It’s a Small World for Random Surfers*†

Abbas Mehrabian[1] and Nick Wormald[2]

1 **Department of Combinatorics and Optimization, University of Waterloo Waterloo, ON, Canada**
`amehrabi@uwaterloo.ca`

2 **School of Mathematical Sciences, Monash University Clayton, VIC, Australia**
`nick.wormald@monash.edu`

Abstract

We prove logarithmic upper bounds for the diameters of the random-surfer Webgraph model and the PageRank-based selection Webgraph model, confirming the small-world phenomenon holds for them. In the special case when the generated graph is a tree, we get close lower and upper bounds for the diameters of both models.

1998 ACM Subject Classification C.2.1 Network Architecture and Design, G.2.2 Graph Theory, G.3 Probability and Statistics

Keywords and phrases random-surfer webgraph model, PageRank-based selection model, small-world phenomenon, height of random trees, probabilistic analysis, large deviations

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2014.857

1 Introduction

Due to the ever growing interest in social networks, the Webgraph, biological networks, etc., in recent years a great deal of research has been built around modelling real world networks (see, e. g., the monographs [6, 8, 10, 15]). One of the important observations about many real world networks involves the diameter, which is the maximum shortest-path distance between any two nodes. The so-called *small-world phenomenon* is that the diameter of a network is significantly smaller than its size, typically growing as a polylogarithmic function.

The Webgraph is a directed graph whose vertices are the static web pages, and there is an edge joining two vertices if there is a hyperlink in the first page pointing to the second page. Barabási and Albert [1] in 1999 introduced one of the first models for the Webgraph, widely known as the *preferential attachment model*. Their model can be informally described as follows (see [5] for the formal definition). Let d be a positive integer. We start with a fixed small graph, and in each time-step a new vertex appears and is joined to d old vertices, where the probability of joining to each old vertex is proportional to its *degree*. Pandurangan, Raghavan and Upfal [20] in 2002 introduced the *PageRank-based selection model* for the Webgraph. This model is similar to the previous model, except the attachment probabilities are proportional to the *PageRanks* of the vertices rather than their degrees (see Section 2 for the formal definition). Blum, Chan, and Rwebangira [4] in 2006 introduced a *random-surfer model* for the Webgraph, in which the d out-neighbours of the new vertex are chosen by doing d independent random walks that start from random vertices and whose lengths are

* The first author was supported by the Vanier Canada Graduate Scholarships program, and the second author was supported by Australian Laureate Fellowships grant FL120100125.

† The full version of this paper is available in `http://arxiv.org/abs/1404.7189`.

17th Int’l Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX’14) / 18th Int’l Workshop on Randomization and Computation (RANDOM’14).
Editors: Klaus Jansen, José Rolim, Nikhil Devanur, and Cristopher Moore; pp. 857–871
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

geometric random variables with parameter p (see Section 2 for the formal definition). It was shown that under certain conditions (see Section 2), the previous two models are equivalent.

The directed models considered here generate directed acyclic graphs (new vertices create edges to old vertices), so it is natural to define the *diameter* of a directed graph as the maximum shortest-path distance between any two vertices in its underlying undirected graph. The diameter of the preferential attachment model was analysed by Bollobás and Riordan [5]. Previous work on the PageRank-based selection and random-surfer models has focused on their degree distributions. To the best of our knowledge, the diameters of these models have not been studied previously, and it is an open question even whether they have logarithmic diameter. One of the main contributions of this paper is giving logarithmic upper bounds for their diameters. We also give close lower and upper bounds in the special case $d = 1$, namely when the generated graph is (almost) a tree. It turns out that the key parameter in this case is the *height* of the generated random tree. We find the asymptotic value of the height for all $p \in [0.21, 1]$, and for $p \in (0, 0.21)$ we provide logarithmic lower and upper bounds. Our results hold *asymptotically almost surely (a.a.s.)*, which means the probability that they are true approaches 1 as the number of vertices grows. As the two models are equivalent, we will focus on the random-surfer model, since it is easier to work with.

1.1 Our Approach and Organization of the Paper

In the preferential attachment model and most of its variations (see, e. g., [1, 13, 14, 17]) the probability that the new vertex attaches to an old vertex v, called the *attraction* of v, is proportional to a deterministic function of the degree of v. In other variations (see, e. g., [3, 16]) the attraction also depends on the so-called 'fitness' of v, which is a random variable generated independently for each vertex and does not depend on the structure of the graph. For analysing such models when they generate trees, a typical technique is to approximate them with population-dependent branching processes and prove that results on the corresponding branching processes carry over to the original models. A classical example is Pittel [22] who estimated the height of random recursive trees. Bhamidi [2] used this technique to show that the height of a variety of preferential attachment trees is asymptotic to a constant times the logarithm of the number of vertices, where the constant depends on the parameters of the model.

In the random-surfer Webgraph model, however, the attraction of a vertex does not depend only on its degree, but rather on the graph's general structure, so the branching processes techniques cannot apply directly, and new ideas are needed.

The crucial novel idea in our proof is to reduce the attachment rule to a simple one, with the help of introducing (possibly negative) 'weights' for the edges. First, consider the general case, $d \geq 1$. Whenever a new vertex appears, it builds d new edges to old vertices; suppose that we mark the first new edge. Then the marked edges induce a spanning tree whose diameter we bound, and thus we get an upper bound for the diameter of the random-surfer Webgraph model.

In the special case $d = 1$, we obtain a *random recursive tree* with edge weights, and then we adapt a powerful technique developed by Broutin and Devroye [7] (that uses branching processes) to study its weighted height. This technique is based on large deviations. The main theorem of Broutin and Devroye [7, Theorem 1] is not applicable here for two reasons. Firstly, the weights of edges on the path from the root to each vertex are not independent, and secondly, the weights can be negative.

We define the models and state our main results in Section 2. In Section 3 we prove Theorem 2, giving a logarithmic upper bound for the diameter of the random-surfer Webgraph model in the general case $d \geq 1$. In Sections 4–6 we focus on the special case $d = 1$ and we prove Theorems 3 and 4. Section 4 contains the main technical contribution of this paper, where we explain how to transform the random-surfer tree model into one that is easier to analyse. The lower and upper bounds are proved in Sections 5 and 6, respectively. Concluding remarks appear in Section 7. All proofs omitted from this extended abstract can be found in the full version.

2 Definitions and Main Results

Given $p \in (0, 1]$, let $\mathrm{Geo}(p)$ denote a geometric random variable with parameter p; namely for every nonnegative integer k, $\mathbb{P}[\mathrm{Geo}(p) = k] = (1 - p)^k p$.

▶ **Definition 1** (Random-Surfer Webgraph Model [4])**.** Let d be a positive integer and let $p \in (0, 1]$. Generate a random directed rooted n-vertex multigraph, with all vertices having out-degree d. Start with a single vertex v_0, the root, with d self-loops. At each subsequent step s, where $1 \leq s \leq n-1$, a new vertex v_s appears and d edges are created from it to vertices in $\{v_0, v_1, \dots, v_{s-1}\}$, by doing the following probabilistic procedure d times, independently: choose a vertex u uniformly at random from $\{v_0, v_1, \dots, v_{s-1}\}$, and a fresh random variable $X = \mathrm{Geo}(p)$; perform a simple random walk of length X starting from u, and join v_s to the last vertex of the walk. Note that the random walk is performed on the *directed* graph.

The motivation behind this definition is as follows. Think of the vertex v_s as a new web page that is being set up. Say the owner wants to put d links in her web page. To build each link, she does the following: she goes to a random page. With probability p she likes the page and puts a link to that page. Otherwise, she clicks on a random link on that page, and follows the link to a new page. Again, with probability p she likes the new page and puts a link to that, otherwise clicks on a random link etc., until she finds a desirable page to link to. The geometric random variables correspond to this selection process.

Our main result regarding the diameter of the random-surfer Webgraph model is the following theorem (recall that the diameter of a directed graph is defined as the diameter of its underlying undirected graph).

▶ **Theorem 2.** *Let d be a positive integer and let $p \in (0, 1]$. A.a.s. as $n \to \infty$ the diameter of the random-surfer Webgraph model with parameters p and d is at most $8e^p(\log n)/p$.*

Notice that the upper bound in Theorem 2 does not depend on d (whereas one would expect that the diameter must decrease asymptotically as d increases). This independence is because in our argument we employ only the first edge created by each new vertex to bound the diameter. Obtaining better bounds as d grows is related to analysing the first order statistic of several intricate random variables, and seems to be much harder.

▶ Remark. In considering the undirected version of diameter of directed graphs, we follow [18]. The directed version of diameter, i. e. the largest directed distance between any two vertices, is much harder to study for the following reason. Let v_s be a new vertex, and let u be the start vertex of a corresponding random walk. The random walk from u could potentially move to the *worst* neighbour of u, i. e. a neighbour whose directed depth is much larger than u, causing a great increase in the directed depth of v_s (compared to the case that it attaches to u). However, in the case of undirected distances, each step of the walk can increase any undirected distance by at most 1.

A *random-surfer tree* is an undirected tree obtained from a random-surfer Webgraph with $d = 1$ by deleting the self-loops of the root and ignoring the edge directions. The *height*

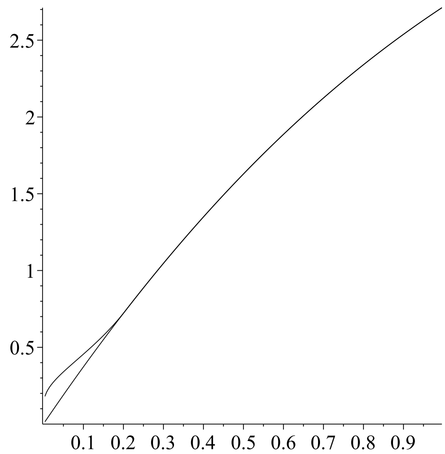

Figure 1 The functions c_L and c_U in Theorems 3 and 4.

of a tree is defined as the maximum graph distance between a vertex and the root. Our main result regarding the height of the random-surfer tree model is the following theorem.

▶ **Theorem 3.** *For $p \in (0,1)$, let s be the unique solution in $(0,1)$ to*

$$s \log\left(\frac{(1-p)(2-s)}{1-s}\right) = 1 \,. \tag{1}$$

Let $p_0 \approx 0.206$ be the unique solution in $(0,1/2)$ to

$$\log\left(\frac{1-p}{p}\right) = \frac{1-p}{1-2p} \,. \tag{2}$$

Define the functions $c_L, c_U : (0,1) \to \mathbb{R}$ as

$$c_L(p) = \exp(1/s)s(2-s)p \,,$$

and

$$c_U(p) = \begin{cases} c_L(p) & \text{if } p_0 \le p < 1 \\ \left(\log\left(\frac{1-p}{p}\right)\right)^{-1} & \text{if } 0 < p < p_0 \,. \end{cases}$$

For every fixed $\varepsilon > 0$, a.a.s. as $n \to \infty$ the height of the random-surfer tree model with parameter p is between $(c_L(p) - \varepsilon)\log n$ and $(c_U(p) + \varepsilon)\log n$.

It is easy to see that the value p_0 and the functions c_L and c_U (plotted in Figure 1) are well defined. Also, c_L and c_U are continuous, and $\lim_{p\to 0} c_L(p) = \lim_{p\to 0} c_U(p) = 0$ and $\lim_{p\to 1} c_L(p) = e$. We suspect that the gap between our bounds when $p < p_0$ is an artefact of our proof technique, and we do not expect a phase transition in the behaviour of the height at $p = p_0$.

We also prove lower and upper bounds for the diameter, which are close to being tight.

▶ **Theorem 4.** *Let c_L and c_U be defined as in Theorem 3. For every fixed $\varepsilon > 0$, a.a.s. as $n \to \infty$ the diameter of the random-surfer tree model with parameter $p \in (0,1)$ is between $(2c_L(p) - \varepsilon)\log n$ and $(2c_U(p) + \varepsilon)\log n$.*

Immediately, we have the following corollary.

▶ **Corollary 5.** *Let c_L and p_0 be defined as in Theorem 3. For any $p \in [p_0, 1)$, the height of the random-surfer tree model with parameter p is a.a.s. asymptotic to $c_L(p) \log n$ as $n \to \infty$, and its diameter is a.a.s. asymptotic to $2c_L(p) \log n$.*

We remark that Theorem 4 does not imply Theorem 2; in fact it does not even imply the diameter of the random-surfer Webgraph model (with $d > 1$) is logarithmic. The reason is that in the random-surfer tree model, due to the tree structure, the random walk for each vertex always moves closer to the root, whereas in the random-surfer Webgraph model, this is not the case, and the random walk could move further from the root.

We now define the PageRank-based selection model introduced in [20, 21].

▶ **Definition 6** (PageRank and the PageRank-based Selection Webgraph Model [20, 21])**.** Let d be a positive integer and let $p, \beta \in [0, 1]$. The *PageRank* of a directed graph is a probability distribution over its vertices, which is the stationary distribution of the following random walk. The random walk starts from a vertex chosen uniformly at random. In each step, with probability p it jumps to a vertex chosen uniformly at random, and with probability $1 - p$ it walks to a random out-neighbour of the current vertex.

We generate a random n-vertex directed multigraph with all vertices having out-degree d. We start with a single vertex with d self-loops. At each subsequent step a new vertex appears, chooses d old vertices and attaches to them (where a vertex can be chosen multiple times). These choices are independent and the head of each edge is a uniformly random vertex of the existing graph with probability β, and is a vertex chosen according to the PageRank distribution on the existing graph with probability $1 - \beta$.

The motivation behind this definition is as follows. Consider the case $\beta = 0$. Think of the vertex v_s as a new web page that is being set up. Say the owner wants to put d links in her web page. She finds the destination pages using d independent Google searches. Since Google sorts the search results according to their PageRank (see [19]), under some suitable randomness assumptions, we may imagine that the probability that a given page is linked to is close to its PageRank.

Chebolu and Melsted [9] showed the PageRank-based selection Webgraph model with $\beta = 0$ is equivalent to the random-surfer Webgraph model. Hence the conclusions of Theorems 2–4 apply to the former model with $\beta = 0$. Moreover, the proof of Theorem 2 easily extends to the PageRank-based selection Webgraph model, giving the same conclusion for all $\beta \in [0, 1]$

In Theorems 3 and 4 we have assumed that $p < 1$, since the situation for $p = 1$ has been clarified in previous work. Let $p = 1$. Then a random-surfer tree has the same distribution as a so-called random recursive tree, the height of which is a.a.s. asymptotic to $e \log n$ as proved by Pittel [22]. It is not hard to alter the argument in [22] to prove that the diameter is a.a.s. asymptotic to $2e \log n$. The diameter of a random-surfer Webgraph thus has also an asymptotically almost sure upper bound of $2e \log n$. For the rest of the paper, we fix $p \in (0, 1)$.

We will need two large deviation inequalities. Define the function $\Upsilon : (0, \infty) \to \mathbb{R}$ as

$$\Upsilon(x) = \begin{cases} x - 1 - \log(x) & \text{if } 0 < x \leq 1 \\ 0 & \text{if } 1 < x \,. \end{cases} \tag{3}$$

The following proposition follows easily from Cramér's Theorem (see, e. g., [11, Theorem 2.2.3, p. 27]) and the calculations in [7, p. 279].

▶ **Proposition 7.** *Let $E_1, E_2, \dots, E_m$ be independent exponential random variables with mean 1. For any fixed $x > 0$, as $m \to \infty$ we have*

$$\exp\left(-\Upsilon(x)m - o(m)\right) \le \mathbb{P}\left[E_1 + E_2 + \cdots + E_m \le xm\right] \le \exp(-\Upsilon(x)m)\,.$$

We include some definitions here. We define the *depth* of a vertex as the length of a shortest path (ignoring edge directions) connecting the vertex to the root, and the *height* of a graph G, denoted by $\mathrm{ht}(G)$, as the maximum depth of its vertices. Clearly the diameter is at most twice the height. In a weighted tree (a tree whose *edges* are weighted), we define the *weight* of a vertex to be the sum of the weights of the edges connecting the vertex to the root, and the *weighted height* of tree T, written $\mathrm{wht}(T)$, to be the maximum weight of its vertices. We view an unweighted tree as a weighted tree with unit edge weights, in which case the weight of a vertex is its depth, and the notion of weighted height is the same as the usual height.

3 Upper bound for the random-surfer Webgraph model

In this section we prove Theorem 2, giving an upper bound for the diameter of the random-surfer Webgraph model. Define the function $f : (-\infty, 1] \to \mathbb{R}$ as

$$f(x) = (2-x)^{2-x} p(1-p)^{1-x}(1-x)^{x-1}\,. \tag{4}$$

We will need a technical lemma which follows from straightforward calculations.

▶ **Lemma 8.** *Let η, c be positive numbers satisfying $\eta \ge 4e^p/p$ and $c \le p\eta$. Then we have $-c\Upsilon(1/c) + c\log f(2-\eta/c) < \eta(1-p)\log(1-p^3) - 1$.*

Proof of Theorem 2. We define an auxiliary weighted tree whose vertex set equals the vertex set of the graph generated by the random-surfer Webgraph model, and whose weighted height dominates the height of this graph. Then we bound the weighted height of this tree. Initially the tree has just one vertex v_0. Recall the growth of the random-surfer Webgraph model at each subsequent step $s \in \{1, 2, \dots, n-1\}$: "a new vertex v_s appears and d edges are created from it to vertices in $\{v_0, v_1, \dots, v_{s-1}\}$, by doing the following probabilistic procedure d times, independently: choose a vertex u uniformly at random from $\{v_0, v_1, \dots, v_{s-1}\}$, and a fresh random variable $X = \mathrm{Geo}(p)$; perform a simple random walk of length X starting from u, and join v_s to the last vertex of the walk." Consider a step s and the first chosen $u \in \{v_0, \dots, v_{s-1}\}$ and $X = \mathrm{Geo}(p)$. In the tree, we join the vertex v_s to u and set the weight of the edge $v_s u$ to be $X + 1$. Note that the edge weights are mutually independent. Clearly, the weight of v_s in the auxiliary tree is greater than or equal to the depth of v_s in the graph. Let $\eta = 4e^p/p$. Hence, to prove the theorem it suffices to show that a.a.s. the weighted height of the auxiliary tree is at most $\eta \log n$. We work with this tree in the rest of the proof.

We consider an alternative way to grow the tree, used in [12], which results in the same distribution. Let $U_1, U_2, \dots$ be i.i.d. uniform random variables in $(0, 1)$. Then for each new vertex v_s, we join it to the vertex $v_{\lfloor sU_s \rfloor}$, which is indeed a vertex uniformly chosen from $\{v_0, \dots, v_{s-1}\}$.

For convenience, we consider the tree when it has $n+1$ vertices $v_0, v_1, \dots, v_n$. Define

$$\mathcal{A}(\ell) = n\mathbb{P}\left[D(n) = \ell\right]\mathbb{P}\left[W(n) > \eta\log n | D(n) = \ell\right]\,,$$

where $D(s)$ and $W(s)$ denote the depth and the weight of vertex v_s, respectively. We have

$$\mathbb{P}\left[\mathrm{wht}(\mathrm{tree}) > \eta \log n\right] \leq \sum_{s=1}^{n} \mathbb{P}\left[W(s) > \eta \log n\right] \leq n\mathbb{P}\left[W(n) > \eta \log n\right] = \sum_{\ell=1}^{n} \mathcal{A}(\ell) ,$$

so to complete the proof it suffices to show $\sum_{\ell=1}^{n} \mathcal{A}(\ell) = o(1)$.

Let $P(0) = 0$ and for $s = 1, \dots, n$, let $P(s)$ denote the index of the parent of v_s. We have

$$\mathbb{P}\left[D(n) \geq \ell\right] = \mathbb{P}\left[D(P(n)) \geq \ell - 1\right] = \cdots = \mathbb{P}\left[D(P^{\ell-1}(n)) \geq 1\right] = \mathbb{P}\left[P^{\ell-1}(n) \geq 1\right] .$$

Since $P(m) = \lfloor mU_m \rfloor \leq mU_m$ for each m and since the U_i are i.i.d., we have

$$\mathbb{P}\left[D(n) \geq \ell\right] = \mathbb{P}\left[P^{\ell-1}(n) \geq 1\right] \leq \mathbb{P}\left[nU_1U_2 \dots U_{\ell-1} \geq 1\right] \leq \exp\left(-(\ell-1)\Upsilon\left(\frac{\log n}{\ell-1}\right)\right) , \tag{5}$$

where we have used Proposition 7 and the fact that $\log(1/U_i)$ is an exponential random variable with mean 1 for each i. The right-hand side is $o(1/n)$ for $\ell = 1.1e \log n$. Hence to complete the proof we need only show that

$$\mathcal{A}(\ell) = o(1/\log n) \qquad \forall \ell \in (0, 1.1e \log n) . \tag{6}$$

Fix an arbitrary positive integer $\ell \in (0, 1.1e \log n)$. The random variable $W(n)$, conditional on $D(n) = \ell$, is a sum of ℓ i.i.d. $1 + \mathrm{Geo}(p)$ random variables. By using Chernoff's technique of bounding the moment generating function of geometric random variables, we get

$$\mathbb{P}\left[W(n) > \eta \log n | D(n) = \ell\right] \leq f(2 - \eta \log n/\ell)^{\ell} , \tag{7}$$

where f is defined in (4). Combining (5) and (7), we get

$$\mathcal{A}(\ell) \leq \exp\left[\log n - (\ell-1)\Upsilon\left(\frac{\log n}{\ell-1}\right) + \ell \log f(2 - \eta \log n/\ell)\right] . \tag{8}$$

Let $c = \ell/\log n$ and $c_1 = c - 1/\log n$. By Lemma 8 and since the function $c\Upsilon(1/c)$ is uniformly continuous in $c \in [0, 2e]$, we find that for large enough n,

$$1 - c_1\Upsilon(1/c_1) + c \log f(2 - \eta/c) < \frac{1}{2}\eta(1-p)\log(1-p^3) = -\Omega(1) .$$

Together with (8), this gives $\mathcal{A}(\ell) = \exp(-\Omega(\log n))$, and (6) follows. ◀

4 Transformations of the Random-surfer Tree Model

In Sections 4–6 we study the random-surfer tree model and we prove Theorems 3 and 4. In this section we show how to transform the random-surfer tree model three times to eventually obtain a new random tree model, which we analyse in subsequent sections. The first transformation is novel. The second one was used by Broutin and Devroye [7], and the third one by Pittel [22].

Let us call the random-surfer tree model the *first model*. First, we will replace the attachment rule with a simpler one by introducing *weights* for the edges. In the first model, the edges are unweighted and in every step s a new vertex v_s appears, chooses an old vertex u, and attaches to a vertex in the path connecting u to the root, according to some rule. We introduce a second model that is weighted, and such that there is a one to one correspondence between the vertices in the second model and in the first model. For a vertex v in the first model, we denote its corresponding vertex in the second model by $\overline{v}$. In the second model, in every step s a new vertex $\overline{v_s}$ appears, chooses an old vertex $\overline{u}$ and attaches to $\overline{u}$, and the weight $w(\overline{u}\,\overline{v_s})$ of the new edge $\overline{u}\,\overline{v_s}$ is chosen such that the *weight* of $\overline{v_s}$ equals the *depth* of v_s in the first model. Let $w\,(\overline{u})$ denote the weight of vertex $\overline{u}$. Then it follows from the definition of the random-surfer tree model that $w(\overline{u}\,\overline{v_s})$ is distributed as $\max\{1 - \mathrm{Geo}(p), 1 - w\,(\overline{u})\}$. The term $\mathrm{Geo}(p)$ here precisely corresponds to the length of the random walk corresponding to v_s in Definition 1, and the term $1 - w\,(\overline{u})$ appears here solely because the weight of $\overline{v_s}$ is at least 1 (in the first model, the depth of v_s is at least 1, since it cannot attach to a vertex higher than the root). Let us emphasize that $w(\overline{u}\,\overline{v_s})$ can be negative. Because the depth of v in the first model equals the weight of $\overline{v}$ in the second model, the height of the first model equals the weighted height of the second model.

We will need to make the degrees of the tree bounded, so we define a third model. In this model, the new vertex can attach just to the leaves. In step s a new vertex v_s appears, chooses a random leaf u and attaches to u using an edge with weight distributed as $\max\{1 - \mathrm{Geo}(p), 1 - w\,(\overline{u})\}$. Simultaneously, a new vertex u' appears and attaches to u using an edge with weight 0. Then we have $w(u) = w(u')$ and henceforth u' plays the role of u, i. e. the next vertex wanting to attach to u, but cannot do so because u is no longer a leaf, may attach to u' instead. Clearly there exists a coupling between the second and third models in which the weighted height of the third model, when it has $2n - 1$ vertices, equals the weighted height of the second model with n vertices. In fact the second model may be obtained from the third one by contracting all zero-weight edges. We can thus study the weighted height of the first model by studying it in the third model.

All the above models were defined using discrete time steps. We now define a fourth model using the following continuous time branching process. At time 0 the root is born. From this moment onwards, whenever a new vertex v is born (say at time κ), it waits for a random time E, which is distributed exponentially with mean 1, and after time E has passed (namely, at absolute time $\kappa + E$) gives birth to two children v_1 and v_2. The weights of the edges vv_1 and vv_2 are generated as follows: vertex v chooses $i \in \{1, 2\}$ independently and uniformly at random. The weight of vv_i is distributed as $\max\{1 - \mathrm{Geo}(p), 1 - w\,(v)\}$ and the weight of vv_{3-i} is 0. All vertices live forever, and each vertex gives birth to exactly two children during its lifetime. Given $t \geq 0$, we denote by T_t the (almost surely finite) random tree obtained by taking a snapshot of this process at time t. By the memorylessness of the exponential distribution, if one starts looking at this process at any deterministic moment, the next leaf to give birth is chosen uniformly at random. Hence for any stopping time τ, the distribution of T_τ, conditional on T_τ having $2n - 1$ vertices, is the same as the distribution of the third model when it has $2n - 1$ vertices.

The following lemma implies that certain results for T_t carry over to results for the random-surfer tree model. The proof is by a coupling argument and uses [7, Proposition 2].

▶ **Lemma 9.** *Assume that there exist constants θ_L, θ_U such that for every fixed $\varepsilon > 0$,*

$$\mathbb{P}\left[\theta_L(1-\varepsilon)t \leq \mathrm{wht}(T_t) \leq \theta_U(1+\varepsilon)t\right] \to 1$$

as $t \to \infty$. Then for every fixed $\varepsilon > 0$, a.a.s. as $n \to \infty$ the height of the random-surfer tree model is between $\theta_L(1-\varepsilon)\log n$ and $\theta_U(1+\varepsilon)\log n$.

We define T_t in a static way, which is equivalent to the dynamic definition above.

▶ **Definition 10** (T_∞, T_t)**.** Let T_∞ denote an infinite binary tree. To every edge e is associated a random pair (E_e, W_e) and to every vertex v a random variable W_v, where the W_e's and

W_v's are the *weights*. The law for $\{E_e\}_{e\in E(T)}$ is easy: first with every vertex v we associate independently an exponential random variable with mean 1, and we let the values of E on the edges joining v to its two children be equal to this variable. In the dynamic interpretation, this random variable denotes the age of v when it gives birth. Generation of the weights is done in a top-down manner, where we think of the root as the top vertex. Let the weight of the root be zero. Let v be a vertex whose weight has been determined, and let v_1, v_2 be its two children. Choose $i \in \{1,2\}$ independently and uniformly at random, and then choose $Y = 1 - \mathrm{Geo}(p)$ independently of previous choices. Then let

$$W_{vv_i} = \max\{Y, 1 - W_v\}, \quad W_{v_i} = W_v + W_{vv_i}\,, \tag{9}$$

and $W_{vv_j} = 0, W_{v_j} = W_v$ for $j = 3 - i$.

For a vertex v, let $\pi(v)$ be the set of edges of the path connecting v to the root. It is easy to check that the weight of any vertex v equals $\sum_{e\in\pi(v)} W_e$. Finally, given $t \geq 0$ we define T_t as the subtree of T_∞ induced by vertices with birth time at most t. Note that T_t is connected by definition, and is finite almost surely.

5 Lower Bounds for the Random-surfer Tree Model

Here we prove the lower bounds in Theorems 3 and 4. For this, we consider another infinite binary tree T'_∞ which is very similar to T_∞, except for the generation rules for the weights, which are as follows. Let the weight of the root be zero. Let v be a vertex whose weight has been determined, and let v_1, v_2 be its two children. Choose $i \in \{1,2\}$ independently and uniformly at random, and choose $Y = 1 - \mathrm{Geo}(p)$ independently of previous choices. Then let

$$W_{vv_i} = Y \text{ and } W_{v_i} = W_v + W_{vv_i} \tag{10}$$

and $W_{vv_j} = 0$ and $W_{v_j} = W_v$ for $j = 3 - i$. Comparing (10) with (9), we find that the weight of every vertex in T'_∞ is stochastically less than or equal to that of its corresponding vertex in T_∞. The tree T'_t is defined as before. Clearly probabilistic lower bounds for $\mathrm{wht}(T'_t)$ are also probabilistic lower bounds for $\mathrm{wht}(T_t)$.

The proof of the following lemma is similar to that of [7, Lemma 4] except a small twist is needed at the end to handle the negative weights. Distinct vertices u and v in a tree are called *antipodal* if the unique (u, v)-path in the tree passes through the root.

▶ **Lemma 11.** *Consider the tree T'_∞. Let $\gamma_L : (0,1) \to \mathbb{R}$ be such that for every $a \in (0,1)$, each vertex u and each descendent v of u that is m levels deeper,*

$$\mathbb{P}[W_v - W_u \geq am] \geq \exp(-m\gamma_L(a) + o(m)) \tag{11}$$

as $m \to \infty$. Assume that there exist $\alpha^*, \rho^* \in (0,1)$ with $\gamma_L(\alpha^*) + \Upsilon(\rho^*) = \log 2$, where Υ is defined in (3). Then for every fixed $\varepsilon > 0$, a.a.s. there exist antipodal vertices u, v of T'_t with weights at least $\frac{\alpha^*}{\rho^*}(1-\varepsilon)t$.

Proof. Let $c = \frac{\alpha^*}{\rho^*}$, and let $\varepsilon, \delta > 0$ be arbitrary. We prove that with probability at least $1 - \delta$ for all large enough t there exists a pair (u, v) of antipodal vertices of T'_∞ with $\max\{B_u, B_v\} < t$ and $\min\{W_u, W_v\} > (1 - 2\varepsilon)\,ct$.

Let L be a constant positive integer that will be determined later, and let $\alpha = \alpha^*$ and $\rho = \frac{\alpha}{c(1-\epsilon)} > \rho^*$. Since $\gamma_L(\alpha^*) + \Upsilon(\rho^*) = \log 2$ and $\rho^* < 1$ and Υ is strictly decreasing on $(0,1]$, we have

$$\gamma_L(\alpha) + \Upsilon(\rho) < \log 2\,.$$

Build a Galton-Watson process from T'_∞ whose particles are a subset of vertices of T'_∞, as follows. Start with the root as the initial particle of the process. If a given vertex u is a particle of the process, then its potential offspring are its 2^L descendants that are L levels deeper. Moreover, such a descendent v is an offspring of u if and only if $W_v - W_u \geq \alpha L$ and $B_v - B_u \leq \rho L$. As these two events are independent, the expected number of children of u is at least

$$2^L \mathbb{P}[W_v - W_u \geq \alpha L]\, \mathbb{P}[B_v - B_u \leq \rho L] \geq \exp[(\log 2 - \gamma_L(\alpha) - \Upsilon(\rho) - o(1))L]$$

as $L \to \infty$, by (11) and Proposition 7. Since we have $\log 2 - \gamma_L(\alpha) - \Upsilon(\rho) > 0$, we may choose L large enough that this expected value is strictly greater than 1. Therefore, this Galton-Watson process survives with probability $q > 0$.

We now boost this probability up to $1 - \delta$, by starting several independent processes, giving more chance that at least one of them survives. Specifically, let b be a constant large enough that

$$(1-q)^{2^{b-1}} < \delta/3\,.$$

Consider 2^b Galton-Watson processes, which have the vertices at depth b of T'_∞ as their initial particles, and reproduce using the same rule as before. Let a be a constant large enough that

$$2^{b+1}(e^{-a} + (1-p)^{a+2}) < \delta/3\,,$$

and let A be the event that all edges e in the top b levels of T'_∞ have $E_e \leq a$ and $W_e \geq -a$. Then

$$1 - \mathbb{P}[A] \leq 2^{b+1}(e^{-a} + (1-p)^{a+2}) < \delta/3\,.$$

Also, let Q be the event that in each of the two branches of the root, at least one of the 2^{b-1} Galton-Watson processes survives. Then

$$1 - \mathbb{P}[Q] \leq 2(1-q)^{2^{b-1}} < 2\delta/3\,,$$

and so with probability at least $1-\delta$ both A and Q occur.

Assume that both A and Q occur. Let

$$m = \left\lfloor \frac{t(1-\varepsilon)}{\rho L} \right\rfloor$$

and let u and v be particles at generation m of surviving processes in distinct branches of the root. Then u and v are antipodal,

$$\max\{B_u, B_v\} \leq ab + m\rho L \leq t(1-\varepsilon) + O(1) < t\,,$$

and

$$\min\{W_u, W_v\} \geq -ab + m\alpha L \geq \frac{(1-\varepsilon)\alpha}{\rho}\, t - O(1) > c(1-2\varepsilon)t$$

for t large enough, as required. ◀

Let $Y_1, Y_2, \ldots$ be i.i.d. with $Y_i = 1 - \mathrm{Geo}(p)$. Let $f : (-\infty, 1] \to \mathbb{R}$ be as defined in (4). Note that $f(1) = p$ and $f\left(2 - p^{-1}\right) = 1$. It is easy to verify that f is continuous in $(-\infty, 1]$ and differentiable in $(-\infty, 1)$. Moreover, f is increasing on $(-\infty, 2-p^{-1}]$ and decreasing on $[2 - p^{-1}, 1]$. Using standard concentration methods, we can show that as $m \to \infty$, uniformly for all $a \in [0, 1]$,

$$\mathbb{P}[Y_1 + \cdots + Y_m \geq am] \geq [f(a) - o(1)]^m\,, \tag{12}$$

and if $p \geq 1/2$, then uniformly for all $a \in [0, 2 - \frac{1}{p}]$ we have

$$\mathbb{P}\left[Y_1 + \cdots + Y_m \geq am\right] \geq [1 - o(1)]^m \ . \tag{13}$$

We define a two variable function $\Phi(a, s) = p(1-p)(2-s)^2(s-a) - a(1-s)$, and we define a function $\phi : [0,1] \to [0,1]$ implicitly by the equation

$$\Phi(a, \phi(a)) = p(1-p)(2-\phi(a))^2(\phi(a) - a) - a(1 - \phi(a)) = 0\,. \tag{14}$$

It is not hard to see that ϕ is well defined, increasing and invertible on $[0,1]$, and differentiable on $(0,1)$. Moreover, if $a \in \{0,1\}$ then $\phi(a) = a$, and otherwise, $0 < a < \phi(a) < 1$.

Next let $\hat{Y}_1, \hat{Y}_2, \dots$ be independent and distributed as follows: for every $i = 1, 2, \dots$ we flip an unbiased coin, if it comes up heads, then $\hat{Y}_i = Y_i$, otherwise $\hat{Y}_i = 0$. Define $g_L : (0,1) \to \mathbb{R}$ as

$$g_L(a) = \begin{cases} 1/2 & \text{if } p > 1/2 \text{ and } 0 < a < 1 - \frac{1}{2p} \\ \frac{\phi(a)-a}{\phi(a)} \left(\frac{(1-p)(2-\phi(a))}{1-\phi(a)}\right)^a & \text{otherwise.} \end{cases}$$

The inequalities (12) and (13) together with Stirling's formula imply the following.

▶ **Lemma 12.** *We have the following large deviation inequality for every fixed $a \in (0,1)$ as $m \to \infty$.*

$$\mathbb{P}\left[\hat{Y}_1 + \cdots + \hat{Y}_m \geq am\right] \geq (2g_L(a) - o(1))^{-m}\,.$$

The lower bound in Theorem 3 is obtained easily from the following lemma and Lemma 9.

▶ **Lemma 13.** *Given $\varepsilon > 0$, a.a.s as $t \to \infty$ there exist two antipodal vertices u, v of T'_t with weights at least $c_L(p)(1-\varepsilon)t$. In particular, a.a.s. the weighted height of T'_t is at least $c_L(p)(1-\varepsilon)t$.*

Proof. It is easy to see that there is a unique solution $s \in (0,1)$ to $(1-p)(2-s) = \exp(1/s)(1-s)$, and if $p > 1/2$ then $s > 2 - 1/p$. By definition, $c_L = c_L(p) = \exp(1/s)s(2-s)p$. By Lemma 12, the assumption (11) of Lemma 11 holds for $\gamma_L(a) = \log(2g_L(a))$. Let $a = \phi^{-1}(s)$ and let $\rho = 1 - \frac{a}{s}$. Since $s \in (0,1)$ we have $0 < a < s < 1$ and thus $\rho \in (0,1)$ as well. Moreover, since $\Phi(a,s) = 0$, we have $c_L = a/\rho$. It is easy to verify that $\Phi(1 - \frac{1}{2p}, 2 - \frac{1}{p}) = 0$. If $p > 1/2$, then we have $s > 2 - \frac{1}{p}$. Since ϕ^{-1} is increasing, we have $a = \phi^{-1}(s) \geq 1 - \frac{1}{2p}$. This implies that $g_L(a) = \frac{s-a}{s}\exp(a/s)$, hence $\log(2g_L(a)) + \rho - 1 - \log(\rho) = \log 2$, and Lemma 11 completes the proof. ◀

Proof of the lower bound in Theorem 4. Fix $\varepsilon > 0$. Let us define the *semi-diameter* of a tree as the maximum weighted distance between any two antipodal vertices. Clearly, semi-diameter is a lower bound for the diameter, so we just need to show a.a.s. as $n \to \infty$ the semi-diameter of the random-surfer tree with n vertices is at least $(2c_L(p) - \varepsilon)\log n$. By Lemma 13, a.a.s as $t \to \infty$ the semi-diameter of T'_t is at least $(2c_L(p) - \varepsilon)t$. Using an argument similar to the proof of Lemma 9 we may conclude that a.a.s. as $n \to \infty$ the semi-diameter of the third model (of Section 4) with $2n - 1$ vertices is at least $(2c_L(p) - \varepsilon)\log n$. Then it is easy to observe that the same is true for the random-surfer tree with n vertices, and the proof is complete. ◀

6 Upper Bounds for the Random-surfer Tree Model

In this section we prove the upper bounds in Theorems 3 and 4. We start with a lemma which complements Lemma 11.

▶ **Lemma 14.** *Let $\gamma_U : [0,1] \to [0,\infty)$ be a continuous function such that for every fixed $a \in [0,1]$ and every vertex v of T_∞ at depth m,*

$$\mathbb{P}\left[\sum_{e\in\pi(v)} W_e > am\right] \leq \exp(-m\gamma_U(a) + o(m)) \tag{15}$$

as $m \to \infty$. Define

$$\theta = \sup\left\{\frac{a}{\rho} : \gamma_U(a) + \Upsilon(\rho) = \log 2 : a \in [0,1], \rho \in (0,\infty)\right\} .$$

Then for every fixed $\varepsilon > 0$ we have $\mathbb{P}[\mathrm{wht}(T_t) > \theta(1+\varepsilon)t] \to 0$ as $t \to \infty$.

The proof of Lemma 14 is similar to that of [7, Lemma 3], in which the assumption (15) is not needed. In fact, in the model studied in [7], the weights $\{W_e : e \in \pi(v)\}$ are mutually independent, and the authors use Cramér's Theorem to obtain a large deviation inequality for $\sum_{e\in\pi(v)} W_e$.

Let $Y_1, Y_2, \ldots$ be i.i.d. with $Y_i = 1 - \mathrm{Geo}(p)$, and define random variables $X_1, X_2, \ldots$ as follows: $X_1 = \max\{Y_1, 1\}$, and for $i \geq 1$, $X_{i+1} = \max\{Y_{i+1}, 1 - (X_1 + \cdots + X_i)\}$. Define $h : [0,1] \to \mathbb{R}$ as

$$h(x) = \begin{cases} 1 & \text{if } p \geq \frac{1}{2} \text{ and } 0 \leq x \leq 2 - \frac{1}{p} \\ \left(\frac{p}{1-p}\right)^x & \text{if } p < \frac{1}{2} \text{ and } 0 \leq x \leq \frac{1-2p}{1-p} \\ f(x) & \text{otherwise} . \end{cases}$$

A careful but straightforward analysis using exact formulae gives that there exists an absolute constant C such that for every $a \in [0,1]$ and every positive integer m we have

$$\mathbb{P}[X_1 + \cdots + X_m > am] \leq Cm^2 h(a)^m . \tag{16}$$

Next define random variables $\hat{X}_1, \hat{X}_2, \ldots$ as follows: for every $i = 1, 2, \ldots$ we flip an independent unbiased coin, if it comes up heads, then $\hat{X}_i = X_i$, otherwise $\hat{X}_i = 0$. Define $g_U : [0,1] \to \mathbb{R}$ as

$$g_U(a) = \begin{cases} 1/2 & \text{if } p \geq 1/2 \text{ and } 0 \leq a \leq 1 - 1/2p \\ (\frac{1-p}{p})^a/2 & \text{if } p < 1/2 \text{ and } 0 \leq a \leq \frac{1-2p}{2-2p} \\ 1/p & \text{if } a = 1 \\ \frac{\phi(a)-a}{\phi(a)}\left(\frac{(1-p)(2-\phi(a))}{1-\phi(a)}\right)^a & \text{otherwise} , \end{cases}$$

where ϕ is defined by (14). Recall that we have $0 < a < \phi(a) < 1$ for $a \in (0,1)$, so g_U is well defined. The following lemma can be proved using (16), Stirling's formula and standard arguments.

▶ **Lemma 15.** *We have the following large deviation inequality for every $a \in [0,1]$ and every positive integer m, where C' is an absolute constant: $\mathbb{P}\left[\hat{X}_1 + \cdots + \hat{X}_m > am\right] \leq C'm^3(2g_U(a))^{-m}$.*

We are ready to prove the upper bound in Theorem 3. The upper bound in Theorem 4 follows immediately as in every tree the diameter is at most twice the height.

Proof of the upper bound in Theorem 3. Let $c_U = c_U(p)$. By Lemma 9 we just need to show that given $\varepsilon > 0$, a.a.s as $t \to \infty$ the weighted height of T_t is at most $(1+\varepsilon)c_U t$. For proving this we use Lemma 14. Lemma 15 implies that condition (15) of Lemma 14 holds with $\gamma_U(a) = \log(2g_U(a))$, so we need only show that

$$c_U = \sup\left\{\frac{a}{\rho} : \log(g_U(a)) + \rho - 1 - \log(\rho) = 0 : a \in [0,1], \rho \in (0,\infty)\right\} \tag{17}$$

Observe that if $0 < g_U(a) \le 1$, there is a unique $\rho \in (0,1]$ satisfying $\log(g_U(a)) + \rho - 1 - \log(\rho) = 0$; and if $g_U(a) > 1$, there is no $\rho \in (0,\infty)$ satisfying this equation. Let $a_{\max} \in [0,1]$ denote the unique solution to $g_U(x) = 1$. Define the function $\tau : [0, a_{\max}] \to (0,1]$ as follows. Let $\tau(a_{\max}) = 1$ and for $x < a_{\max}$ let $\tau(x)$ be the unique number satisfying

$$\log(g_U(x)) + \tau(x) - 1 - \log \tau(x) = 0\,. \tag{18}$$

Hence to prove (17) it is enough to show that

$$c_U = \sup\left\{\frac{x}{\tau(x)} : x \in [0, a_{\max}]\right\}\,. \tag{19}$$

Differentiating (18) and using the implicit function theorem, we find that τ is differentiable and

$$\tau'(x) = \frac{\tau(x)}{1-\tau(x)} \times \frac{g'_U(x)}{g_U(x)}\,.$$

It can be verified that $\log(g_U(a))$ is increasing and convex and that $\rho - 1 - \log(\rho)$ is decreasing and convex. Using standard convexity arguments it can be proved that the supremum in (19) occurs at a point $x^* \in (0, a_{\max})$ with $\tau(x^*) = x^* \tau'(x^*)$, hence it is enough to find $x^* \in (0, a_{\max})$ satisfying

$$c_U = \frac{x^*}{\tau(x^*)} = \frac{1-\tau(x^*)}{\tau(x^*)} \frac{g_U(x^*)}{g'_U(x^*)}\,. \tag{20}$$

Recall that $p_0 \approx 0.206$ is the unique solution to (2). If $0 < p \le p_0$, we let $x^* = \left[2\log\left(\frac{1-p}{p}\right)\right]^{-1}$. Then it is not hard to see that

$$g_U(x^*) = \exp\left(\frac{1}{2} - \log 2\right), \quad g'_U(x^*) = \log\left(\frac{1-p}{p}\right) g_U(x^*), \text{ and } \quad \tau(x^*) = 1/2\,,$$

and (20) follows. If $p_0 < p < 1$, we let $x^* = \phi^{-1}(s^*)$, where $s^* \in (0,1)$ is the unique solution to (1). Then (20) follows from the following equations, which are not hard to verify:

$$g_U(x^*) = \left(1 - \frac{x^*}{s^*}\right)\exp\left(x^*/s^*\right), \quad g'_U(x^*) = g_U(x^*)/s^*, \text{ and } \quad \tau(x^*) = 1 - \frac{x^*}{s^*}\,. \qquad \blacktriangleleft$$

7 Concluding Remarks

One natural open problem is to close the gap between the lower and upper bounds in Theorems 3 and 4 when $p < p_0$. It seems that for solving this problem new ideas are required. In Theorem 2 we gave logarithmic upper bounds for the diameter of the random-surfer

Webgraph model for all $d \geq 1$, which are close to being tight when $d = 1$. Another interesting open problem is to give lower bounds for $d > 1$. This problem seems to need quite different techniques. In fact, the diameter for $d > 1$ might be of a smaller order, e. g. $\Theta(\log n / \log\log n)$, as is the case for the preferential attachment model (see [5, Theorem 1]).

There is a common generalization of random recursive trees, preferential attachment trees, and random-surfer trees. Consider i.i.d. random variables $X_1, X_2, \ldots \in \{0, 1, 2, \ldots\}$. Start with a single vertex v_0. At each step s a new vertex v_s appears, chooses a random vertex u in the present graph, and then walks X_s steps from u towards v_0, attaching to the last vertex in the walk (if it reaches v_0 before X_s steps, it attaches to v_0). Random recursive trees correspond to $X_i = 0$, preferential attachment trees correspond to $X_i = \text{Bernoulli}(1/2)$ (see, e. g., [4, Theorem 3.1]), and random-surfer trees correspond to $X_i = \text{Geo}(p)$. Using the ideas of this paper, it is possible to obtain lower and upper bounds for the height and the diameter of this general model (similar to Theorems 3 and 4), provided one can prove large deviation inequalities (similar to (12) and (13)) for the sum of X_i's and also large deviation inequalities (similar to (16)) for the sum of random variables X'_i defined as $X'_1 = 1$ and $X'_{i+1} = \max\{1 - X_i, 1 - (X'_1 + \cdots + X'_i)\}$ for $i > 0$.

References

1 A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

2 S. Bhamidi. Universal techniques to analyze preferential attachment trees: global and local analysis. preprint, available via `http://www.unc.edu/~bhamidi/`, 2007.

3 G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. *Europhys. Lett.*, 54(4):436–442, 2001.

4 A. Blum, T.-H. H. Chan, and M. R. Rwebangira. A random-surfer web-graph model. In *Proc. of 8th Workshop on Algorithm Engineering and Experiments and 3rd Workshop on Analytic Algorithmics and Combinatorics*, pages 238–246, 2006.

5 B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, January 2004.

6 A. Bonato. *A course on the web graph*, volume 89 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

7 N. Broutin and L. Devroye. Large deviations for the weighted height of an extended class of trees. *Algorithmica*, 46(3-4):271–297, 2006.

8 D. Chakrabarti and C. Faloutsos. *Graph Mining: Laws, Tools, and Case Studies*. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers, 2012.

9 P. Chebolu and P. Melsted. Pagerank and the random surfer model. In *Proceedings of the 19th annual ACM-SIAM symposium on Discrete algorithms*, SODA'08, pages 1010–1018, Philadelphia, PA, USA, 2008.

10 F. Chung and L. Lu. *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2006.

11 A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

12 L. Devroye, O. Fawzi, and N. Fraiman. Depth properties of scaled attachment random recursive trees. *Random Structures Algorithms*, 41(1):66–98, 2012.

13 S. Dommers, R. van der Hofstad, and G. Hooghiemstra. Diameters in preferential attachment models. *Journal of Statistical Physics*, 139(1):72–107, 2010.

14 E. Drinea, A. Frieze, and M. Mitzenmacher. Balls and bins models with feedback. In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA'02, pages 308–315, Philadelphia, PA, USA, 2002.

15 R. Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

16 G. Ergün and G.J. Rodgers. Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications*, 303(1–2):261–272, 2002.

17 P. L. Krapivsky and S. Redner. Organization of growing random networks. *Phys. Rev. E*, 63:066123, May 2001.

18 J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1), March 2007.

19 L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999.

20 G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. In *Proceedings of the 8th Annual International Conference on Computing and Combinatorics*, COCOON'02, pages 330–339, London, UK, UK, 2002.

21 G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. *Internet Mathematics*, 3(1):1–20, 2006.

22 B. Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures and Algorithms*, 5(2):337–347, 1994.