PROTEAN GRAPHS WITH A VARIETY OF RANKING SCHEMES

PAWEŁ PRAŁAT

Abstract. The World Wide Web may be viewed as a graph each of whose vertices corresponds to a static HTML web page, and each of whose edges corresponds to a hyperlink from one web page to another. Recently there has been considerable interest in using random graphs to model complex real-world networks to gain an insight into their properties. In this paper, we propose a generalized version of the protean graph (a random model of the web graph) in which the degree of a vertex depends on its age. Classic protean graphs can be seen as a special case of the rank-based approach where vertices are ranked according to age. Here, we investigate graph generation models based on other ranking schemes and show that these models lead to graphs with a power law degree distribution.

1. Introduction

Recently many new random graphs models have been introduced and analyzed by certain common features observed in many large-scale real-world networks such as the 'web graph' (see, for instance, the survey [1]). The web may be viewed as a directed graph whose nodes correspond to static pages on the web, and whose arcs correspond to links between these pages.

One of the most characteristic features of this graph is its degree sequence. Broder et al. [2] noticed that the distribution of degrees follows a power law: the fraction of vertices with degree k is proportional to $k^{-\gamma}$, where γ is a constant independent of the size of the network (more precisely, $\gamma \approx 2.1$ for in-degrees, $\gamma \approx 2.7$ for out-degrees). These observations suggest that the web is not well modeled by traditional random graph models such as $G_{n,p}$ (see, for instance [4]).

Łuczak and the author of this paper introduced in [6] another random graph model of the undirected 'web graphs', the protean graph $\mathcal{P}_n(d,\eta)$, which is controlled by two additional parameters ($d \in \mathbb{N}$ and $0 < \eta < 1$). The major feature of this model is that older vertices are preferred when joining a new vertex into the graph. In [6] it is proved that the degrees of the $\mathcal{P}_n(d,\eta)$ are distributed according to the power law and the behaviour near the connectivity threshold is studied. The author of this paper showed also in [8] that the protean graph $\mathcal{P}_n(d,\eta)$ asymptotically almost surely (a.a.s.) has one giant component, containing a positive fraction of all vertices, whose diameter is equal to $\Theta(\log n)$. (See also [9] where the growing protean graphs are studied.)

Classic protean graphs can be viewed as a special case of the rank-based approach where vertices are ranked according to age. The general approach was first proposed by Fortunato, Flammini and Menczer in [3], and the occurrence of a power law was postulated based on simulations (Janssen and the author of this paper provided rigorous proofs in [5]). In this approach, the vertices are ranked from 1 to n according to

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

some ranking scheme (so the vertex with highest degree has rank 1, etc.), and the link probability of a given vertex is proportional to its rank, raised to the power $-\eta$ for some $\eta \in (0,1)$; we will refer to η as the *attachment strength*. (Negative powers are chosen since a low value for rank should result in a higher link probability.)

As we will show, protean graphs with rank-based attachment leads to power law graphs for a variety of different ranking schemes. One obvious ranking scheme is to rank vertices by age (*the old get richer*); as we already mentioned, this model was studied in [6, 8] and this leads to a power law with the exponent $1+1/\eta$. In this paper, we study a ranking scheme where an external prestige label for each vertex is given and vertices are ranked according to their prestige label. In order to allow for a different distribution of "prestige" over the vertices, we considered also a *random ranking* scheme. 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 $\mathbb{P}(R_i \le k) = (k/n)^s$. First we show that, if $s = 1$, then the situation is similar to the one described previously, and vertices with initial rank R_i exhibit behaviour as if they had received fitness R_i/n. We also consider the case where $s > 1$, so the rank of new vertices is biased towards the higher ranks.

These results suggest an explanation for the power law degree distribution often observed in real-life networks such as the web graph, protein interaction networks, and social networks. The growth of 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.

In order to establish the right attachment strength to model a given real-life network we should consider the following. In a graph in which the number of vertices of degree k decreases roughly as $k^{-\gamma}$ the fraction of vertices of degree at least k changes roughly as

$$\sum_{\ell \ge k} O(\ell^{-\gamma}) = O(k^{1-\gamma})\,.$$

Thus, in order to imitate this distribution the attachment strength η should be set to $\eta \sim 1/(\gamma - 1)$.

2. Definitions

In this section, we formally define the graph generation model based on rank-based attachment. 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 $d \in N$, and attachment strength $\eta \in (0,1)$. 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 initialization and update of the ranking is done according to a *ranking scheme*. Various ranking schemes can be considered; we first give the general model, and then list the ranking schemes.

Let $G_0 = (V_0, E_0)$ be any graph on n vertices and $r_0 = r(\cdot, 0) : V_0 \to [n]$ any initial rank function. (For random labeling scheme we take any function $l : V_0 \to (0, 1)$ and the initial rank function is a function of l; for degree scheme $r_0 = r_0(G_0)$.) For $t \geq 1$ we form G_t from G_{t-1} according to the following rules:

- Choose uniformly at random a vertex $u \in V_{t-1}$, delete u together with all edges incident to it.
- Add a new vertex v_t together with d edges from v_t 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 v is chosen as its endpoint (the link probability), equals

$$\frac{r(v, t-1)^{-\eta}}{\sum_{i=1}^{n} i^{-\eta}} = \frac{1-\eta}{n^{1-\eta} + O(1)} r(v, t-1)^{-\eta}.$$

- Update the ranking function $r(\cdot, t) : V_t \to [n]$ according to the ranking scheme.

Our model allows for loops and multiple edges; there seems no reason to exclude them. However, there will not in general be very many of these, so excluding them can be shown not to affect our conclusions in any significant way.

We now define the different ranking schemes.

- **Ranking by age**: The vertex added at time t obtains an initial rank n; its rank decreases by one each time a vertex with smaller rank is removed.
- **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.
- **Ranking by random labeling**: The vertex added at time t obtains a label $l(v_t) \in (0, 1)$ chosen uniformly at random. Vertices are ranked according to their labels: if $l(v_i) < l(v_j)$, then $r(v_i, t) < r(v_j, t)$.
- **Random ranking**: The vertex added at time t obtains an initial rank R_t which is randomly chosen from $[n]$ according to a prescribed distribution. Formally, let $F : [0, 1] \to [0, 1]$ be any cumulative distribution function. Then for all $k \in [t]$,

$$\mathbb{P}(R_t \leq k) = F(k/t).$$

- **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 $\deg(v_i, t) < \deg(v_j, t)$ then $r(v_i, t) < r(v_j, t)$, and if $\deg(v_i, t) = \deg(v_j, t)$ then $r(v_i, t) < r(v_j, t)$ if $i < j$.

In this paper, due to the space limitations, we focus on ranking by random labeling and random ranking with $F(x) = x^s$ for $s \geq 1$. The other ranking schemes will be studied in a journal version of this paper. In particular, it is interesting and non-trivial task to investigate the ranking by degree scheme; in this case, it is not even clear how long we have to wait to obtain a stationary distribution. For the other schemes (except the random labeling case), it is enough to wait L steps for all vertices to be 'renewed' (for the random labeling case we have to wait two times longer: the first round is needed to have labels distributed uniformly at random, during the second one the process 'forgets' about the initial graph) and from that time the protean process is the Markov chain that is in the stationary distribution (that is, the distribution determined by G_t on the set of all ordered graphs on n vertices is identical for all t.) By the coupon collector problem, a.a.s. $L = n(\log n + O(\omega(n)))$ where $\omega(n)$ is any function tending to infinity with n (for random labeling scheme, clearly $L = 2n(\log n + O(\omega(n)))$ a.a.s.). Furthermore, this distribution does not depend on the choice of G_0 and r_0. The random graph G_L corresponding to this distribution is called a protean graph $\mathcal{P}_n(d, \eta)$.

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 η 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 a.a.s., 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 [10].

3. Ranking by random labeling

In this scheme, each new vertex v_t obtains a label $l(v_t) \in (0,1)$ chosen uniformly at random. (Note that the probability that two vertices receive the same label is zero.) Vertices are ranked by their labels: if $l(v_i) < l(v_j)$, then $r(v_i,t) < r(v_j,t)$.

First we note that the process of choosing a label *uar* from $(0,1)$ does not imply loss of generality. 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 $\mathcal{P}(F(l(v_i)) \le x) = \mathcal{P}(l(v_i) \le F^{-1}(x)) = F(F^{-1}(x)) = x$, so the values of labels $F(l(v_i))$ are chosen from $(0,1)$ according to the uniform distribution.

First we investigate the expected degree of a vertex v at time L with a given age-rank and a label. We use $a(\cdot,t)$ for a ranking by age and stay with $r(\cdot,t)$ for a ranking by random labeling.

Theorem 3.1. *Let $0 < \eta < 1$, $d \in \mathbb{N}$, $i = i(n) \in [n]$, and $0 < l(v_i) = l(v_i)(n) < 1$. If $n \cdot l(v_i) > \log^3 n$, then the expected degree of a vertex v_i with an age-rank $a(v_i,L) = i$ that obtained a label $l(v_i)$, is given by*

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

and wep

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

Proof. It is clear that the expected rank of v_i is equal to $l(v_i)n$ at each step of the process. Moreover, we can use the fact that a sum of independent random variables with large enough expected value is not too far from its mean (see, for example, Theorem 2.8 in [4]). From this it follows that, if $\varepsilon \le 3/2$, then the following inequality, known as a Chernoff bound, holds

$$\mathbb{P}\left(|r(v_i,t) - \mathbb{E}r(v_i,t)| \ge \varepsilon\mathbb{E}r(v_i,t)\right) \le 2\exp\left(-\frac{\varepsilon^2}{3}\mathbb{E}r(v_i,t)\right).$$

Therefore, *wep* $r(v_i,t) = l(v_i)n(1+O(\log^{-1/2} n))$ during the whole period (since $L = O(n\log n)$).

Let $X(t,j)$ be a random indicator variable for an event that vertex v_t (for which $a(v_t,L) = t$) joins v_i at substep j of step when v_t was born ($i < t \le n$, $j \in [d]$). It is clear that

$$\begin{aligned}\mathbb{P}(X(t,j)=1)=1-\mathbb{P}(X(t,j)=0) &= \frac{\Big(l(v_i)n(1+O(\log^{-1/2} n))\Big)^{-\eta}}{n^{1-\eta}/(1-\eta)+O(1)}\\ &= (1+O(\log^{-1/2} n))(1-\eta)l(v_i)^{-\eta}/n.\end{aligned}$$

The number of neighbours v_t of v_i such that $t > i$ is a random variable and can be expressed as a sum $\sum_{t=i+1}^{n}\sum_{j=1}^{d} X(t,j)$ of independent random variables. Note also that vertex v_i generated exactly d edges at the time it was born but only i vertices (including v_i) have not been 'renewed' since then. Thus,

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

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. □

Let $Z_k = Z_k(n,d,\eta)$ 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 $1/\eta$. Since the $Z_{\geq k}$'s represent the cumulative degree distribution, this implies that the degree distribution follows a power law with exponent $1+1/\eta$.

Theorem 3.2. *Let* $0<\eta<1$ *and* $d\in\mathbb{N}$, $\log^4 n \leq k \leq n^{\eta}/\log^{4\eta} n$. *Then* wep

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

Proof. This theorem is a simple consequence of Theorem 3.1. One can show that *wep* each vertex v_i such that

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

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

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

has more than k neighbours.

Thus,

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

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

4. Randomly chosen initial rank

Next, we consider the case where the rank of the new vertex v_i, $R_i = r(v_i, i)$, is chosen at random from $[n]$. As described earlier, 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 $1 \le k \le n$, we let

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

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

$$F(x) = x^s, \text{ where } s \ge 1.$$

(The case $0 < s < 1$ will be studied in the journal version of this paper.)

We start from a special case $s = 1$, where the distribution of each R_i is uniform. We will show that this case is similar to the random labeling case with a label equal to R_i/n. Hence, our aim is to show that the random variable $r(v_i, t)$ is sharply concentrated around R_i. In fact, $r(v_i, t) - r(v_i, i)$ is the sum of the differences $r(v_i, j) - r(v_i, j-1) = X_j$, $i+1 \le j \le t$. If the differences are independent, then the Chernoff bounds are very useful. When the differences are not independent but there is a large degree of independence, results can be often obtained by using large deviation inequalities for corresponding martingales. It is exactly the case here.

Our proofs use the supermartingale method of Pittel et al. [7], as described in [11, Corollary 4.1]. We need the following lemma.

Lemma 4.1. *Let $G_0, G_1, \ldots, G_n$ be a random process and X_t a random variable determined by $G_0, G_1, \ldots, G_t$, $0 \le t \le n$. Suppose that for some real β and constants γ_t,*

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

and

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

for $1 \le t \le n$. Then for all $\alpha > 0$,

$$\mathbb{P}\big(\textit{For some } t \textit{ with } 0 \le t \le n : X_t - X_0 \ge t\beta + \alpha\big) \le \exp\left(-\frac{\alpha^2}{2\sum_{j=1}^{n} \gamma_t^2}\right).$$

Lemma 4.2. *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. Note that $r(v, t+1) - r(v, t) = -1$ (conditionally on the fact that v is not deleted at time $t+1$) with probability $(r(v, t) - 1)(n - r(v, t))/(n-1)n$ and $r(v, t+1) - r(v, t) = 1$ 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 ($\gamma_t = 1$) so we can use Lemma 4.1 with $\alpha = \sqrt{n} \log^{3/2} n$ 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 and $R \ge \sqrt{n} \log^2 n$). $\square$

From the previous lemma it follows that the random ranking case for $s = 1$ is very similar to the random labeling case. The proof of the following theorem is the same as the proof of the Theorem 3.1 so it is omitted. (Note that the range for k is slightly different due to the stronger condition for the initial rank.)

Theorem 4.3. *Let $0 < \eta < 1$ and $d \in \mathbb{N}$, $\log^4 n \le k \le n^{\eta/2}/\log^{3\eta} n$. Then* wep

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

Next, we consider the case where $s > 1$, but before we move to investigating the rank of vertex v after t steps of the process, we study its age-rank. In other words, we would like to know how many vertices have not been 'renewed' after t steps of the process. For this, we use the differential equations method [11]. Without loss of generality, we can assume that the vertex was born at time 0. It is clear that $a(v, 0) = n$ and $a(v, t)$, $t > 0$, is a random variable, which in time step $t + 1$ decreases by one precisely when vertex u for which $a(u, t) < a(v, t)$ is deleted. So, working in the conditional space under consideration, we obtain

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

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 $n \exp(-t/n)$. We will show that it represents the "shape" of a typical process.

Theorem 4.4. *Let $a(v, t)$ be defined as above. Then* wep*, for every t in the range $0 \le t \le t_f = \frac{1}{2} n \log n - 2n \log\log n$, 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

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

and the stopping time

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

(A stopping time is any random variable T with values in $\{0, 1, \dots\} \cup \{\infty\}$ for which it is 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 $(H(\mathbf{w}_t) : 0 \le t \le t_f)$. Note that the second-order partial derivatives of H with respect to $a(v, t)$ and

t are $O(1/a(v,t)^2) = O(1/n\log^4 n)$, 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,

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

provided $T > t$, since H was chosen so that $H(\mathbf{w})$ is close to a constant along every trajectory $\mathbf{w}$ of the differential equation (1).

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

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

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 4.1 to the sequence $(H(\mathbf{w}_{t\wedge T}) : 0 \le t \le t_f)$, and symmetrically to $(-H(\mathbf{w}_{t\wedge T}) : 0 \le t \le t_f)$, with $\alpha = 1/\log^{1/2} n$, $\beta = O(1/n\log^4 n)$, 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 $0 \le t \le T$.

To complete the proof we need to show that *wep*, $T = t_f$. The events asserted by (2) hold with this probability 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*. $\square$

Exactly the same approach can be used to study the rank of 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 4.5. *Suppose that a vertex v obtained an initial rank $r(v,0) = R < (1 - 1/\sqrt{n}\log^2 n)n$ at time 0. Then* wep*, for every $t > 0$ 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. Defining a real function $z(x)$ to model the behaviour of $r(v, xn)/n$, we get $z'(x) = -z(x) + z(x)^s$ with the initial condition $z(0) = R/n$. The general solution is $z(x) = (Ce^{(s-1)x} + 1)^{1/(1-s)}$, $C \in \mathbb{R}$ and the particular solution is

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

□

Now we are ready to state the main theorem in this section. The proof is rather straightforward but again we omit the details in this extended abstract.

Theorem 4.6. *Let* $0 < \eta < 1$ *and* $d \in \mathbb{N}$, $\log^4 n \le k \le n^{\eta/2}/\log^{3\eta} n$. *Then* wep

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

Proof. Consider vertices v_i ($i = xn$) and v_j ($j = yn$) with the age-ranks $a(v_i, L) = i$ and $a(v_j, L) = j$, respectively. Suppose that v_i obtained an initial rank of R. By Theorem 4.4, *wep* vertices v_i and v_j were born $(1 + o(1))n \log(1/x)$ and, respectively, $(1 + o(1))n \log(1/y)$ steps ago. By Theorem 4.5, *wep* v_i had the following rank at that time

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

Thus,

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

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

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

Therefore, we get a threshold $R_0 = R_0(k, x)$ on the initial rank for heaving degree at least $k \ge \log^4 n$, namely,

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

Finally, one can show that the expected number of vertices of degree at least k is asymptotic to

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

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

□

References

[1] A. Bonato, *A survey of web graph models*, Proceedings of Combinatorial and Algorithm Aspects of Networking, 2004.

[2] A. Broder, R. Kumar, F. Maghoul, P. Rahaghavan, S. Rajagopalan, R. State, A. Tomkins and J. Wiener, *Graph structure in the web*, Proc. 9th International World-Wide Web Conference (WWW), 2000, pp. 309–320.

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

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

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

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

[7] 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.

[8] P. Prałat, *A note on the diameter of protean graphs*, Discrete Mathematics, accepted, 10pp.

[9] P. Prałat and N. Wormald, *Growing protean graphs*, Internet Mathematics, accepted, 13pp.

[10] 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.

[11] 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