obtain a proper ranking, the rank function $r(\cdot, t) : V_t \to [n]$ is a bijection for all t, so every vertex has a unique rank. In agreement with the common use of the word "rank", high rank refers to a vertex v for which $r(v, t)$ is small: the highest ranked vertex is ranked number one, so has rank equal to 1; the lowest ranked vertex has rank n. The probability that v receives an edge is proportional to $r(v, t)^{-\alpha}$; the negative exponent guarantees that vertices with higher ranks ($r(v, t)$ close to 1) are more likely to receive new edges than lower ranks. The initialization and update of the ranking is done according to a *ranking scheme*. Various ranking schemes can be considered, and will lead to different protean graphs. We first give the general model, and then list the ranking schemes.

To initialize the model, let $G_0 = (V_0, E_0)$ be any graph on n vertices and let $r_0 = r(\cdot, 0) : V_0 \to [n]$ be any initial rank function which is consistent with the ranking scheme. (For the random labeling scheme we assign any set of labels $l : V_0 \to (0, 1)$ and form the initial rank function accordingly.) For $t \geq 1$ we form G_t from G_{t-1} according to the following rules:

- (i) Add a new vertex v together with d edges from v to existing vertices chosen randomly with weighted probabilities. The edges are added in d substeps. In each substep, one edge is added, and the probability that a vertex w is chosen as its endpoint is proportional to $r(w, t-1)^{-\alpha}$.
- (ii) Choose uniformly at random a vertex $u \in V_{t-1}$, delete u and all edges incident to u.
- (iii) Update the ranking function $r(\cdot, t) : V_t \to [n]$ according to the ranking scheme.

We refer to the time step t in which vertex v was added to the graph as time in which v was born. Since all results refer to the steady state of the process, no vertices of G_0 remain at the time L when the limiting graph is analyzed.

Our model allows for loops and multiple edges; there seems no reason to exclude them. However, with high probability there will not be many of these, so removing them after the process ends can be shown not to affect our conclusions in any significant way.

We now define the different ranking schemes.

- (i) **Ranking by age**: The newly added vertex v obtains an initial rank n; its rank decreases by one each time a vertex with smaller rank is removed. Formally, $r(v, t) = r(v, t-1) - 1 - \gamma$, where $\gamma = 1$ if the rank of the vertex deleted in step t is smaller than $r(v, t-1)$, and 0 otherwise.
- (ii) **Ranking by inverse age**: The vertex added at time t obtains an initial rank 1; its rank increases by one each time a vertex with higher rank is removed. Formally, $r(v, t) = r(v, t-1) + 1 - \gamma$, where $\gamma = 1$ if the rank of the vertex deleted in step t is smaller than $r(v, t-1)$, and 0 otherwise.
- (iii) **Ranking by prestige label**: The vertex v added at time t obtains a label $l(v) \in (0, 1)$ chosen uniformly at random. Vertices are ranked according to their labels: if $l(u) < l(w)$, then $r(u, t) < r(w, t)$.
- (iv) **Random initial rank**: The vertex added at time t obtains an initial rank R_t which is randomly chosen from $[n]$ according to a prescribed distribution. Ranks of all vertices are adjusted accordingly. Formally, for each $v \in V_{t-1}$, $r(v, t) = r(v, t-1) + \delta - \gamma$, where $\delta = 1$ if $r(v, t-1) > R_t$ and 0 otherwise, and $\gamma = 1$ if where the rank of the vertex deleted in step t is smaller than $r(v, t-1)$, and 0 otherwise.
- (v) **Ranking by degree**: After each time step t, vertices are ranked according to their degrees in G_t, and ties are broken by age. Precisely, if $\deg(u, t) < \deg(w, t)$ then $r(u, t) < r(w, t)$, and if $\deg(u, t) = \deg(w, t)$ then $r(u, t) < r(w, t)$ if u was born before w.

Since the process is an ergodic Markov chain, it will converge to a stationary distribution. The random graph G_L corresponding to this distribution is called a protean graph $\mathcal{P}_n(d, \alpha, scheme)$, where *scheme* indicates the ranking scheme used. The coupon collector problem can give us insight into when the stationary state will be reached. Namely, let $L = n(\log n + O(\omega(n)))$ where $\omega(n)$ is any function tending to infinity with n. It is a well-known result that *aas* after L steps all original vertices will have been deleted. In the case of ranking by age, inverse age or random initial rank this implies that after L steps, the stationary distribution has been reached. In the case of ranking by prestige label, we need $L = 2n(\log n + O(\omega(n)))$ steps for the process to converge: the first $L/2$ steps will remove the initial prestige labels, and another $L/2$ steps will eliminate all vertices that were possibly influenced by prestige labels of the initial vertices.

In the rest of the paper, $\{G_t\}_{t=1}^{\infty}$ is assumed to be a graph sequence generated by the rank-based attachment model, with ranking scheme as defined in each particular section, and d and α are assumed to be the initial degree and attachment strength parameters of the model as defined above. The results are generally about the degree distribution in G_L, where the asymptotics are based on n tending to infinity.

We will use the stronger notion of *wep* in favour of the more commonly used *aas*, since it simplifies some of our proofs. We say that an event holds *with extreme probability* (*wep*), if it holds with probability at least $1 - \exp(-\Theta(\log^2 n))$ as $n \to \infty$. Thus, if we consider a polynomial number of events that each holds *wep*, then *wep* all events hold. To combine this notion with asymptotic notations such as $O()$ and $o()$, we follow the conventions in [13].

For any $0 < \alpha < 1$, we define the function $g_\alpha : \mathbb{N} \to \mathbb{R}$:

$$g_\alpha(n) = \sum_{j=1}^{n} j^{-\alpha} = \frac{n^{1-\alpha}}{1-\alpha} + O(1)\,.$$

Thus, the probability that a vertex v is chosen as a neighbour of v_t in a substep of step 1 of the generation process equals

$$\frac{r(v,t-1)^{-\alpha}}{g_\alpha(n)} = \frac{1-\alpha}{n^{1-\alpha} + O(1)} r(v,t-1)^{-\alpha}.$$

Finally, we will make frequent use of the following standard result about the sum of independent random variables, known as the *Chernoff bound*:

Theorem 2.1 (Chernoff bound, see for example Theorem 2.8 [5])**.** *Let X be a random variable that can be expressed as a sum $X = \sum_{i=1}^{n} X_i$ of independent random indicator variables where $X_i \in \mathrm{Be}(p_i)$ with (possibly) different $p_i = \mathbb{P}(X_i = 1) = \mathbb{E}X_i$. Then the following holds for $t \geq 0$:*

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}X + t) &\leq \exp\left(-\frac{t^2}{2(\mathbb{E}X + t/3)}\right), \\ \mathbb{P}(X \leq \mathbb{E}X - t) &\leq \exp\left(-\frac{t^2}{2\mathbb{E}X}\right). \end{aligned}$$

In particular, if $\varepsilon \le 3/2$, *then*

$$\mathbb{P}(|X - \mathbb{E}X| \ge \varepsilon \mathbb{E}x) \le 2\exp\left(-\frac{\varepsilon^2 \mathbb{E}X}{3}\right).$$

Moreover, if $\mathbb{E}X \le \log^2 n$, *then* wep $X = O(\log^2 n)$.

3. Ranking by age and inverse age

We start by discussing two deterministic ranking schemes: the rank of the new vertex is independent of the stochastic process. In the case of ranking based on age, the new vertex is assigned rank n. This ranking scheme was used in the paper that introduced protean graphs [7]. It was shown there that the degree distribution follows a power law with exponent $1 + 1/\alpha$. In fact, the variable representing the degree of a vertex over time is concentrated around its mean (if this mean is sufficiently large), and the mean shows a "midlife crisis": starting from a fixed initial degree d, a given vertex first tends to loose neighbours and gain few new ones; as higher ranked vertices are deleted the vertex drifts towards to higher ranks and start attracting more new neighbours.

To investigate the other extreme, we introduce ranking according to inverse age: each new vertex is assigned rank 1. Thus, right after birth, a given vertex attracts many new neighbours. However, over time new vertices are added to the front of the line, and the vertex will drift toward the lower ranks, eventually loosing more neighbours through deletion process than gaining new ones.

Thus, the degree of a given vertex is determined by its age. To understand the influence of age, we introduce the following concept.

Definition 3.1. *The* age rank $a(v, t)$ *of vertex* v *at time* t *is the rank of* v *if the vertices in* G_t *are ranked by age. In other words,* $a(v, t) - 1$ *equals the number of vertices in* G_t *that were born earlier than* v.

Consider vertices v_i and v_j with ranks $r(v_i, L) = i$ and $r(v_j, L) = j$, respectively. Assume that $j < i$. Because of the ranking scheme, this implies that v_j is younger. It is clear that the rank of v_i when v_j was born is at least $i - j$, so

$$\begin{aligned} \mathbb{E}\deg(v_i, L) &\le d + d\sum_{j=1}^{i-1} \frac{(i-j)^{-\alpha}}{g_\alpha(n)} \\ &= d + (1 + o(1))d\left(\frac{i}{n}\right)^{1-\alpha} \le 2.1d. \end{aligned}$$

Thus, the expected degree is bounded by a constant, and, since the degree is the sum of independent variables, we can use the Chernoff bound (see Theorem 2.1) to show that *wep* the degree of any vertex is at most $\log^2 n$. Thus, vertices do not remain in the high ranks long enough to accumulate a large number of neighbours, and we do not get a power law.

To gain a better understanding of the behaviour of the degree over time, we consider the case where $i = xn$ and $j = yn$, and x, y are two different constants in $(0, 1)$. Since vertices are deleted uniformly at random, the rank $r(v_i, t)$ of vertex

v_i at time t behaves exactly like $n - a(v_i, t)$. Using the results from [7] (Facts 3.2 and 3.3), we find that the expected rank of v_i at the time v_j was born equals:

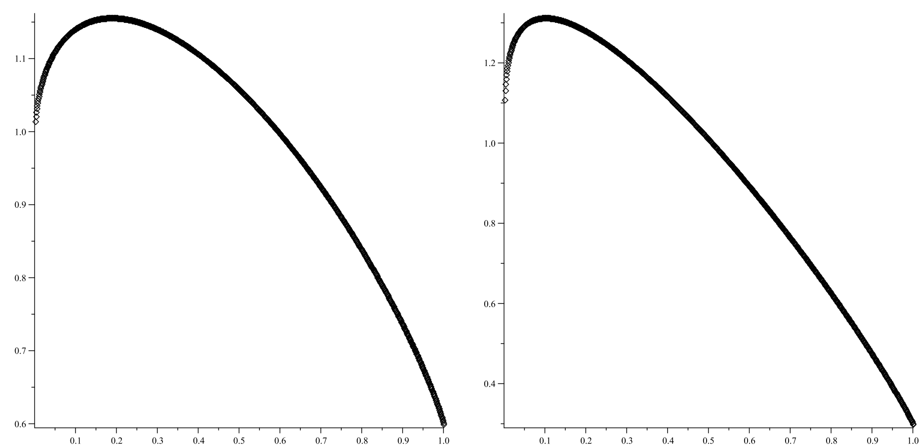
$$n - (n-i)\frac{n}{n-j} = n\frac{i-j}{n-j} = n\frac{x-y}{1-y},$$

and it is possible to show concentration for this random variable.

Therefore,

$$\mathbb{E}\deg(v_{xn}, L) = (1+o(1))d\left(1 - x + (1-\alpha)\int_0^x \left(\frac{x-y}{1-y}\right)^{-\alpha} dy\right).$$

By numerical approximation of the integral we obtained the following figures (see Figure 1), for different values of α. We see that, in this case, instead of a "midlife crisis" we have a "midlife peak": the degree of a vertex initially increases, but reaches a maximum fairly soon, and then starts a steady decline. As the figures show, the place where the maximum is reached depends on α.



(a) $\alpha = 0.4$ (b) $\alpha = 0.7$

FIGURE 1. Function $f(x) = \mathbb{E}\deg(v_{xn}, L)$ for $d = 1$.

4. Ranking by random labeling

In this scheme, each new vertex v obtains a *prestige label* $l(v) \in \mathbb{R}$ chosen randomly according to any distribution with the property that the probability that two vertices receive the same label is zero. Prestige labels for different vertices are chosen independently. Vertices are ranked by their labels: if $l(u) < l(w)$, then $r(u,t) < r(w,t)$.

We assume (without loss of generality) that the prestige labels are chosen uniformly at random from $(0,1)$. Namely, suppose that the labels are chosen from $\mathbb{R}$ according to any probability distribution with a strictly increasing *cumulative distribution function* F. Since F is an increasing function, labels $F(l(v_i))$ lead to exactly the same ranking as labels $l(v_i)$. But $\mathcal{P}(F(l(v_i)) \leq x) = \mathcal{P}(l(v_i) \leq F^{-1}(x)) = F(F^{-1}(x)) = x$, so the values of labels $F(l(v_i))$ are chosen from $(0,1)$ according to the uniform distribution.

The following theorem shows how the expected degree depends on the age rank and prestige label, and gives a concentration result for the case where the expected degree grows sufficiently fast with n.

Theorem 4.1. *Let $0 < \alpha < 1$, $d \in \mathbb{N}$, $i = i(n) \in [n]$, and let v_i be the vertex whose age rank at time L equals $a(v_i, L) = i$. Let $l(v_i)$ be the prestige label of v_i, and assume that $n \cdot l(v_i) > \log^3 n$. Then the expected degree of v_i is given by*

$$\mathbb{E}\deg(v_i, L) = d\frac{i-1}{n-1} + (1 + O(\log^{-1/2} n))d(1-\alpha)l(v_i)^{-\alpha}(1 - i/n).$$

Moreover, if $\mathbb{E}\deg(v_i, L) \geq \log^2 n$, we have that wep

$$\deg(v_i, L) = \mathbb{E}\deg(v_i, L) + O(\sqrt{\mathbb{E}\deg(v_i, L)}\log n).$$

If $\mathbb{E}\deg(v_i, L) < \log^2 n$, then wep $\deg(v_i, L) = O(\log^2 n)$.

Proof. Since $l(v_i)$ is chosen *uar*, at any time during the process the expected rank of v_i is equal to $l(v_i)n$. Since the prestige labels are chosen independently, the rank is the sum of independent random variables, so the Chernoff bound (see Theorem 2.1) applies. Therefore, *wep* $r(v_i, t) = l(v_i)n(1 + O(\log^{-1/2} n))$ during the entire period (since *wep* $L = O(n \log n)$ and the sum of L exponentially small probabilities is still exponentially small).

Next we consider the contribution to the degree of v_i of vertices that are younger than v_i. Let v_t be the vertex with age rank t at time L, and assume $i < t \leq n$. Let $X(t, j)$ be a random indicator variable for the event that vertex v_t chooses v_i as a neighbour at substep j of the time step when v_t was born ($j \in [d]$). Then

$$\begin{aligned}\mathbb{P}(X(t,j) = 1) &= \frac{\Big(l(v_i)n(1 + O(\log^{-1/2} n))\Big)^{-\alpha}}{g_\alpha(n)} \\ &= (1 + O(\log^{-1/2} n))(1-\alpha)l(v_i)^{-\alpha}/n.\end{aligned}$$

The number of younger neighbours of v_i can thus be expressed as a sum $\sum_{t=i+1}^{n}\sum_{j=1}^{d} X(t,j)$ of independent random variables.

For the number of older neighbours, note that vertex v_i had exactly d older neighbours at the time it was born. From the $n-1$ vertices that were older than v_i at the time it was born, only $i-1$ remain. Since vertices are deleted *uar*, this means that the expected number of older neighbours remaining equals $d(i-1)/(n-1)$. Combining the expected number of older and younger neighbours, we obtain:

$$\begin{aligned}\mathbb{E}\deg(v_i, L) &= d\frac{i-1}{n-1} + d(n-i)\mathbb{E}X(t,j) \\ &= d\frac{i-1}{n-1} + (1 + O(\log^{-1/2} n))d(1-\alpha)l(v_i)^{-\alpha}(1 - i/n)\,.\end{aligned}$$

Finally, since the number of younger neighbours of v_i is expressed as a sum of independent random variables, we can use the Chernoff bound (see Theorem 2.1) to show the concentration result. □

Let $Z_k = Z_k(n, d, \alpha)$ denote the number of vertices of degree k and $Z_{\geq k} = \sum_{l \geq k} Z_l$. The following theorem shows that the $Z_{\geq k}$'s follow a power law with exponent $1/\alpha$. Since the $Z_{\geq k}$'s represent the cumulative degree distribution, this implies that the degree distribution follows a power law with exponent $1 + 1/\alpha$.

Theorem 4.2. *Let* $0 < \alpha < 1$ *and* $d \in \mathbb{N}$, $\log^4 n \leq k \leq n^\alpha / \log^{4\alpha} n$. *Then* wep

$$Z_{\geq k} = \left(1 + O(\log^{-1/3} n)\right) \frac{\alpha}{1+\alpha} \left(\frac{d(1-\alpha)}{k}\right)^{1/\alpha} n.$$

Proof. This theorem is a simple consequence of Theorem 4.1. One can show that *wep* each vertex v_i such that

$$l(v_i) \geq \left(1 + \log^{-1/3} n\right) \left(\frac{d(1-\alpha)(1-i/n)}{k}\right)^{1/\alpha}$$

has fewer than k neighbours, and each vertex v_i for which

$$l(v_i) \leq \left(1 - \log^{-1/3} n\right) \left(\frac{d(1-\alpha)(1-i/n)}{k}\right)^{1/\alpha}$$

has more than k neighbours.

Thus,

$$\begin{aligned}
\mathbb{E} Z_{\geq k} &= \sum_{i=1}^{n} \left(1 + O(\log^{-1/3} n)\right) \left(\frac{d(1-\alpha)(1-i/n)}{k}\right)^{1/\alpha} \\
&= \left(1 + O(\log^{-1/3} n)\right) \left(\frac{d(1-\alpha)}{k}\right)^{1/\alpha} n \int_0^1 (1-x)^{1/\alpha} dx \\
&= \left(1 + O(\log^{-1/3} n)\right) \frac{\alpha}{1+\alpha} \left(\frac{d(1-\alpha)}{k}\right)^{1/\alpha} n
\end{aligned}$$

and the assertion follows from Chernoff bound since $\mathbb{E} Z_{\geq k} = \Omega(\log^4 n)$. □

5. Randomly chosen initial rank

Next, we consider the case where the rank R_i of the vertex v added at time i is chosen at random from $[n]$. As in the previous case, the ranks of existing vertices are adjusted accordingly. In contrast to the previous scheme, in this case it does matter according to which distribution R_i is chosen. We make the assumption that all initial ranks are chosen according to a similar distribution. In particular, we fix a continuous bijective function $F : [0, 1] \to [0, 1]$, and for all integers $1 \leq k \leq n$, we let

$$\mathbb{P}(R_i \leq k) = F\left(\frac{k}{n}\right).$$

Thus, F represents the limit, for n going to infinity, of the cumulative distribution functions of the variables R_i. To simplify the calculations while exploring a wide array of possibilities for F, we assume F to be of the form

$$F(x) = x^s, \text{ where } s > 0.$$

We will distinguish three cases: $s = 1$, $s > 1$, and $0 < s < 1$. When $s = 1$, the initial rank is chosen *uar*, and the behaviour mimics that of the random labeling scheme. In the second case, we again obtain power law behaviour with exponent $1/\alpha$, as in the previously studied schemes. For the third case, the behaviour is more complex, and depends on the relative values of s and α.

5.1. The case $s = 1$. The case $s = 1$ represents the uniform distribution of the R_i. As with random labeling, the random variable $r(v, t)$ is sharply concentrated around a fixed value, in this case the initial rank R_i.

Our proofs use the supermartingale method of Pittel et al. [8], as described in [14, Corollary 4.1]. We need the following lemma.

Lemma 5.1. *Let $G_0, G_1, \dots, G_L$ be a random process and X_t a random variable determined by $G_0, G_1, \dots, G_t$, $0 \le t \le L$. Suppose that for some real β and γ,*

$$\mathbb{E}(X_t - X_{t-1} \mid G_0, G_1, \dots, G_{t-1}) < \beta$$

and

$$|X_t - X_{t-1} - \beta| \le \gamma$$

for $1 \le t \le L$. Then for all $\varepsilon > 0$,

$$\mathbb{P}\big(\textit{For some } t \textit{ with } 0 \le t \le L : X_t - X_0 \ge t\beta + \varepsilon\big) \le \exp\Big(- \frac{\varepsilon^2}{2L\gamma^2}\Big) .$$

Lemma 5.2. *Suppose that vertex v obtained an initial rank $R \ge \sqrt{n}\log^2 n$. Then,* wep *$r(v, t) = R(1 + O(\log^{-1/2} n))$ to the end of its life.*

Proof. Fix t so that v is alive at times t and $t+1$. Then $r(v, t+1) - r(v, t) = -1$ if a vertex of rank lower than $r(v, t)$ is deleted, and the new vertex receives rank higher than $r(v, t)$. This happens with probability $(r(v, t)-1)(n-r(v, t))/(n-1)n$. Similarly $r(v, t+1)-r(v, t) = 1$ with probability $(n-r(v, t))r(v, t)/(n-1)n$. Thus,

$$\beta = \mathbb{E}(r(v, t + 1) - r(v, t) \mid r(v, t)) = O(1/n).$$

Clearly, the rank can change by at most one ($\gamma = 1$) so we can use Lemma 5.1 with $\varepsilon = \sqrt{n}\log^{3/2} n$ to get that *wep* $r(v, t) = R(1 + O(\log^{-1/2} n))$ during the whole life of that vertex (note that *wep* v will be deleted after $O(n \log n)$ steps, so $L = O(n \log n)$, and the condition on R implies that $\varepsilon/R \le \log^{-1/2} n$). □

From the previous lemma it follows that the random ranking case for $s = 1$ is very similar to the random labeling case, where an initial rank of R corresponds to a prestige label of R/n. Since we have similar behaviour of the rank, the following theorem is an exact analogue of Theorem 4.1, so its proof is omitted.

Theorem 5.3. *Let $0 < \alpha < 1$, $d \in \mathbb{N}$, $i = i(n) \in [n]$, and let v_i be the vertex whose age rank at time L equals $a(v_i, L) = i$. Let R be the initial rank of v_i, and assume that $R \ge \sqrt{n}\log^2 n$. Then the expected degree of v_i is given by*

$$\mathbb{E}\deg(v_i, L) = d\frac{i-1}{n-1} + (1 + O(\log^{-1/2} n))d(1 - \alpha)(R/n)^{-\alpha}(1 - i/n).$$

Moreover, if $\mathbb{E}\deg(v_i, L) \geq \log^2 n$*, then* wep

$$\deg(v_i, L) = \mathbb{E}\deg(v_i, L) + O(\sqrt{\mathbb{E}\deg(v_i, L)}\log n),$$

and if $\mathbb{E}\deg(v_i, L) < \log^2 n$*, then* wep $\deg(v_i, L) = O(\log^2 n)$.

As a corollary of the behaviour of the degree of each vertex, we obtain our result about the degree distribution, as expressed in the following theorem, which is an analogue of Theorem 4.2. Again, the proof is omitted. Note that the range for k is slightly different due to the condition on the initial rank $R \geq \sqrt{n}\log^2 n$ in Lemma 5.2, which is stronger than the corresponding condition $l(v_i)n > \log^3 n$ in Theorem 4.1 for the prestige label.

Theorem 5.4. *Let* $0 < \alpha < 1$ *and* $d \in \mathbb{N}$, $\log^4 n \leq k \leq n^{\alpha/2}/\log^{3\alpha} n$*. Then* wep

$$Z_{\geq k} = \left(1 - O(\log^{-1/3} n)\right)\frac{\alpha}{1+\alpha}\left(\frac{d(1-\alpha)}{k}\right)^{1/\alpha} n.$$

5.2. The case $s > 1$.

In this case, the initial rank is biased towards the lower ranks. Thus, this behaviour tends towards age-based ranking, addressed in Section 3. Vertices tend to receive an initial rank near the "end of the line", but will drift towards the front over time. Thus the age rank is not concentrated around a constant value, as in previous cases, but tends to decrease with time. The rank function also exhibits more complex behaviour in this case.

We first study the age rank of a vertex v. We assume without loss of generality that v was born at time 0, so $a(v, 0) = n$. For $t > 0$, $a(v, t)$ decreases by one precisely when in time step $t + 1$, the vertex u which is deleted was older than v, so $a(u, t) < a(v, t)$. So, we obtain that

$$\mathbb{E}(a(v, t+1) - a(v, t) \mid G_t) = -\frac{a(v, t) - 1}{n - 1},$$

conditional on the fact that v is not deleted.

To analyze the age rank, we use the differential equations method [14]. Defining a real function $z(x)$ to model the behaviour of $a(v, xn)/n$, the above relation implies the following differential equation

$$z'(x) = -z(x) \tag{1}$$

with the initial condition $z(0) = 1$.

The general solution is $z(x) = \exp(-x + C)$, $C \in \mathbb{R}$ and the particular solution is $z(x) = \exp(-x)$. This *suggests* that a random variable $a(v, t)$ should be close to a deterministic function $n\exp(-t/n)$. The following theorem precisely states the conditions under which this holds.

Theorem 5.5. *Let* $a(v, t)$ *be the age rank of vertex* v *at time* t*. Then* wep*, for every* t *in the range* $0 \leq t \leq t_f = \frac{1}{2}n\log n - 2n\log\log n$*, we have*

$$a(v, t) = n\exp(-t/n)(1 + O(\log^{-1/2} n)) \tag{2}$$

conditional upon the vertex v *surviving until time* t_f.

Proof. We transform $a(v,t)$ into something close to a martingale. Consider the following real-valued function

$$H(a,t) = \log a + t/n \tag{3}$$

and the stopping time

$$T = \min\{t \geq 0 : a(v,t) < (1/2)\sqrt{n}\log^2 n \vee t = t_f\} .$$

(A stopping time is any random variable T with values in $\{0,1,\dots\} \cup \{\infty\}$ for which it can be determined whether $T = \hat{t}$ for any time $\hat{t}$ from knowledge of the process up to and including time $\hat{t}$.)

Let $\mathbf{w}_t = (a(v,t),t)$, and consider the sequence of random variables $(H(\mathbf{w}_t) : 0 \leq t \leq t_f)$. H is chosen so that $H(\mathbf{w})$ is close to a constant along every trajectory $\mathbf{w}$ of the differential equation (1). It is easy to check that the second-order partial derivatives of H are $O(a^{-2}) = O(n^{-1}\log^{-4} n)$ along the trajectory $\mathbf{w}_t$, provided $T > t$. Therefore, with $i \wedge T$ denoting $\min\{i,T\}$, we have

$$\begin{aligned} &H(\mathbf{w}_{(t+1)\wedge T}) - H(\mathbf{w}_{t\wedge T}) \\ &\quad = (\mathbf{w}_{(t+1)\wedge T} - \mathbf{w}_{t\wedge T}) \cdot \text{ grad } \ H(\mathbf{w}_{t\wedge T}) + O(1/n\log^4 n) . \end{aligned} \tag{4}$$

Observe also that,

$$\begin{aligned} &\mathbb{E}(\mathbf{w}_{t+1} - \mathbf{w}_t \mid G_t) \cdot \text{ grad } H(\mathbf{w}_t) \\ &\quad = \left(-\frac{a(v,t)-1}{n-1}, 1\right) \cdot (1/a(v,t), 1/n) = O((a(v,t)n)^{-1}) = O(n^{-3/2}\log^{-2} n), \end{aligned}$$

provided $T > t$.

Taking the expectation of (4) conditional on $G_{t\wedge T}$, we obtain that

$$\mathbb{E}(H(\mathbf{w}_{(t+1)\wedge T}) - H(\mathbf{w}_{t\wedge T}) \mid G_{t\wedge T}) = O(1/n\log^4 n) .$$

From (4), noting that grad $H(\mathbf{w}_t) = (O(1/a(v,t)), 1/n)$, and using the fact that the rank changes by at most one in each step,

$$|H(\mathbf{w}_{(t+1)\wedge T}) - H(\mathbf{w}_{t\wedge T})| = O(1/a(v,t\wedge T)) + O(1/n) + O(1/n\log^4 n) = O(1/\sqrt{n}\log^2 n) .$$

Now we may apply Lemma 5.1 to the sequence $(H(\mathbf{w}_{t\wedge T}) : 0 \leq t \leq t_f)$, and symmetrically to $(-H(\mathbf{w}_{t\wedge T}) : 0 \leq t \leq t_f)$, with $\varepsilon = 1/\log^{1/2} n$, $\beta = O(1/n\log^4 n)$, and $\gamma_t = O(1/\sqrt{n}\log^2 n)$ to show that *wep*

$$|H(\mathbf{w}_{t\wedge T}) - H(\mathbf{w}_{t_0})| = O(\log^{-1/2} n).$$

As $H(\mathbf{w}_0) = \log n$, this implies from the definition (3) of the function H, that *wep* equation (2) holds for every $0 \leq t \leq T$.

To complete the proof we need to show that *wep*, $T = t_f$. The events asserted by (2) hold *wep* up until time T, as shown above. Thus, in particular, *wep* $a(v,T) = (1+o(1))n\exp(-T/n) > (1+o(1))\sqrt{n}\log^2 n$ which implies that $T = t_f$ *wep*. □

Exactly the same approach can be used to study the rank of a vertex after t steps of the process, given that its initial rank is equal to R. We present a sketch of the proof only.

Theorem 5.6. *Suppose that a vertex v obtained an initial rank $r(v,0) = R < 0.99n$ at time 0. Then* wep, *for every t in the range $0 \le t \le t_f = \frac{1}{2}n\log n - 2n\log\log n$ conditional upon the vertex v surviving until time t,*

$$r(v,t) = n\left(\left(\left(\frac{R}{n}\right)^{1-s} - 1\right)e^{(s-1)t/n} + 1\right)^{\frac{1}{1-s}} (1 + O(\log^{-1/2} n))$$

provided

$$n\left(\left(\left(\frac{R}{n}\right)^{1-s} - 1\right)e^{(s-1)t/n} + 1\right)^{\frac{1}{1-s}} \ge \sqrt{n}\log^2 n\,.$$

Proof. The conditional expected change in $r(v,t)$ in time step $t+1$, conditional on vertex v surviving to time $t+1$, is given by:

$$\mathbb{E}(r(v,t+1) - r(v,t) \mid G_t) = -\frac{r(v,t)-1}{n-1} + \left(\frac{r(v,t)}{n}\right)^s.$$

Defining a real function $z(x)$ to model the behaviour of $r(v,xn)/n$, this suggests the differential equation

$$z'(x) = -z(x) + z(x)^s,$$

with the initial condition $z(0) = R/n$. The general solution is

$$z(x) = \left(Ce^{(s-1)x} + 1\right)^{\frac{1}{1-s}}, \quad C \in \mathbb{R},$$

and the particular solution is

$$z(x) = \left(\left(\left(\frac{R}{n}\right)^{1-s} - 1\right)e^{(s-1)x} + 1\right)^{\frac{1}{1-s}}.$$

Define the function

$$H(r,t) = \log\left(\left(\frac{n}{r}\right)^{s-1} - 1\right) - (s-1)\frac{t}{n}$$

and the stopping time

$$T = \min\{t \ge 0 : r(v,t) < (1/2)\sqrt{n}\log^2 n \ \vee \ t = t_f\}\,.$$

Let $\mathbf{w}_t = (r(v,t),t)$. As in the analysis of the age rank, H is chosen to be close to a constant along every trajectory of the differential equation. Specifically, it can be shown that

$$\begin{aligned} |H(\mathbf{w}_{(t+1)\wedge T}) - H(\mathbf{w}_{t\wedge T})| &= O(1/r(v,t\wedge T)) = O(1/\sqrt{n}\log^2 n) \\ \mathbb{E}(H(\mathbf{w}_{(t+1)\wedge T}) - H(\mathbf{w}_{t\wedge T}) \mid G_{t\wedge T}) &= O(1/n\log^4 n), \end{aligned}$$

and $H(\mathbf{w}_0) = \log((R/n)^{1-s} - 1)$. Using Lemma 5.1 in a similar way as in the previous proof, we can then show that *wep* $H(\mathbf{w}_{t\wedge T}) = H(\mathbf{w}_0) + O(\log^{-1/2} n)$.

Solving for r in the expression for H, we obtain that that *wep*

$$\begin{aligned} r(v,t) &= n\left(e^{H(\mathbf{w}_0)}e^{(s-1)t/n}+1\right)^{\frac{1}{1-s}}(1+O(\log^{-1/2} n)) \\ &= n\left(\left(\left(\frac{R}{n}\right)^{1-s}-1\right)e^{(s-1)t/n}+1\right)^{\frac{1}{1-s}}(1+O(\log^{-1/2} n)). \end{aligned}$$

□

Now we are ready to state the main theorem in this section.

Theorem 5.7. *Let* $0<\alpha<1$ *and* $d\in\mathbb{N}$, $\log^4 n\le k\le n^{\alpha/2}\log^{-3\alpha} n$. *Then* wep

$$Z_{\ge k}=(1+o(1))\left(\frac{d(1-\alpha)}{k(1+\alpha)}\right)^{1/\alpha} n.$$

Proof. Consider vertices v_i and v_j with age-ranks $a(v_i,L)=i$ and $a(v_j,L)=j$, respectively, and let $i=xn$ and $j=yn$ $(i<j)$. Suppose that v_i obtained an initial rank of R. Let t_i and t_j be the times that vertices v_i and v_j were born, respectively. By Theorem 5.5, *wep* $t_i=L-(1+O(\log^{-1/2} n))n\log(1/x)$ and $t_j=L-(1+O(\log^{-1/2} n))n\log(1/y)$, and $t_j-t_i=(1+O(\log^{-1/2} n))n\log(y/x)$. By Theorem 5.6, *wep* v_i had the following rank when v_j was born:

$$r(v_i,t_j)=n\left(\left(\left(\frac{R}{n}\right)^{1-s}-1\right)\left(\frac{y}{x}\right)^{s-1}+1\right)^{\frac{1}{1-s}}(1+O(\log^{-1/2} n)).$$

Thus, the contribution to the degree of v_i of vertices born after v_i is the sum of independent indicator variables of the event that a vertex v_j links to v_i in a particular substep of time step t_j. The probability of this event is $r(v_i,t_j)^{-\alpha}/g_\alpha(n)$. Since every vertex has initial degree d, the contribution to the degree of v_i by older vertices is $O(d)$. Combining this, we obtain the following expression for the expected degree:

$$\mathbb{E}\deg(v_i,L)=O(d)+(1+O(\log^{-1/2} n))d(1-\alpha)\int_x^1\left(\left(\left(\frac{R}{n}\right)^{1-s}-1\right)\left(\frac{y}{x}\right)^{s-1}+1\right)^{\frac{-\alpha}{1-s}}dy.$$

If $x=\Omega(1)$ and $R/n=\Omega(1)$ then the expected degree is a constant and the degree is smaller than $\log^2 n$ *wep*. Otherwise it simplifies to

$$\begin{aligned} \mathbb{E}\deg(v_i,L) &= (1+O(\log^{-1/2} n))d(1-\alpha)\left(\left(\frac{R}{n}\right)^{1-s}-1\right)^{\frac{-\alpha}{1-s}}x^{-\alpha}\int_x^1 y^\alpha dy \\ &= (1+O(\log^{-1/2} n))\frac{d(1-\alpha)}{1+\alpha}\left(\left(\frac{R}{n}\right)^{1-s}-1\right)^{\frac{-\alpha}{1-s}}\left(x^{-\alpha}-x\right), \end{aligned}$$

and, provided $\mathbb{E}\deg(v_i,L)=\Omega(\log^4 n)$, *wep* $\deg(v_i,L)=\mathbb{E}\deg(v_i,L)(1+O(\log^{-1/2} n))$, by the Chernoff bound.

Therefore, we get a threshold $R_0 = R_0(k,x)$ on the initial rank which causes the vertex to have degree at least k, namely,

$$R_0(k,x) = n\left(\left(\frac{d(1-\alpha)}{k(1+\alpha)}\left(x^{-\alpha}-x\right)\right)^{\frac{1-s}{\alpha}}+1\right)^{\frac{1}{1-s}}.$$

Precisely, a vertex v_i with initial rank R has degree at least k (provided $k \geq \log^4 n$) if $R \leq R_0(k,i/n)(1-\log^{-1/3} n)$, and degree at most $k-1$ if $R \geq R_0(k,i/n)(1+\log^{-1/3} n)$.

The expected number of vertices of degree at least k is

$$\begin{aligned}\sum_{i=1}^{n}\left(\frac{R_0(k,i/n)}{n}\right)^s(1+o(1)) &= (1+o(1))n\int_0^1\left(\left(\frac{d(1-\alpha)}{k(1+\alpha)}\left(x^{-\alpha}-x\right)\right)^{\frac{1-s}{\alpha}}+1\right)^{\frac{s}{1-s}}dx\\ &= (1+o(1))\left(\frac{d(1-\alpha)}{k(1+\alpha)}\right)^{1/\alpha}n.\end{aligned}$$

To see the last step, let $A = \frac{d(1-\alpha)}{k(1+\alpha)}$. Using the substitution $x = A^{1/\alpha}z$, and noting that $A = O(1/k) = o(1)$, we obtain

$$\begin{aligned}\left(A\left(x^{-\alpha}-x\right)\right)^{\frac{1-s}{\alpha}} &= \left(A\left(A^{-1}z^{-\alpha}-A^{1/\alpha}z\right)\right)^{\frac{1-s}{\alpha}}\\ &= \left(z^{-\alpha}-A^{1+1/\alpha}z\right)^{\frac{1-s}{\alpha}}\\ &= z^{s-1}(1-A^{1+1/\alpha}z^{1+\alpha})^{\frac{1-s}{\alpha}}\\ &= z^{s-1}(1+o(1)).\end{aligned}$$

Thus,

$$\begin{aligned}\int_0^1\left(\left(A\left(x^{-\alpha}-x\right)\right)^{\frac{1-s}{\alpha}}+1\right)^{\frac{s}{1-s}}dx &= A^{1/\alpha}(1+o(1))\int_0^{A^{-1/\alpha}}\left(z^{s-1}+1\right)^{\frac{s}{1-s}}dz\\ &= A^{1/\alpha}(1+o(1)),\end{aligned}$$

since the antiderivative of $(z^{s-1}+1)^{\frac{s}{1-s}}$ is $z(z^{s-1}+1)^{\frac{1}{1-s}}$. The assertion follows from the Chernoff bound. □

5.3. The case $0 < s < 1$. In this case, the distribution of the initial rank R_i is biased towards the higher ranks. Thus, the behaviour tends somewhat towards a rank-based process based on *inverse age*, where new vertices are ranked first.

The results on the behaviour of age rank and rank for the case where $s > 1$, as given in Theorems 5.5 and 5.6, do not depend on s, and thus hold for this case as well. Using these theorems, we can derive the degree of a vertex with age rank i. The interesting fact is that this degree depends on both s and α. In particular, there are two regimes for s: if $s > 1-\alpha$, then the degree depends on the initial rank R with an exponent that depends both on α and on s, and if $s \leq 1-\alpha$, the behaviour mimics that of the inverse age case, and *wep* the degree of each vertex is bounded from above by $\log^2 n$.

Theorem 5.8. *Let $0 < \alpha < 1$, $d \in \mathbb{N}$, $i = i(n) \in [n]$ so that $i = xn$ for $x \in (0, 1)$. Let v_i be the vertex whose age rank at time L equals $a(v_i, L) = i$. Let R be the initial rank of v_i, and assume that $R \geq \sqrt{n} \log^2 n$.*

If $1 - \alpha < s < 1$ and $R \leq n \log^{-3/(s+\alpha-1)} n$, then wep

$$\deg(v_i, L) = (1 + o(1))d(1 - \alpha) \left(\frac{n}{R}\right)^{s+\alpha-1} \frac{x}{s + \alpha - 1}.$$

If $0 < s \leq 1 - \alpha$, then wep $\deg(v_i, L) = O(\log^2 n)$.

Proof. Let v_i be a vertex with age rank i and initial rank R as in the statement of the theorem. For any vertex v_j with age rank $j > i$ (so v_i is older than v_j), let t_j be the time when vertex v_j is born. By Theorem 5.5,

$$t_j = L - n \log(n/i)(1 + O(\log^{-1/2} n)).$$

By the above, $t_j - t_i = n \log(\frac{j}{i})(1 + O(\log^{-1/2} n))$, so by Theorem 5.6, *wep* the rank $r(i, t_j, R)$ of v_i when v_j was born is

$$r(i, t_j, R) = n \left(\left(\left(\frac{R}{n}\right)^{1-s} - 1 \right) \left(\frac{j}{i}\right)^{(s-1)(1+O(\log^{-1/2} n))} + 1 \right)^{\frac{1}{1-s}} (1+O(\log^{-1/2} n)).$$

Now suppose that $j = i + t$, where $t \leq T = n \log^{-1} n$. Since $t/n = O(\log^{-1} n)$,

$$\begin{aligned} \left(\frac{j}{i}\right)^{(s-1)(1+O(\log^{-1/2} n))} &= \left(1 + \frac{t}{xn}\right)^{(s-1)(1+O(\log^{-1/2} n))} \\ &= 1 - (1 - s)\frac{t}{xn}(1 + O(\log^{-1/2} n)). \end{aligned}$$

If $R/n = o(1)$, then

$$\begin{aligned} r(i, t_j, R) &= n \left(\left(\frac{R}{n}\right)^{1-s} + (1 - s)\frac{t}{xn} \right)^{\frac{1}{1-s}} (1 + o(1)) \\ &= R \left(1 + (1 - s) \frac{t}{xn(R/n)^{1-s}} \right)^{\frac{1}{1-s}} (1 + o(1)). \end{aligned} \tag{5}$$

If $R = \Theta(n)$, then $r(i, t_j, R) = R(1+o(1))$, and thus formula (5) correctly expresses $r(i, t_j, R)$ for the case where $t < T$, and thus $t/n = o(1)$.

We can use expression (5) to estimate the number of edges from vertices v_j with $j = i + t$ and $t \leq T$.

$$
\begin{aligned}
& d\sum_{t=1}^{T}\frac{r(i,t_j,R)^{-\alpha}}{g_\alpha(n)} \\
&= (1+o(1))\frac{dR^{-\alpha}}{g_\alpha(n)}\sum_{t=1}^{T}\left(1+(1-s)\frac{t}{xn(R/n)^{1-s}}\right)^{\frac{-\alpha}{1-s}} \\
&= (1+o(1))\frac{d(1-\alpha)R^{-\alpha}}{n^{1-\alpha}}n\left(\frac{R}{n}\right)^{1-s}\int_0^y\left(1+(1-s)\frac{z}{x}\right)^{\frac{-\alpha}{1-s}}dz \\
&= (1+o(1))d(1-\alpha)\left(\frac{R}{n}\right)^{1-s-\alpha}\frac{x}{1-s}\int_1^{1+(1-s)y/x}w^{\frac{-\alpha}{1-s}}\,dw, \qquad (6)
\end{aligned}
$$

where $y = (T/n)(R/n)^{s-1} = \log^{-1} n(R/n)^{s-1} = \Omega(\log n)$. (The second step was obtained by estimating the sum by an integral, and making the substitution $z = (t/n)(R/n)^{s-1}$.)

Now suppose first that $1 - s - \alpha < 0$. Then the integral is bounded, specifically

$$
\begin{aligned}
\int_1^{1+(1-s)y/x}w^{\frac{-\alpha}{1-s}}\,dw &= \frac{1-s}{1-s-\alpha}\left((1+(1-s)y/x)^{\frac{1-s-\alpha}{1-s}}-1\right) \\
&= \frac{1-s}{s+\alpha-1}(1+o(1)).
\end{aligned}
$$

Thus the contribution to the expected degree of v_i from vertices v_{i+t} with $t \leq T$ equals

$$
d(1-\alpha)\left(\frac{n}{R}\right)^{s+\alpha-1}\frac{x}{s+\alpha-1}(1+o(1)).
$$

For the case where $t > T$, we use the fact that

$$
\left(1+\frac{t}{xn}\right)^{s-1} = 1+O\left(\frac{t}{n}\right),
$$

since t/n is bounded from above by 1. From (5) and the fact that $(R/n)^{1-s} \leq t/(yn) = o(t/n)$ we conclude that

$$
r(i,t_j,R) = \Omega\left(n\left(\frac{t}{n}\right)^{\frac{1}{1-s}}\right).
$$

Then

$$\begin{aligned} d\sum_{t=T+1}^{n-i} \frac{r(i,t_j,R)^{-\alpha}}{g_\alpha(n)} &= O(1)\sum_{t=T+1}^{n-i} \frac{\left(n\left(\frac{t}{n}\right)^{\frac{1}{1-s}}\right)^{-\alpha}}{n^{1-\alpha}} \\ &= O(n^{-(1-\alpha-s)/(1-s)})\sum_{t=T+1}^{n-i} t^{-\alpha/(1-s)} \\ &= O(n^{-(1-\alpha-s)/(1-s)})\int_T^\infty t^{-\alpha/(1-s)} \\ &= O\left(\left(\frac{T}{n}\right)^{(1-s-\alpha)/(1-s)}\right) \\ &= O\left(\left(y\left(\frac{R}{n}\right)^{1-s}\right)^{(1-s-\alpha)/(1-s)}\right) \\ &= o((R/n)^{1-s-\alpha}), \end{aligned} \tag{7}$$

since $y^{(1-s-\alpha)/(1-s)} = o(1)$. Thus this part of the sum does not substantially contribute to the expected degree of v_i.

If $1-s-\alpha = 0$, then

$$\int_1^{1+(1-s)y/x} w^{\frac{-\alpha}{1-s}}\, dw = \log(1+(1-s)y/x) = O(\log\log n).$$

Also, the terms before the integral in (6) are now $O(1)$, so the contribution to the expected degree of v_i from vertices v_{i+t} with $t \le T$ is $O(\log\log n)$. For the second part of the sum, note that

$$\sum_{t=T+1}^{n-i} t^{-\alpha/(1-s)} = \sum_{t=T+1}^{n-i} t^{-1} = O(\log n).$$

Thus, we have that the expected degree of each vertex is $O(\log n)$.

If $1-s-\alpha > 0$, then

$$\int_1^{1+(1-s)y/x} w^{\frac{-\alpha}{1-s}}\, dw = O(y^{(1-s-\alpha)/(1-s)}) = O(\log^{(1-s-\alpha)/(1-s)} n) = O(\log n),$$

$(R/n)^{1-s-\alpha} = O(\log^{-3} n)$, and the contribution from vertices v_{i+t} with $t \le T$ is $o(1)$. In this case,

$$\sum_{t=T+1}^{n-i} t^{-\alpha/(1-s)} = O(n^{(1-s-\alpha)/(1-s)}),$$

so from (7), we see that $d\sum_{t=T+1}^{n-i} \frac{r(i,t_j,R)^{-\alpha}}{g_\alpha(n)} = O(1)$.

Finally, since $\deg(v_i, L)$ is expressed as a sum of independent random variables, we can use the Chernoff bound to show the concentration result. □

Theorem 5.9. *Let* $0 < \alpha < 1$, $1 - \alpha < s < 1$. *Let* $d \in \mathbb{N}$ *and*

$$\log^4 n \leq k \leq \left(\frac{n}{\log^3 n} \right)^{\frac{s+\alpha-1}{s}}.$$

Then wep

$$Z_{\geq k} = (1 + o(1)) \frac{s + \alpha - 1}{2s + \alpha - 1} \left(\frac{k(s + \alpha - 1)}{d(1 - \alpha)} \right)^{-\frac{s}{s+\alpha-1}} n.$$

The proof is similar to the proof of Theorem 5.7 so technical details are omitted.

Proof. Consider vertex v_i ($i = xn$) with age-rank $a(v_i, L) = i$. From Theorem 5.8, we obtain the following threshold $R_0(k, x)$ on the initial rank for this vertex having degree k (for values of k as stated in the theorem):

$$R_0(k, x) = (1 + o(1)) n \left(\frac{d(1 - \alpha) x}{k(s + \alpha - 1)} \right)^{\frac{1}{s+\alpha-1}}.$$

Therefore, the expected number of vertices of degree at least k is

$$\begin{aligned} \mathbb{E} Z_{\geq k} = \sum_{i=1}^{n} \left(\frac{R_0(k, i/n)}{n} \right)^s &= (1 + o(1)) n \left(\frac{d(1 - \alpha) x}{k(s + \alpha - 1)} \right)^{\frac{s}{s+\alpha-1}} \int_0^1 x^{\frac{s}{s+\alpha-1}} dx \\ &= (1 + o(1)) n \left(\frac{d(1 - \alpha) x}{k(s + \alpha - 1)} \right)^{\frac{s}{s+\alpha-1}} \frac{s + \alpha - 1}{2s + \alpha - 1}. \end{aligned}$$

The assertion follows from the Chernoff bound since $\mathbb{E} Z_{\geq k} = \Omega(\log^3 n)$. □

6. Ranking by degree

The final ranking scheme is based on the same principle as preferential attachment: vertices with higher degree are ranked higher, and thus have a higher probability of receiving a link. Precisely, the rank function $r(\cdot, t) : V_t \to [t]$ is determined by the degree sequence at time t: if $\deg(v_i, t) > \deg(v_j, t)$, then $r(v_i, t) < r(v_j, t)$; otherwise (that is, if $\deg(v_i, t) = \deg(v_j, t)$) $r(v_i, t) < r(v_j, t)$ if $i < j$. Our results in this case are more tenuous than in the previous cases; we can only conjecture that the degree distribution variables converge in this case as well. Because of the importance of the preferential attachment principle in the modelling of real-world networks, we decided to include our results even in they are somewhat inconclusive.

For all $t \geq 1$ and $k \geq 0$, let $Z_k(t)$ denote the number of vertices of degree k in G_t, and let $Z_{\geq k}(t) = \sum_{j \geq k} Z_j(t)$ (in particular, $Z_{\geq 0}(t) = n$). At time t, the vertices of degree k have ranks starting at $Z_{\geq k+1}(t) + 1$, and ending at $Z_{\geq k}(t)$. In this section, we assume that $d = 1$. When a vertex v_i is deleted, and a new vertex v_{t+1} and an edge $v_j v_{t+1}$ is added at time $t + 1$, the change in any Z_k has contributions from six possible sources: if v_i is a vertex of degree k or a neighbour of a vertex of degree k, Z_k decreases, but if v_i is a neighbour of a vertex of degree $k + 1$, Z_k increases. The expected net increase in Z_k due to the deletion of v_i is thus $((k + 1) Z_{k+1}(t) - (1 + k) Z_k(t))/n$.

The probability that a vertex of degree k receives a link in step $t+1$ equals

$$\sum_{j=Z_{\geq k+1}(t)+1}^{Z_{\geq k}(t)} \frac{j^{-\alpha}}{g_\alpha(n)} = \frac{g_\alpha(Z_{\geq k}(t)) - g_\alpha(Z_{\geq k+1}(t))}{g_\alpha(n)}.$$

Thus, the following equations express the expected change in each time step:

$$\begin{aligned}
\mathbb{E}(Z_0(t+1) - Z_0(t) \mid G_t) &= -\frac{g_\alpha(n) - g_\alpha(Z_{\geq 1}(t))}{g_\alpha(n)} + \frac{Z_1(t)}{n} - \frac{Z_0(t)}{n}, \\
\mathbb{E}(Z_1(t+1) - Z_1(t) \mid G_t) &= 1 + \frac{g_\alpha(n) - g_\alpha(Z_{\geq 1}(t))}{g_\alpha(n)} - \frac{g_\alpha(Z_{\geq 1}(t)) - g_\alpha(Z_{\geq 2}(t))}{g_\alpha(n)} \\
&\quad + \frac{2Z_2(t)}{n} - \frac{2Z_1(t)}{n},
\end{aligned}$$

and similarly, for all $k \geq 2$,

$$\begin{aligned}
\mathbb{E}(Z_k(t+1) - Z_k(t) \mid G_t) &= \frac{g_\alpha(Z_{\geq k-1}(t)) - g_\alpha(Z_{\geq k}(n))}{g_\alpha(n)} \\
&\quad - \frac{g_\alpha(Z_{\geq k}(t)) - g_\alpha(Z_{\geq k+1}(t))}{g_\alpha(n)} \\
&\quad + \frac{(k+1)Z_{k+1}(t)}{n} - \frac{(k+1)Z_k(t)}{n}.
\end{aligned}$$

Since the process is an ergodic Markov chain, each random variable $Z_k(t)$ will tend to a limiting random variable Z_k as t grows large, where Z_k represents the value of the number of vertices of degree k in the limiting protean graph. Considering the other results in this paper, and the results in [6] for similar graph processes, it seems reasonable to assume that, for a fixed value of k, Z_k is concentrated and Z_k/n converges as n grows large, in other words, that *wep* $Z_k = c_k n + o(n)$. Under this assumption, $\mathbb{E}(Z_k(t+1) - Z_k(t) \mid G_t) \to 0$, and we can use the equations above to find a recurrence relation for the c_k.

To express this recurrence, we define $C_k = \sum_{i=k}^{\infty} c_i = 1 - \sum_{i=0}^{k-1} c_i$, and observe that $c_k = C_k - C_{k+1}$ and $C_0 = 1$. Then, using that $g_\alpha(cn) = \frac{1}{1-\alpha}(cn)^{1-\alpha} + O(1)$, we obtain the following recurrence relations between the C_k:

$$\begin{aligned}
0 &= -(1 - C_1^{1-\alpha}) + (C_1 - C_2) - (1 - C_1) \\
0 &= 1 + (1 - C_1^{1-\alpha}) - (C_1^{1-\alpha} - C_2^{1-\alpha}) + 2(C_2 - C_3) - 2(C_1 - C_2) \\
0 &= (C_{k-1}^{1-\alpha} - C_k^{1-\alpha}) - (C_k^{1-\alpha} - C_{k+1}^{1-\alpha}) \\
&\quad + (k+1)(C_{k+1} - C_{k+2} - (k+1)(C_k - C_{k+1}) \qquad \text{for } k \geq 2.
\end{aligned}$$

The last recurrence is telescoping, so for $k \geq 2$,

$$\begin{aligned}
& (C_k^{1-\alpha} - C_{k+1}^{1-\alpha}) - (k+1)(C_{k+1} - C_{k+2}) \\
= \; & (C_{k-1}^{1-\alpha} - C_k^{1-\alpha}) - k(C_k - C_{k+1}) - (C_k - C_{k+1}) \\
\vdots \; & \\
= \; & (C_1^{1-\alpha} - C_2^{1-\alpha}) - 2(C_2 - C_3) - (C_2 - C_{k+1}) \\
= \; & 1 + (1 - C_1^{1-\alpha}) - 2(C_1 - C_2) - (C_2 - C_{k+1}) \\
= \; & C_{k+1}.
\end{aligned}$$

For $k = 1$, the same relation holds, and thus the C_k satisfy the following recurrence relation:

$$\begin{aligned}
C_2 &= 2C_1 - (1 - C_1^{1-\alpha}) - 1 \\
C_{k+1} &= 1/k(C_k^{1-\alpha} - C_{k-1}^{1-\alpha} + (k+1)C_k) \text{ for } k \geq 2.
\end{aligned}$$

Note that this recurrence leaves the value of C_1 undetermined. We have not been able to solve the recurrence; however, the recurrence relation is consistent, in order, with the expression $C_k = ck^{-1/\alpha}(1 + o(1))$. We conjecture that, in fact, the Z_k are concentrated and follow a power law with exponent $1/\alpha$.

References

[1] A. Bonato, *A Course on the Web Graph*, American Mathematical Society Graduate Studies Series in Mathematics, Providence, Rhode Island, 2008.

[2] C. Borgs, J.T. Chayes, C. Daskalakis, and S. Roch, *First to Market is not Everything: an Analysis of Preferential Attachment with Fitness*, Proceedings of the 39rd annual ACM Symposium on the Theory of Computing (STOC), 135–144 (2007).

[3] F.R.K. Chung, L. Lu, *Complex graphs and networks*, American Mathematical Society, U.S.A., 2004.

[4] S. Fortunato, A. Flammini, and F. Menczer, *Scale-free network growth by ranking*, Phys. Rev. Lett. **96**(21): 218701 (2006).

[5] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[6] J. Janssen and P. Prałat, *Rank-based attachment leads to power law graphs*, preprint.

[7] T. Łuczak and P. Prałat, *Protean graphs*, Internet Mathematics **3** (2006), 21–40.

[8] B. Pittel, J. Spencer and N. Wormald, *Sudden emergence of a giant k-core in a random graph*, J. Combinatorial Theory, Series B **67** (1996), 111–151.

[9] P. Prałat, *A note on the diameter of protean graphs*, Discrete Mathematics **308** (2008), 3399–3406.

[10] P. Pralat, *Connectivity threshold and recovery time in rank-based models for complex networks*, preprint.

[11] P. Prałat, *Protean graphs with a variety of ranking schemes*, Proceedings of the 2nd Annual International Conference on Combinatorial Optimization and Applications (COCOA'08), Lecture Notes in Computer Science **5165**, Springer, 2008, 149–159.

[12] P. Prałat and N. Wormald, *Growing protean graphs*, Internet Mathematics **4** (2009), 1–16.

[13] N.C. Wormald, Random graphs and asymptotics. Section 8.2 in *Handbook of Graph Theory*, J.L. Gross and J. Yellen (eds), pp. 817–836. CRC, Boca Raton, 2004.

[14] N. Wormald, *The differential equation method for random graph processes and greedy algorithms* in Lectures on Approximation and Randomized Algorithms, eds. M. Karoński and H. J. Prömel, PWN, Warsaw, pp. 73–155, 1999.

Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada B3H 3J5
E-mail address: pralat@mathstat.dal.ca

Department of Mathematics and Statistics, Dalhousie University, Halifax, NS, Canada B3H 3J5
E-mail address: janssen@mathstat.dal.ca