Internet Mathematics Vol. 3, No. 1: 21-40

Protean Graphs

Tomasz Łuczak and Paweł Prałat

Abstract. We propose a new random model of web graphs in which the degree of a vertex depends on its age. We characterize the degree sequence of this model and study its behaviour near the connectivity threshold.

1. Introduction

One of the most characteristic features of the internet graph, which corresponds to internet web connections (or more generally, of the so-called web graphs), is its degree sequence, in which the fraction of vertices of degree larger than k decreases as a power of k. Since in the standard models of sparse random graphs the fraction of vertices of large degree decreases exponentially with k, a number of new probabilistic models of web graphs for which degree sequence obeys the power law have been proposed (see, for instance, [Bollobás et al. 03, Bollobás et al. 01, Cooper and Frieze 03, Kumar et al. 00]). In this note we introduce yet another random graph model in which the shape of the degree sequence is controlled by some additional parameters. The formal definition of the model, given in the next section, is somewhat technical, but the idea behind it is simple and natural. We start with any graph G on n vertices, and in each step we pick randomly one of the vertices v to be *renewed*. Thus, we delete from G all edges incident to v (this corresponds to a removal of a random node from the network). Then, we generate new edges incident to v according to some prescribed distribution $\mathbf{X} = \mathbf{X}_{n-1}$ (which can be viewed as a *new* node

1542-7951/05 $0.50 per page

that establishes connections with some nodes in the network). We allow the probability that v is joined to w to depend on the *age* of w, i.e., the last time the vertex w had been renewed (it seems natural to assume that *older* nodes are more attractive for newcomers). Let $\mathcal{P}^k(\mathbf{X}) = \mathcal{P}^k(G; \mathbf{X})$ denote the graph obtained by applying this procedure k times. Note that if by times k_1 and k_2 each vertex of a graph G is renewed at least once, the random graphs $\mathcal{P}^{k_1}(G; \mathbf{X})$, $\mathcal{P}^{k_2}(G; \mathbf{X})$ are identical random objects whose properties do not depend on the graph G with which we started, but only on the distribution $\mathbf{X}$ with which we generate edges incident to a renewed vertex. We call the graph generated in this process a *protean graph* and denote it by $\mathcal{P}(\mathbf{X})$.

Note that, unlike in most theoretical models of the internet graph, the number of vertices of the protean graph is large but fixed and does not grow during the protean process. One may view it as a weakness of our approach since the internet graph is, at least at this moment, rapidly expanding. Let us, however, point out a few features of protean graphs that, in our opinion, make this model interesting. Protean graphs generalize some standard models of random graphs (as $G(n, p)$) but could also be used to imitate real networks such as web graphs or, after some modifications, peer-to-peer networks (see a remark in Section 2). This is, of course, done by an appropriate selection of the distribution $\mathbf{X}$; for a particularly natural choice of $\mathbf{X}$, the degree sequence of $\mathcal{P}(\mathbf{X})$ is studied in Section 4. Our model takes into account an additional parameter of a vertex, its age, and predicts how it influences the degree of a vertex. Moreover, it seems that protean graphs not only are interesting as models of the web graphs but also are attractive from a theoretical point of view: they have a very rich dependence structure, and, unlike many other models of random graphs, $\mathcal{P}(\mathbf{X})$ can be viewed as the stationary distribution of the protean process $\{\mathcal{P}^k(\mathbf{X})\}_k$. In order to show similarities and differences between the behaviour of protean graphs and other random graph models, in the last section we study how the threshold of connectivity is affected by the dependence structure of the protean graph and characterize the limit distribution of the *recovery time* for connectivity near the connectivity threshold.

Finally, let us mention that most of the arguments that we use are fairly long and technical. Thus, we often omit details of the proofs that can be easily filled in by the reader (a much more elaborate treatment of the subject can be found in the second author's PhD dissertation [Prałat 04]).

2. Definition

Let $\mathbf{X}_{n-1} = (X_1, \ldots, X_{n-1})$ be an $(n-1)$-dimensional nonnegative integer-valued random variable, G be a graph with vertex set $[n] = \{1, 2, \ldots, n\}$, and $\sigma :$

$[n] \to [n]$ denote a permutation of $[n]$. Consider a Markov chain $\{(\tilde{G}_k, \tilde{\sigma}_k, A_k)\}_{k=0}^{\infty}$ whose stages are triples $(\tilde{G}_k, \tilde{\sigma}_k, A_k)$, where $\tilde{G}_k$ is a (multi)graph with vertex set $[n]$, $\tilde{\sigma}_k : [n] \to [n]$ is a permutation of set $[n]$, and $A_k \subseteq [n]$. The process starts with $(\tilde{G}_0, \tilde{\sigma}_0, A_0) = (G, \sigma, \emptyset)$. In the kth step of the process, we choose a random vertex $i \in [n]$ and move i to the end of the permutation $\tilde{\sigma}_{k-1}$, i.e., we set

$$\tilde{\sigma}_k(j) = \begin{cases} \tilde{\sigma}_{k-1}(j) & \text{for } \tilde{\sigma}_{k-1}(j) < \tilde{\sigma}_{k-1}(i) \\ \tilde{\sigma}_{k-1}(j) - 1 & \text{for } \tilde{\sigma}_{k-1}(j) > \tilde{\sigma}_{k-1}(i) \\ n & \text{for } j = i \,. \end{cases}$$

Then, we remove all edges incident to i from $\tilde{G}_{k-1}$ and generate randomly new edges incident to it, so that the vector

$$(d^i(\tilde{\sigma}_k^{-1}(1)), d^i(\tilde{\sigma}_k^{-1}(2)), \ldots, d^i(\tilde{\sigma}_k^{-1}(n-1)))\,,$$

where $d^i(\tilde{\sigma}_k^{-1}(\ell))$, $\ell = 1, 2, \ldots, n-1$, counts the number of edges joining i and vertex $\tilde{\sigma}_k^{-1}(\ell)$, is distributed with distribution $\mathbf{X}_{n-1}$. Thus, roughly speaking, to get $\tilde{G}_k$ we delete a random vertex, update edges of the remaining vertices accordingly, add a new vertex n, and join it to the other vertices according to the distribution $\mathbf{X}_{n-1}$. Note that the weight of a vertex v depends only on its position $\tilde{\sigma}_k(v)$ in permutation $\tilde{\sigma}_k$; we call $\tilde{\sigma}_k(v)$ the *rank* of v. Finally, we modify the set A_k of the vertices of $\tilde{G}_k$ that have been renewed so far, putting $A_k = A_{k-1} \cup \{i\}$.

Let

$$L = \min\{k \colon A_k = [n]\}\,.$$

The *protean process* $\mathfrak{P}(\mathbf{X}_{n-1})$ is defined as the Markov chain $\{(G_i, \sigma_i)\}_{i=0}^{\infty}$ whose stages are pairs (G_i, σ_i), where $G_i = \tilde{G}_{i+L}$ and $\sigma_i = \tilde{\sigma}_{i+L}$. Note that the chain $\mathfrak{P}(\mathbf{X}_{n-1}) = \{(G_i, \sigma_i)\}_{i=0}^{\infty}$ is already in the stationary distribution, i.e., the distribution determined by G_i on the set of all ordered graphs with vertex set $[n]$ is identical for all $i \geq 0$. Furthermore, this distribution does not depend on the choice of G or σ. The random graph corresponding to this distribution is a *protean graph.* In order to make some of the notation below slightly easier, we shall assume that the ranks of the vertices of the protean graph coincide with their labels, i.e., we set

$$\tilde{L} = \min\{i \geq L : \tilde{\sigma}_i \text{ is an identity}\}$$

and identify the protean graph $\mathcal{P}(\mathbf{X}_{n-1})$ with $G_{\tilde{L}}$.

Clearly, the behaviour of $\mathcal{P}(\mathbf{X}_{n-1})$ is largely affected by the distributions of $\mathbf{X}_{n-1}$. Thus, for instance, if every coordinate of $\mathbf{X}_{n-1}$ has the binomial distribution $B(1, p)$, then $\mathcal{P}(\mathbf{X}_{n-1})$ can be identified with a standard random graph model $G(n, p)$. Another simple distribution is when a renewed vertex selects at random precisely d neighbours from the remaining $n-1$ vertices. This model is somewhat similar to the models of peer-to-peer networks, where a new vertex is connected to some number of cache vertices; in the case of peer-to-peer networks, however, one needs to establish additional links to ensure that the network is connected (see, for instance, [Pandurangan et al. 01]). Here, we concentrate on certain multigraph variants of the above model, which can be used to study (undirected) web graphs.

Let $n, d \in \mathbb{N}$ and $\eta \geq 0$. We shall consider only a special type of protean (multi)graphs with vertex set $[n]$, where in each step a new vertex chooses a neighbour among the existing vertices d times, and in each of these d independent choices each vertex v is chosen with the probability proportional to the rank of v raised to $-\eta$ (e.g., we assume that old vertices of small ranks are more attractive to newcomers). As we see shortly (Theorem 4.2), the distribution of the degrees of the vertices in these graphs also obeys the power law, provided that $\eta \in (0,1)$. In order to make this definition precise, for $1 \leq s \leq n-1$ and

$$\delta_s = (0, 0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^{n-1} \qquad (1 \text{ at the } s\text{th position}),$$

let $\mathbf{X}_{n-1}^{\eta}$ and $\mathbf{X}_{n-1}^{\eta,d}$ be nonnegative integer-valued random variables such that

$$\mathbb{P}(\mathbf{X}_{n-1}^{\eta} = \delta_s) = s^{-\eta} \Big/ \sum_{i=1}^{n-1} i^{-\eta} .$$

Furthermore, for $d \in \mathbb{N}$, let $\mathbf{X}_{n-1}^{\eta}(i)$, $i = 1, 2, \ldots, d$, denote independent copies of $\mathbf{X}_{n-1}^{\eta}$, and finally let

$$\mathbf{X}_{n-1}^{\eta,d} = \sum_{i=1}^{d} \mathbf{X}_{n-1}^{\eta}(i) .$$

Then, by $\mathcal{P}_n(d, \eta)$ we denote the protean graph $\mathcal{P}(\mathbf{X}_{n-1}^{\eta,d})$, while for the protean process $\mathfrak{P}(\mathbf{X}_{n-1}^{\eta,d})$ we write $\mathfrak{P}_n(d, \eta) = \{(\mathcal{P}_n^t(d, \eta), \sigma_t)\}_{t=0}^{\infty}$.

3. Basic Lemma

In this section we introduce the main tool in dealing with protean graphs: Lemma 3.5, which, roughly speaking, states that $\mathcal{P}_n(d, \eta)$ is, in a way, related to a random graph on the set of vertices $[n] = \{1, 2, \ldots, n\}$, in which a pair of vertices i, j, $1 \leq i < j \leq n$, is adjacent with probability $(1-\eta)\frac{d}{n}(\frac{j}{i})^{\eta}$, independently for each such pair.

Let $\tilde{\xi} = \{\xi_i\}_{i=-\infty}^{0}$ denote a random sequence of integers, where $\xi_i \in [n]$ and $\mathbb{P}(\xi_i = r) = 1/n$ for each $i \le 0$ and $1 \le r \le n$ (the ξ_i correspond to the labels of the vertices as they are renewed during the process). For $i \in [n]$ and $t \le 0$, we define $T(t,i)$ by setting

$$T(t,i) = \max\{j \le t : \xi_j = i\};$$

if such a j does not appear in $\{\xi_i\}_{i=-\infty}^{t}$ at all, which happens with probability 0, we put $T(t,i) = -\infty$. Now, let $\hat{\xi} = \{\xi_i\}_{i=-\infty}^{0}$ be a random sequence obtained from $\tilde{\xi}$ by conditioning on the event that

$$-\infty < T(0,1) < T(0,2) < \cdots < T(0,n) = 0\,.$$

Let $\sigma_n^{(j)}$ denote a permutation of the set $[n]$ obtained from a uniform random permutation by conditioning on the event that the elements $1,\ldots,j$ appear in it in the correct order and $\sigma_n^{(j)}(j) = n$, i.e., that

$$\sigma_n^{(j)}(1) < \sigma_n^{(j)}(2) < \cdots < \sigma_n^{(j)}(j) = n\,. \tag{3.1}$$

Our further argument is based on the following elementary observation, which states that at the moment when the jth vertex of $\mathcal{P}_n(d,\eta)$ is renewed for the last time, the rank of vertex i has been given by $\sigma_n^{(j)}(i)$.

Fact 3.1. *Let $j \in [n]$. Define a permutation $\bar{\sigma}^j$, setting $\bar{\sigma}^j(i) = k$, whenever the kth smallest element in the sequence*

$$T(T(0,j),1), T(T(0,j),2), \ldots, T(T(0,j),n)$$

is equal to $T(T(0,j),i)$. Then, $\bar{\sigma}^j$ has the same distribution as the random permutation $\sigma_n^{(j)}$.

Let us start with the distribution of the random variable $\sigma_n^{(j)}(i)$.

Fact 3.2. *For $1 \le i < j \le n$ and $i \le k \le n-j+i$, we have*

$$\mathbb{P}(\sigma_n^{(j)}(i) = k) = \frac{\binom{k-1}{i-1}\binom{n-k-1}{j-i-1}}{\binom{n-1}{j-1}}\,. \tag{3.2}$$

In particular, $\mathbb{E}\sigma_n^{(j)}(i) = in/j$.

Proof. The number of permutations $\sigma : [n] \to [n]$ for which (3.1) holds is equal to $\binom{n-1}{j-1}(n-j)!$; among them there are $\binom{k-1}{i-1}\binom{n-k-1}{j-i-1}(n-j)!$ for which we have

$\sigma(i) = k$. Hence (3.2) holds, and

$$\mathbb{E}\sigma_n^{(j)}(i) = \sum_k k \frac{\binom{k-1}{i-1}\binom{n-k-1}{j-i-1}}{\binom{n-1}{j-1}} = \frac{in}{j} \sum_k \frac{\binom{k}{i}\binom{n-k-1}{j-i-1}}{\binom{n}{j}} = \frac{in}{j}.$$

The last equation follows from the fact that $\binom{k}{i}\binom{n-k-1}{j-i-1}/\binom{n}{j}$ is a hypergeometric probability function, so the very last sum is equal to one. ☐

It turns out that the random variable $\sigma_n^{(j)}(i)$ is sharply concentrated around its mean.

Fact 3.3. *Let $0 < \varepsilon < 1/4$, $1 \le i < j \le n$, and*

$$k_\varepsilon^{\pm} = \frac{in}{j}(1 \pm \varepsilon)\,.$$

Then,

$$\mathbb{P}(\sigma_n^{(j)}(i) \le k_\varepsilon^-) \le 3i \exp\Big(- \frac{\varepsilon^2 i}{4}\Big)\,,$$

and

$$\mathbb{P}(\sigma_n^{(j)}(i) \ge k_\varepsilon^+) \le 2j \exp\Big(- \frac{\varepsilon^2 i}{12}\Big)\,.$$

Proof. Note that

$$\mathbb{P}(\sigma_n^{(j)}(i) = k) = \frac{\binom{k-1}{i-1}\binom{n-k-1}{j-i-1}}{\binom{n-1}{j-1}} = \frac{i(j-i)n}{jk(n-k)} \frac{\binom{k}{i}\binom{n-k}{j-i}}{\binom{n}{j}}\,.$$

Thus, one can estimate the above probability by applying a well-known bounds for the tails of the hypergeometric distribution (see, for instance, Theorem 2.10 in [Janson et al. 00]). Since the calculations are standard, we omit the technical details. ☐

Now let us consider a generalization of the well-known balls-into-bins model, which will be useful to prove Lemma 3.5. Suppose that we sequentially put d balls into m bins by placing each ball into a bin independently and the probability that we choose a bin k, $1 \le k \le m$, is equal to ρ_k, where $\sum_{i=1}^m \rho_k = 1$. Let $S_1, S_2 \subseteq [m]$, $S_1 \cap S_2 = \emptyset$, $|S_1| \le d$, and $p(S_1, S_2)$ denote the probability that every bin from set S_1 has at least one ball and bins from set S_2 have no balls.

Fact 3.4. *Using the notations described above, we have*

$$p(S_1, S_2) \ge (1 - \sum_{j \in S_1 \cup S_2} \rho_j)^{d-|S_1|} d(d-1)\dots(d-|S_1|+1) \prod_{i \in S_1} \rho_i$$

and

$$p(S_1, S_2) \le (1 - \sum_{j \in S_2} \rho_j)^{d-|S_1|} d(d-1)\dots(d-|S_1|+1) \prod_{i \in S_1} \rho_i \,.$$

Proof. In the first inequality we estimate $p(S_1, S_2)$ by the probability that each bin from S_1 contains precisely one ball; in the second we count some configurations more than once. ☐

Now we state the main result of this section, Lemma 3.5. Although its precise formulation is rather technical, the lemma is a straightforward consequence of the definition of $\mathcal{P}_n(d,\eta)$ and the large deviation inequalities stated in Fact 3.3.

Let $0 < \eta < 1$, $d \in \mathbb{N}$, and

$$E_1, E_2 \subseteq \{\{i,j\} : \log^3 n < i < j \le n\}, \quad E_1 \cap E_2 = \emptyset \,.$$

For every $i, j \in [n]$ and $r = 1, 2$, let

$$V_r(j) = \{i : i < j \text{ and } \{i,j\} \in E_r\} \,,$$

$$w(i,j) = (1-\eta)\frac{1}{n}\Big(\frac{j}{i}\Big)^{\eta} = \big(1 + O\big(n^{\eta-1}\big)\big)\frac{(in/j)^{-\eta}}{\sum_{s=1}^{n} s^{-\eta}} \,,$$

and

$$w_r(j) = \sum_{i \in V_r(j)} w(i,j) \,.$$

Then, the following holds.

Lemma 3.5. *Let $0 < \eta < 1$, d, E_1, E_2, $V_1(j)$, $w(i,j)$, $w_1(j)$, and $w_2(j)$ be defined as above, and let $|V_1(j)| \le d$ for every $j \in [n]$.*

Let $P_n(E_1, E_2, d, \eta)$ denote the probability that all pairs from E_1 are edges of $\mathcal{P}_n(d,\eta)$ and no pair from E_2 is an edge of $\mathcal{P}_n(d,\eta)$. Then,

$$\begin{aligned} P_n(E_1, E_2, d, \eta) \le{} & o(\exp(-\log^{3/2} n)) \\ & + \prod_{j=1}^{n} [1 - (1 + O(\log^{-1/2} n)) w_2(j)]^{d-|V_1(j)|} \\ & \times d(d-1)\dots(d-|V_1(j)|+1) \prod_{i \in V_1(j)} (1 + O(\log^{-1/2} n)) w(i,j), \end{aligned}$$

and

$$\begin{aligned}P_n(E_1, E_2, d,\eta) \geq\ & o(\exp(-\log^{3/2} n))\\ &+ \prod_{j=1}^{n} [1-(1+O(\log^{-1/2} n))(w_1(j)+w_2(j))]^{d-|V_1(j)|}\\ &\times d(d-1)\ldots(d-|V_1(j)|+1) \prod_{i\in V_1(j)} (1+O(\log^{-1/2} n))w(i,j).\end{aligned}$$

Proof. As we have already mentioned, Lemma 3.5 is a simple consequence of Facts 3.1, 3.3, and 3.4. Indeed, Fact 3.1 says that at the moment when vertex j is renewed for the last time, the rank of vertex i has the same distribution as the random variable $\sigma_n^{(j)}(i)$. Let $\varepsilon = \log^{-1/2} n$. Fact 3.3 implies that with probability $1 - o(\exp(-\log^{3/2} n))$ for every i, j, $\log^3 n \leq i < j \leq n$, the rank of i at the moment when j is refreshed for the last time is contained between k_ε^- and k_ε^+. Now the assertion follows from Fact 3.4. □

An analogous result for $\eta = 0$ has a particularly simple form. Indeed, in this case we have $w(i,j) = 1/n$ for every $1 \leq i < j \leq n$, and so we do not need to invoke Fact 3.3.

Lemma 3.6. *Let d, E_1, E_2, $V_1(j)$, and $V_2(j)$ be defined as above, and let $|V_1(j)| \leq d$ for every $j \in [n]$.*

Let $P_n(E_1, E_2, d, 0)$ denote the probability that all pairs from E_1 are edges of $\mathcal{P}_n(d, 0)$ and no pair from E_2 is an edge of $\mathcal{P}_n(d, 0)$. Then,

$$\begin{aligned}P_n(E_1, E_2, d, 0) \leq\ & \prod_{j=1}^{n} (1-|V_2(j)|/n)^{d-|V_1(j)|}\\ &\times\ d(d-1)\ldots(d-|V_1(j)|+1)n^{-|V_1(j)|},\end{aligned}$$

and

$$\begin{aligned}P_n(E_1, E_2, d, 0) \geq\ & \prod_{j=1}^{n} (1-|V_1(j)|/n-|V_2(j)|/n)^{d-|V_1(j)|}\\ &\times\ d(d-1)\ldots(d-|V_1(j)|+1)n^{-|V_1(j)|}.\end{aligned}$$

From the above lemmas it follows that the behaviour of the protean graph $\mathcal{P}_n(d, \eta)$ is related to that of the random graph with vertex set $[n]$ in which two vertices i, j, $\log^3 n \le i < j \le n$, are adjacent with probability

$$p(i,j) = dw(i,j) = (1-\eta)\frac{d}{n}\left(\frac{j}{i}\right)^{\eta}, \tag{3.3}$$

independently for each such pair.

Indeed, if $|V_1(j)| = o(d)$ for every $j \in [n]$, then Lemma 3.5 gives

$$\begin{aligned} P_n(E_1, E_2, d, \eta) &\sim \prod_{j=1}^{n} \Big(1 - \sum_{i \in V_2(j)} w(i,j)\Big)^{d} d^{|V_1(j)|} \prod_{i \in V_1(j)} w(i,j) \\ &= (1+o(1)) \exp\Big(- \sum_{\{i,j\} \in E_2} p(i,j)\Big) \prod_{\{i,j\} \in E_1} p(i,j)\,, \end{aligned}$$

whereas if we consider a graph with independent edges, the probability that an analogous event holds is equal to

$$\begin{aligned} &\prod_{\{i,j\} \in E_2} \Big(1 - p(i,j)\Big) \prod_{\{i,j\} \in E_1} p(i,j) \\ &\qquad = (1+o(1)) \exp\Big(- \sum_{\{i,j\} \in E_2} p(i,j)\Big) \prod_{\{i,j\} \in E_1} p(i,j)\,. \end{aligned}$$

Lemmas 3.5 and 3.6 state that the protean graph behaves just like the graph in which two vertices i, j, $i < j$, are adjacent with probability $p(i,j)$, independently for each pair. We prove that for some properties this is indeed the case. However, since we claim nothing about edges between "small vertices" i, $1 \le i < \log^3 n$, we cannot show a general theorem which relates, say, monotone properties of our model with the one with independent edges (as it is done, for instance, in [Chung and Lu 04]). Nonetheless, our lemmas are strong enough to show that any property of the independent model which, roughly speaking, does not depend on the behaviour of the first $\log^3 n$ vertices *and can be proved by computing moments* holds also for the protean graph (see Theorems 5.1 and 5.2). We also emphasize that, in our opinion, it is its *dynamical behaviour* (cf., Theorem 5.3) which makes the protean graph model interesting.

4. Degrees of Vertices

In this section we study the shape of degrees sequence of $\mathcal{P}_n(d, \eta)$. Let us start with the following result on the expected degree of vertex i.

Theorem 4.1. *Let* $0 \leq \eta < 1$ *and* $d = o\big(n^{(1-\eta)/2}\big)$. *Then, the expected degree of a vertex* $i = i(n)$, $\log^4 n \leq i \leq n$, *is given by*

$$\mathbb{E}d(i) = (1+o(1))d\frac{1-\eta}{1+\eta}\Big(\Big(\frac{n}{i}\Big)^{\eta} + \frac{2\eta}{1-\eta}\frac{i}{n}\Big).$$

Proof. Let $d^{<}(i)$ and $d^{>}(i)$ denote the number of neighbours j of i such that $j < i$ and $j > i$, respectively. We look first at the expectation of $d^{<}(i)$. It is easy to see that in every moment of the process any set of $\log^3 n$ vertices has the total weight $O(n^{\eta-1}\log^3 n)$; thus, the expected number of neighbours of i among the first $\log^3 n$ vertices is $o(1)$. Furthermore, from Fact 3.3 we infer that with probability $1-o(\exp(-\log^{3/2} n))$ the total weight of all vertices j, $\log^3 n \leq j < i$, is

$$(1+O(\log^{-1/2} n)) \sum_{j=\log^3 n}^{i} w(i,j) = (1+O(\log^{-1/2} n))i/n\,,$$

so $\mathbb{E}d^{<}(i) = (1+o(1))di/n + o(1)$.

On the other hand, one can use Lemma 3.5 (see also the remark made at the end of the previous section) to show that

$$\begin{aligned}\mathbb{E}d^{>}(i) &= \big(1+o(1)\big)\sum_{j=i+1}^{n} p(i,j)\\ &= \big(1+o(1)\big)d(1-\eta)\Big(\frac{n}{i}\Big)^{\eta}\int_{i/n}^{1} x^{\eta}dx\\ &= \big(1+o(1)\big)d\frac{1-\eta}{1+\eta}\Big(\Big(\frac{n}{i}\Big)^{\eta} - \frac{i}{n}\Big)\,.\end{aligned}$$

Since $d(i) = d^{<}(i) + d^{>}(i)$, the assertion follows. □

Note that for $\eta \in (0,1)$ the above expectation is minimized for $i = (1+o(1))c_{\min}(\eta)n$, where the constant

$$c_{\min}(\eta) = [(1-\eta)/2]^{1/(\eta+1)}$$

depends only on η but not on d. We comment on this fact in the following section, devoted to the threshold for the connectivity for $\mathcal{P}_n(d,\eta)$.

Observe also that, for small i, the expected degree of a vertex i is dominated by the factor $d\,\frac{1-\eta}{1+\eta}\left(\frac{n}{i}\right)^{\eta}$. Consequently, the degrees of the protean graph $\mathcal{P}_n(d,\eta)$ are distributed according to the power law. More specifically, let $Z_k = Z_k(n;d;\eta)$ denote the number of vertices of degree k in $\mathcal{P}_n(d,\eta)$ and $Z_{\geq k} = \sum_{\ell \geq k} Z_\ell$. Here and below *a.a.s.* means "with probability tending to 1 as $n \to \infty$."

Theorem 4.2. *Let* $d \in \mathbb{N}$, $0 < \eta < 1$, $\log^2 n \leq k = k(n) = O(n^\eta / \log^3 n)$, *and* $d = d(n) = o(\log^2 n)$. *Then, a.a.s.*

$$Z_{\geq k} = (1 + o(1))n\Big(\frac{1-\eta}{1+\eta} \cdot \frac{d}{k}\Big)^{1/\eta} + O(\log^3 n)\,.$$

Proof. We just outline an argument, omitting technical details. Let $Y_{i,j}$, $\log^3 n \leq i < j \leq n$, be a family of independent random variables such that

$$\mathbb{P}(Y_{i,j} = 1) = 1 - \mathbb{P}(Y_{i,j} = 0) = p(i,j),$$

where $p(i,j)$ is defined by (3.3). Then, from Lemma 3.5 it follows that the probability that i has k neighbours among vertices j, with $j > i$, can be well approximated by the probability that $\sum_{j>i} Y_{i,j} = k$. One can use this fact also to estimate tails of the distribution and show that large deviation inequalities for $\sum_{j>i} Y_{i,j}$ imply that for every $\varepsilon > 0$ a.a.s. all vertices i such that

$$i \geq (1+\varepsilon)n\Big(\frac{1-\eta}{1+\eta} \cdot \frac{d}{k}\Big)^{1/\eta}$$

has fewer than $k - d = (1+o(1))k$ neighbours among vertices $j > i$, and each vertex i for which

$$i \leq (1-\varepsilon)n\Big(\frac{1-\eta}{1+\eta} \cdot \frac{d}{k}\Big)^{1/\eta}$$

has more than k neighbours. Since the vertex i has at most $d = o(k)$ neighbours j with $j < i$, the assertion follows. $\square$

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

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

Thus, to imitate this distribution, one should set the parameter η of the protean graph $\mathcal{P}_n(d,\eta)$ to be equal to $\eta \sim 1/(\gamma - 1)$. For the web graph the distribution of total degrees is, at this moment, unknown. However, the number of vertices of in-degree k decreases roughly as $k^{-2.1}$, while the fraction of vertices of out-degree k can be approximated by $k^{-2.7}$ (cf., [Broder et al. 00]). Thus, if the total degree of the graph is close to the distribution of in-degree, then it can be approximated by the protean graph with $\eta_{\rm www} \sim 0.91$.

5. Connectivity

Theorem 4.2 shows that protean graphs can be treated as a (stationary) model of web graphs; some properties interesting for this type of application such as the diameter and the size of the largest component of $\mathcal{P}_n(d,\eta)$ will be treated in the forthcoming paper [Prałat 06] (see also [Prałat 04]). In this section we study the connectivity of $\mathcal{P}_n(d,\eta)$ to illustrate similarities and differences both in results and methods between protean graphs and the standard binomial random graph model $G(n,p)$.

Let $\rho_n(d,\eta)$ denote the probability that $\mathcal{P}_n(d,\eta)$ is connected. First, we deal with the simplest case $\eta = 0$. Then, all vertices have the same weight, and the probability that two vertices are connected by an edge is given by

$$\bar{p}(i,j) = \hat{p}(n) = 1 - (1-1/n)^d = d/n + O(d^2/n^2)\,.$$

Thus, one should expect that the threshold function for connectivity is the same as in the binomial random graph model $G(n,\hat{p})$. Theorem 5.1 shows that it is roughly the case, but the dependence structure of $\mathcal{P}_n(d,0)$ influences the second term of the threshold function.

Theorem 5.1. *Let* $d = d(n) = \log n - \frac{1}{2}\log\log n + a(n)$, *with* $a(n) = o(\log\log n)$. *Then,*

$$\lim_{n\to\infty} \rho_n(d,0) = \begin{cases} 1 & \text{if} \quad a(n) \to \infty \\ \exp\left(-\sqrt{\pi/2}e^{-a}\right) & \text{if} \quad a(n) \to a \\ 0 & \text{if} \quad a(n) \to -\infty\,. \end{cases}$$

Before we prove Theorem 5.1, let us remark that Lemma 3.6 (and Lemma 3.5) are not strong enough to deduce Theorem 5.1 (and the following Theorem 5.2) directly from the corresponding result for an independent model. However, they are sufficient to compute all the moments of an appropriate random variables, so the proof for an independent model can be mimicked in this case.

Proof. Note first that for the expectation of the number Y_n of isolated vertices in $\mathcal{P}_n(d,0)$ we have

$$\begin{aligned} &\sum_{i=1}^{n}\left(1-\frac{i-1}{n}\right)^d\left(1-\frac{1}{n}\right)^{d(n-i)} \\ &\qquad = (1+o(1))ne^{-d}\int_0^1 \exp\left(-\frac{x^2 d}{2} + O(x^3 d)\right)dx \\ &\qquad = (1+o(1))\sqrt{\frac{\pi}{2d}}ne^{-d} = (1+o(1))\sqrt{\frac{\pi}{2}}e^{-a}\,. \end{aligned}$$

One can also check that, for a given $r \geq 2$, the rth factorial moment of Y_n tends to $(\pi/2)^{r/2}e^{-ra}$, so the Y_n tends to a Poisson distribution and, in particular, the probability that $\mathcal{P}_n(d,0)$ contains no isolated vertices tends to $\exp(-\sqrt{\pi/2}e^{-a})$ as $n \to \infty$.

Thus, it is enough to show that if, say, $d(n) > 0.99 \log n$, the protean graph consists of one giant component and, perhaps, some number of isolated vertices. The probability that $\mathcal{P}_n(d,0)$ contains a component of size k, where $2 \leq k \leq 2n/3$, is, by Lemma 3.6 (see also the remark made after this lemma), bounded from above by

$$\sum_{k=2}^{2n/3} \binom{n}{k} k^{k-2} (1-d/n)^{(k-2k/\sqrt{d})(n-k)} (d/n)^{k-1}$$

and tends to 0 as $n \to \infty$. (Since at most $2k/\sqrt{d}$ vertices from a spanning tree have degree more than $\sqrt{d}$, we can estimate the probability that the vertices from a tree have no neighbours outside this component by $(1-d/n)^{(k-2k/\sqrt{d})(n-k)}$.) Thus, a.a.s. the protean graph consists of a giant component and some number of isolated vertices, which completes the proof of the theorem. □

In the case when $\eta \in (0,1)$, the threshold for the connectivity is affected by a constant factor.

Theorem 5.2. *Let $\eta \in (0,1)$ and $d = d(n) = a \log n$, where a is a positive constant. Then,*

$$\lim_{n\to\infty} \rho_n(d,\eta) = \begin{cases} 1 & \text{if} \quad a > 1/g(x_0(\eta)) \\ 0 & \text{if} \quad a < 1/g(x_0(\eta)), \end{cases}$$

where

$$g(x) = \frac{1-\eta}{1+\eta}(x^{-\eta} - x) - \log(1-x) \tag{5.1}$$

and x_0 is the value of x that minimizes $g(x)$ in the interval $(0,1)$, i.e., x_0 is the root of equation

$$(1-\eta)\eta x^{-1-\eta} + 1 - \eta = \frac{1+\eta}{1-x}. \tag{5.2}$$

Proof. The proof basically follows the argument that we use to show Theorem 5.1. We first find the expectation of the number of Y_n of the isolated vertices in $\mathcal{P}_n(d,\eta)$. Using Lemma 3.5, and observing that the probability that a vertex is connected to one of the first $\log^3 n$ vertices is negligible, we get

$$\begin{aligned}
\mathbb{E}Y_n &= o(\exp(-\log^{3/2} n)) \\
&\quad + \sum_{i=1}^{n} \Big(1-(1+O(\log^{-1/2} n))\sum_{j=1}^{i-1}\frac{1-\eta}{n}\Big(\frac{i}{j}\Big)^{\eta}\Big)^{d} \\
&\quad \times \prod_{j=i+1}^{n}\Big[1-(1+O(\log^{-1/2} n))(1-\eta)\frac{d}{n}\Big(\frac{j}{i}\Big)^{\eta}\Big] \\
&= (1+o(1))n\int_0^1 \Big[(1-(1+O(\log^{-1/2} n))x)^d \\
&\qquad\qquad \times \exp\Big(-(1+o(1))\frac{1-\eta}{1+\eta}(x^{-\eta}-x)\Big)^{d}\Big]\,dx \\
&= n^{1+o(1)}\int_0^1 \exp\big(-(1+o(1))\,g(x)\,d\big)\,dx\,.
\end{aligned} \tag{5.3}$$

It is not hard to check that $g''(x)>0$ for $x\in(0,1)$ and $\lim_{x\to 0+} g'(x)=-\infty$ while $\lim_{x\to 1-} g'(x)=\infty$. Hence, $g(x)$ has a unique minimum in the point $x_0(\eta)$, which is the root of Equation (5.2).

If $a<1/g(x_0(\eta))$, then for some $\varepsilon>0$ we have $a<1/g(x)$ for each $x\in(x_0(\eta)-\varepsilon, x_0(\eta)+\varepsilon)$. Consequently,

$$\mathbb{E}Y_n \ \geq\ 2\varepsilon n^{1+o(1)}\exp\big(-(1+o(1))a\log n \max_{x\in(x_0(\eta)-\varepsilon,x_0(\eta)+\varepsilon)}\{g(x)\}\big)\ \to\ \infty\,.$$

One can use Lemma 3.5 to show that for such a we have $\mathrm{Var}Y_n=o((\mathbb{E}Y_n)^2)$ so, if $a<1/g(x_0(\eta))$, then $\mathcal{P}_n(d,\eta)$ a.a.s. contains an isolated vertex.

Now, suppose that $d(n)=a\log n$ and $a>1/g(x_0(\eta))$; then,

$$\mathbb{E}Y_n \leq n^{1+o(1)}\exp(-(1+o(1))ag(x_0)\log n)=o(1)\,,$$

and so a.a.s. $\mathcal{P}_n(d,\eta)$ contains no isolated vertices. Thus, in order to show the assertion, we have to check that for some small $\delta>0$ and $d(n)>(1/g(x_0(\eta))-\delta)\log n$, the protean graph $\mathcal{P}_n(d,\eta)$ consists of a giant component and, perhaps, some number of isolated vertices.

Let c be a small positive constant to be chosen later, and let H denote the subgraph of $\mathcal{P}_n(d,\eta)$ induced by vertices i, $cn<i\leq n$. One can use Lemma 3.5 to show that a.a.s. H consists of one large component and $o(n^{0.5})$ isolated vertices. Again, we just outline the proof, omitting technical details. Estimating the number of isolated vertices $\hat{Y}_n$ in H the same way as $\mathbb{E}Y_n$ above, we infer that $\mathbb{E}\hat{Y}_n=o(n^{0.4})$, and so a.a.s. $\hat{Y}_n=o(n^{0.5})$ by Markov's inequality. Using calculations similar to those in (5.3), one can show that the probability that a

vertex $i = \lceil xn \rceil$, $c < x \le 1$, from a spanning tree has no neighbours outside this component of a fixed size which belong to the subgraph H is equal to

$$\begin{aligned}
&(1+o(1))\Big(1-(1+O(\log^{-1/2} n))\sum_{j=\lceil cn\rceil}^{i-1}\frac{1-\eta}{n}\Big(\frac{i}{j}\Big)^{\eta}\Big)^{d}\\
&\qquad\qquad \times \prod_{j=i+1}^{n}\Big[1-(1+O(\log^{-1/2} n))(1-\eta)\frac{d}{n}\Big(\frac{j}{i}\Big)^{\eta}\Big]\\
&= (1+O(\log^{-1/2} n))^{d}(1-x+c^{1-\eta}x^{\eta})^{d}\\
&\qquad\qquad \times \exp\Big(-(1+o(1))\frac{1-\eta}{1+\eta}d(x^{-\eta}-x)\Big)
\end{aligned}$$

and tends to $n^{o(1)}\exp[-(1+o(1))\,d\,g(x_0(\eta))]$ as $c \to 0$. Thus, this probability is smaller than, say, $n^{o(1)}\exp[-0.75\,d\,g(x_0(\eta))]$ for c small enough. From this we can show that the probability that H contains a component of size k, where $2 \le k \le k_0 = O(1)$, is bounded from above by

$$\sum_{k=2}^{k_0}\binom{(1-c)n}{k}k^{k-2}\Big(n^{o(1)}\exp[-0.75\,d\,g(x_0(\eta))]\Big)^{k}\Big(\frac{d}{cn}\Big)^{k-1} \tag{5.4}$$

and tends to 0 as $n \to \infty$, for δ small enough. Moreover, the expectation of the number of components of H that are larger than k_0 and smaller than $2n/3$ is, for c small enough, bounded from above by

$$\sum_{k=k_0+1}^{2n/3}\binom{(1-c)n}{k}k^{k-2}\Big(1-(1-\eta)\frac{d}{2n}\Big)^{(k-2k/\sqrt{d})((1-c)n-k)}\Big(\frac{d}{cn}\Big)^{k-1} \tag{5.5}$$

and tends to 0 as $n \to \infty$, provided that δ is small enough. Notice that in order to show (5.4) and (5.5) we cannot use Lemma 3.5 since the error term, following from the concentration of random variables, is too big. Instead of this we are using the universal upper bounds for the probabilities that an edge is present or absent in the subgraph H. For example, the probability that an edge is present in H is smaller than or equal to the probability that $\{cn, n\}$ is an edge in G.

Finally, one can show that a.a.s. every vertex $i < cn$ is connected with the large component in H, completing the proof of Theorem 5.2. $\square$

Let us note that, as follows from the argument in the proof, a.a.s. near the threshold all isolated vertices have labels $(1+o(1))x_0(\eta)n$. This is more evidence of the *middle-age crisis* of vertices of the protean graph that we pointed out in the remark following Theorem 4.2: the vertices of medium labels are of smallest degrees since they have lost their old neighbours, which have already been renewed, yet they are not old enough to attract new ones. We also remark that the minimum $c_{\min}(\eta)$ related to the expected degrees of vertices is not equal to $x_0(\eta)$, which identifies the labels of the isolated vertices at the connectivity threshold. For instance, for the *internet constant* $\eta_{\mathrm{www}} = 0.91$, we have $c_{\min}(\eta_{\mathrm{www}}) \sim 0.197$ while $x_0(\eta_{\mathrm{www}}) \sim 0.177$.

We conclude this paper with one more result on the protean process $\mathfrak{P}_n(d,\eta) = \{(\mathcal{P}_n^t(d,n),\sigma_t)\}_{t=0}^{\infty}$ that does not have its counterpart for the random graph process $\{G(n,p)\}_{0\le p\le 1}$. Let $\mathcal{A}$ be a graph property such that $\mathcal{A}$ holds for $\mathcal{P}_n(d,\eta)$ a.a.s. but for $\tau(\mathcal{A})$, defined as

$$\tau(A) = \min\{t : \mathcal{P}_n^t(d,\eta) \text{ has not } \mathcal{A}\}\,,$$

we have $\mathbb{P}(\tau(\mathcal{A}) < \infty) = 1$, i.e., with probability one, at some stage of the protean process $\mathfrak{P}_n(d,\eta) = \{(\mathcal{P}_n^t(d,n),\sigma_t)\}_{t=0}^{\infty}$ the property $\mathcal{A}$ disappears for some time. Then, a *recovery time* $\mathrm{rec}(\mathcal{A})$ for $\mathcal{A}$ is defined as

$$\mathrm{rec}(\mathcal{A}) = \min\{t > \tau(\mathcal{A}) : \mathcal{P}_n^t(d,\eta) \text{ has } \mathcal{A}\} - \tau(\mathcal{A})\,,$$

i.e., $\mathrm{rec}(\mathcal{A})$ tells us how long it takes for a protean graph to regain a typical property $\mathcal{A}$. Note that since $\mathcal{A}$ holds a.a.s., and after $O(n\log n)$ steps, each vertex of $\mathcal{P}_n(d,\eta)$ is renewed at least once, so a.a.s. $\mathrm{rec}(\mathcal{A}) = O(n\log n)$. However, typically, the recovery time is smaller than the above bound implied by the coupon collector problem. The following theorem estimates $\mathrm{rec}(\mathcal{C})$ in the case when the graph is connected.

Theorem 5.3. *Let $\eta \in (0,1)$ and $d = a\log n$, where $a > 1/g(x_0)$ and the function $g(x)$ and x_0 are defined as in (5.1) and (5.2), respectively. Then*

$$\mathrm{rec}(\mathcal{C}) \cdot \frac{(1-\eta)a}{(x_0)^\eta}\frac{\log n}{n} \xrightarrow{D} Z\,,$$

where the random variable Z has the exponential distribution, i.e., for every $z \ge 0$,

$$\mathbb{P}(Z \ge z) = e^{-z}\,.$$

Proof. The main part of the proof is to show that a.a.s. at the moment $\tau(\mathcal{C})$ the protean graph consists of a large component and a single isolated vertex i of rank $w = (1+o(1))x_0 n$ (note that for such vertices w the probability of being isolated is maximized). Note also that x_0 minimizes the continuous function $g(x)$ in the interval $(0,1)$, which means that it is enough to show that $g(w/n) \le$

$(1+o(1))g(x_0)$. Again, we just give a sketch of the proof, leaving out technical details.

Let us consider first $n\log^2 n$ steps of the process. The probabilities that during that time in the protean process

- an isolated vertex of the rank w, where $g(w/n) \le (1+\varepsilon)g(x_0)$,
- an isolated vertex of the rank w, where $g(w/n) > (1+\varepsilon)g(x_0)$,
- a component of size k, $2 \le k \le 2n/3$,

appears, we denote by $\rho_1(\varepsilon)$, $\rho_2(\varepsilon)$, and ρ_3, respectively. In order to estimate them, let us compute first the probability $\rho(i,j,t)$ that a vertex i becomes isolated at time t due to the fact that in this step we chose the only neighbour j of i in the preceding graph to be renewed. Let ℓ_i and ℓ_j denote the ranks of i and j, respectively, in the protean graph constructed after first $t-1$ steps. Then, arguing as in (5.3), we may estimate $\rho(i,j,t)$ by

$$\begin{aligned}
&o(n\exp(-\log^{3/2} n)) + (1+o(1))\frac{1}{n}p(\ell_i,\ell_j)\\
&\qquad\times\prod_{r=1}^{\ell_i-1}\Big[1-(1+O(\log^{-1/2} n))\frac{1-\eta}{n}\Big(\frac{\ell_i}{r}\Big)^{\eta}\Big]^d\\
&\qquad\times\prod_{s=\ell_i+1,s\neq\ell_j}^{n}\Big[1-(1+O(\log^{-1/2} n))\frac{1-\eta}{n}d\Big(\frac{s}{\ell_i}\Big)^{\eta}\Big]\\
&\quad= n^{-2+o(1)}\Big(\frac{\ell_j}{\ell_i}\Big)^{\eta}\exp\Big(-(1+o(1))g\Big(\frac{\ell_i}{n}\Big)d\Big)
\end{aligned}\tag{5.6}$$

for $\ell_i < \ell_j$ and

$$\begin{aligned}
&o(n\exp(-\log^{3/2} n)) + (1+o(1))\frac{1}{n}p(\ell_j,\ell_i)\\
&\qquad\times\prod_{r=1,r\neq\ell_j}^{\ell_i-1}\Big[1-(1+O(\log^{-1/2} n))\frac{1-\eta}{n}\Big(\frac{\ell_i}{r}\Big)^{\eta}\Big]^d\\
&\qquad\times\prod_{s=\ell_i+1}^{n}\Big[1-(1+O(\log^{-1/2} n))\frac{1-\eta}{n}d\Big(\frac{s}{\ell_i}\Big)^{\eta}\Big]\\
&\quad= n^{-2+o(1)}\Big(\frac{\ell_i}{\ell_j}\Big)^{\eta}\exp\Big(-(1+o(1))g\Big(\frac{\ell_i}{n}\Big)d\Big)
\end{aligned}\tag{5.7}$$

for $\ell_i > \ell_j$.

Let $\varepsilon > 0$ be a positive constant. Let us denote by $\mathbf{A}_t(i)$ the event that a vertex i of the rank w becomes isolated at step t of the process and $g(w/n) \le$

$(1+\varepsilon/4)g(x_0)$; moreover, let $\mathbf{A}_t = \bigcup_{i=1}^{n} \mathbf{A}_t(i)$. The events $\mathbf{B}'_t(i)$ and $\mathbf{B}'(i)$ are defined in a similar way, but this time, for the rank w of i at the moment when i becomes isolated, we want to have $g(w/n) > (1+\varepsilon)g(x_0)$. From (5.6) and (5.7) we get

$$n^{-1-(1+o(1))(1+\varepsilon/4)ag(x_0)} \leq \mathbb{P}(\mathbf{A}_t(i)) \leq n^{-1-(1+o(1))ag(x_0)},$$

while

$$\mathbb{P}(\mathbf{B}'_t(i)) \leq n^{-1-(1+o(1))(1+\varepsilon)ag(x_0)},$$

and

$$\mathbb{P}(\mathbf{B}'_t) \leq \sum_{i=1}^{n} \mathbb{P}(\mathbf{B}'_t(i)) \leq n^{-(1+o(1))(1+\varepsilon)ag(x_0)}. \tag{5.8}$$

To estimate $\mathbb{P}(\mathbf{A}_t)$ we have to notice that one can use Lemma 3.5 to prove that for $i \neq i'$ the events $\mathbf{A}_t(i)$ and $\mathbf{A}_t(i')$ are, in a way, *weakly dependent*, i.e.,

$$\mathbb{P}(\mathbf{A}_t(i) \cap \mathbf{A}_t(i')) = \mathbb{P}(\mathbf{A}_t(i))\mathbb{P}(\mathbf{A}_t(i'))n^{o(1)}\Theta(1+n^{2\eta-1}).$$

Thus, Bonferroni's inequality gives

$$\begin{aligned}\mathbb{P}(\mathbf{A}_t) &= \mathbb{P}\Big(\bigcup_{i=1}^{n} \mathbf{A}_t(i)\Big) \\ &\geq \sum_{i=1}^{n} \mathbb{P}(\mathbf{A}_t(i)) - \sum_{1\leq i<i'\leq n} \mathbb{P}(\mathbf{A}_t(i) \cap \mathbf{A}_t(i')) \geq n^{-(1+\varepsilon/3)ag(x_0)}.\end{aligned}$$

From (5.8) we get immediately that

$$\rho_2(\varepsilon) = \sum_{t=1}^{n\log^2 n} \mathbb{P}(\mathbf{B}'_t) \leq n^{1-(1+o(1))(1+\varepsilon)ag(x_0)}. \tag{5.9}$$

Again, one can use Lemma 3.5 to show that for $t_1 < t_2$

$$\mathbb{P}(\mathbf{A}_{t_1} \cap \mathbf{A}_{t_2}) = \mathbb{P}(\mathbf{A}_{t_1})\mathbb{P}(\mathbf{A}_{t_2})n^{o(1)}$$

and from Bonferroni's inequality

$$\rho_1(\varepsilon) \geq \mathbb{P}\Big(\bigcup_{t=1}^{n\log^2 n} \mathbf{A}_t\Big) \geq n^{1-(1+\varepsilon/2)ag(x_0)}. \tag{5.10}$$

Moreover, it can also be proved that

$$\rho_3 \leq n^{1+o(1)}[\mathbb{P}(\mathbf{A}_t)]^2 \leq n^{1-(1+o(1))2ag(x_0)} \leq \rho_2(\varepsilon). \tag{5.11}$$

(In order to do it, one should notice that, since the expected number of large components tends to zero quickly, it is enough to consider only the components of the bounded size. Then, use Lemma 3.5 to estimate the probability that they appear; since the argument is fairly standard, we omit the details.)

Now let us consider the first $n^{(1+\frac{3}{4})ag(x_0)}\log^2 n$ steps of the protean process. From (5.9), (5.10), and (5.11), it follows that if the graph becomes disconnected during this period then, a.a.s., it is due to the appearance of a single isolated vertex of rank w, where $g(w/n) \le (1+\varepsilon)g(x_0)$.

Let $\mathbf{D}_k$, $k=0,1,\dots,k_0$, where $k_0 = n^{(1+\frac{3}{4})ag(x_0)-1}/3$, be an event that between $2kn\log^2 n$ and $(2k+1)n\log^2 n$ steps an isolated vertex of the rank w, where $g(w/n) \le (1+\varepsilon)g(x_0)$, appears. Let $\mathbf{F}$ be the event that every vertex of the graphs was at least one time renewed in time $t \in ((2k-1)n\log^2 n, 2kn\log^2 n))$, for each $k=1,\dots,k_0$. Then, $\mathbb{P}(\mathbf{F}) \ge 1-\exp(-\log^{3/2} n)$. Moreover, $\mathbb{P}(\mathbf{D}_k) = \rho_1(\varepsilon)$, and conditioned on $\mathbf{F}$, all events $\mathbf{D}_k$ are independent. Thus, since $k_0\rho_1(\varepsilon) \to \infty$ as $n \to \infty$, we have

$$\mathbb{P}\Big(\bigcup_{k=0}^{k_0} \mathbf{D}_k\Big) \to 1.$$

Consequently, a.a.s. $\tau(\mathcal{C}) = n^{(1+o(1))ag(x_0))}$, and at the moment $\tau(\mathcal{C})$ the protean graph consists of a giant component and a single isolated vertex i_0 of rank $(1+o(1))x_0 n$.

The rest of the proof is straightforward. Let us consider the first $\Theta(n/\log n)$ steps after the moment when the graph became disconnected. The probability that we renew vertex i_0 at that time tends to 0 as $n \to \infty$, and by the argument similar to the one used to estimate $\rho_1(\varepsilon)$, $\rho_2(\varepsilon)$, and ρ_3 above, so is the probability that we create additional small component of $\mathcal{P}_n^t(d,\eta)$. Thus, the graph becomes connected if one of the renewed vertices will choose i_0 as a neighbour. Since the weight of i_0 can change only slightly during $\Theta(n/\log n)$ steps, the probability that for some $z \ge 0$

$$\mathrm{rec}(\mathcal{C}) \ge z\frac{(x_0)^\eta}{(1-\eta)a}\frac{n}{\log n} = z\frac{(x_0)^\eta}{1-\eta}\frac{n}{d}$$

is given by

$$\begin{aligned}
&\Big[1-(1+o(1))(1-\eta)\frac{d}{n}\Big(\frac{n}{x_0 n}\Big)^\eta\Big]^{z\frac{(x_0)^\eta}{1-\eta}\frac{n}{d}} \\
&\qquad = \Big[1-(1+o(1))(1-\eta)\frac{d}{n}(x_0)^{-\eta}\Big]^{z\frac{(x_0)^\eta}{1-\eta}\frac{n}{d}} \\
&\qquad = (1+o(1))e^{-z},
\end{aligned}$$

so the assertion follows. $\square$

Acknowledgements. Work for this paper was partially supported by KBN grant 1 P03A 025 27.

References

[Bollobás et al. 01] B. Bollobás, O. Riordan, J. Spencer, and G. E. Tusnády. "The Degree Sequence of a Scale-Free Random Graph Process." *Random Structures and Algorithms* 18 (2001), 279–290.

[Bollobás et al. 03] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. "Directed Scale-Free Graphs." In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 132–139. Philadelphia: SIAM, 2003.

[Broder et al. 00] A. Broder, R. Kumar, F. Maghoul, P. Rahaghavan, S. Rajagopalan, R. State, A. Tomkins, and J. Wiener. "Graph Structure in the Web." In *Proceedings of the Ninth International World Wide Web Conference*, pp. 309–320. Reston, VA: Foretec Seminars, Inc., 2000.

[Chung and Lu 04] F. Chung and L. Lu. "Coupling Online and Offline Analyses for Random Power Law Graphs." *Internet Mathematics* 1:4 (2004), 409–461.

[Cooper and Frieze 03] C. Cooper and A. Frieze. "A General Model of Web Graphs." *Random Structures and Algorithms* 22 (2003), 311–335.

[Janson et al. 00] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. New York: Wiley, 2000.

[Kumar et al. 00] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. "Stochastic Models for the Web Graph." In *Proceedings of the 41th IEEE Symposium on Foundations of Computer Science*, pp. 57–65. Los Alamitos, CA: IEEE Press, 2000.

[Pandurangan et al. 01] G. Pandurangan, P. Raghavan, and E. Upfal. "Building Low-Diameter P2P Networks." In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pp. 492–499. Los Alamitos, CA: IEEE Press, 2001

[Prałat 04] P. Prałat. "Grafy proteuszowe" ["Protean Graphs"] (in Polish). PhD dissertation, Adam Mickiewicz University, 2004.

[Prałat 06] P. Prałat. "A Note on the Diameter of Protean Graphs." Submitted to *Discrete Mathematics*, 2006.

Tomasz Łuczak, Adam Mickiewicz University, Collegium Mathematicum, ul. Umultowska 87, 61-614 Poznań, Poland (tomasz@amu.edu.pl)

Paweł Prałat, Department of Combinatorics and Optimization, University of Waterloo, Waterloo ON, N2L 3G1, Canada (ppralat@math.uwaterloo.ca)

Received September 24, 2004; accepted in revised form April 6, 2006.