Rumor Source Detection for Rumor Spreading on Random Increasing Trees

Michael Fuchs* and Pei-Duo Yu
Department of Applied Mathematics
National Chiao Tung University
Hsinchu, 300
Taiwan

14 August, 2014

Abstract

In a recent paper, Shah and Zaman proposed the rumor center as an effective rumor source estimator for rumor spreading on random graphs. They proved for a very general random tree model that the detection probability remains positive as the number of nodes to which the rumor has spread tends to infinity. Moreover, they derived explicit asymptotic formulas for the detection probability of random d-regular trees and random geometric trees. In this paper, we derive asymptotic formulas for the detection probability of grown simple families of random increasing trees. These families of random trees contain important random tree models as special cases, e.g., binary search trees, recursive trees and plane-oriented recursive trees. Our results show that the detection probability varies from 0 to 1 across these families. Moreover, a brief discussion of the rumor center for unordered trees is given as well.

1 Introduction and Results

Rumor spreading on random trees has a long history in the biology, computer science and probability literature and has been investigated from many different angles. In a recent paper, Shah and Zaman [9, 10] added a new angle by putting forth the *rumor source detection problem* which asks for the correct identification of the rumor source when only information about the underlying model and the infected nodes is known. In [9, 10], this problem was discussed for random d-regular trees and random geometric trees. Then, in [11], the authors generalized their approach to obtain results for very general families of random trees. Their studies, even though all of them very recent, have already lead to many follow-up works (e.g., according to a google scholar search from August 14, 2014, the number of citations of the paper [10] had already reached 80).

From now on, we assume that some random tree model is fixed. After some time has elapsed, the rumor has spread to n nodes which form a tree Γ. The main idea in [9, 10] was to assign a score to the nodes of Γ. The so-called *rumor center* is then the node which receives the highest score (where ties are either

ignored or broken uniformly at random). In [9, 10], the authors showed that the rumor source estimator obtained in this way is the maximum likelihood (ML) estimator if the underlying random tree model are random d-regular trees. However, for most other random tree models, the rumor source estimator is not the ML estimator. Nevertheless, it was shown in [11] that for very general families of random trees, the rumor source estimator is still effective in the sense that the detection probability tends to a positive value as the number of infected nodes n tend to infinity.

Precise asymptotic values for detection probabilities have so far only been found in the special cases of d-regular trees and geometric trees. It is the purpose of this work to derive detection probabilities for other classes of random trees, namely, all subclasses of simple families of random increasing trees whose random model arises from a (natural) tree evolution process. These subclasses will contain d-regular trees and, e.g., the following important random tree models:

- *Recursive Trees*: they have been proposed as a simple model for the spread of epidemics (a situation very similar to rumor spreading); see Moon [7]. We will show that they constitute the limiting case of d-regular trees as d tends to infinity.

- *Plane-oriented Recursive Trees*: they are one of the most simplest model for real complex networks; see the important paper of Barabási and Albert [1].

We will give a precise mathematical definition of simple families of random increasing trees below and describe some of their properties; for more information see Bergeron, Flajolet, and Salvy [2].

We now provide some more details in order to be able to state our results. We fix some notations. Recall that Γ denotes the tree of the nodes to which the rumor has spread. We will denote by $V(\Gamma)$ the nodes of Γ and by $E(\Gamma)$ the edges of Γ. If $v \in V(\Gamma)$, Γ^v will denote Γ rooted at v with an (arbitrary) embedding in the plane, where we will draw Γ in such a way that v is at the top (and the subtrees are below). If $u \in V(\Gamma)$, then Γ_u^v will denote the subtree at the fringe of Γ^v rooted at u.

Rumor Center. In this paragraph, we will recall the definition of the rumor center from [9, 10]. For $v \in V(\Gamma)$, we define a score as follows

$$R(v, \Gamma) = n! \prod_{u \in V(\Gamma)} \frac{1}{|\Gamma_u^v|}.$$

This is the so-called *shape functional*; see for instance Fill [4]. In order to explain its meaning, we need to recall some further notation from graph theory. We call a rooted tree *ordered* if it comes with a fixed embedding into the plane (where in this paper, we always draw the root at the top); otherwise, the tree is called *unordered*. Moreover, a rooted *increasing tree* of n nodes is a tree whose nodes are labeled with labels from the set $\{1, \ldots, n\}$ in such a way that every sequence of labels from to the root to a leaf forms an increasing sequence. Now, we can explain the meaning of the shape functional: it gives the number of rooted ordered increasing trees which are isomorphic to Γ^v.

We next recall the definition of rumor center from [9, 10].

Definition 1. Let Γ be a tree. A node $v \in V(\Gamma)$ which maximizes $R(v, \Gamma)$ is called a *rumor center* of Γ.

Thus, a rumor center v of Γ is a node such that the number of rooted ordered increasing trees which are isomorphic to Γ^v is maximal. Every such increasing tree corresponds to a spreading order in which the rumor has spread from the source v. Consequently, if all spreading orders are equally likely (as is the case, e.g., for d-regular trees; see [9, 10] and below), then the rumor center is the most likely rumor source or in other words the rumor center is the ML estimator for the rumor source.

It was shown in [9, 10] that the rumor center has a surprisingly easy characterization. We will give two versions of this characterization. For the first, we need the following definition.

Definition 2. Let Γ be a tree. A node $v \in V(\Gamma)$ is called a *local rumor center* if $R(v, \Gamma) \geq R(u, \Gamma)$ for all $u \in V(\Gamma)$ with $\{u, v\} \in E(\Gamma)$.

Then, Shah and Zaman proved the following result in [9, 10].

Theorem 1 (Shah and Zaman; 2010 - Version 1). *Let Γ be a tree. Then, every local rumor center is a rumor center.*

The second version of Sha and Zaman's result (which is in fact only a more precise version of the first one) characterizes the rumor center by graph-theoretical properties.

Theorem 2 (Shah and Zaman; 2010 - Version 2). *Let Γ be a tree with n nodes. Then, $v \in V(\Gamma)$ is a rumor center of Γ if and only if $|\Gamma_u^v| \leq n/2$ for all $u \in V(\Gamma)$ with $\{u, v\} \in E(\Gamma)$. Moreover, if all inequalities are strict there is only one rumor center; otherwise, there are exactly two adjacent rumor centers.*

The *rumor source estimator* is now defined as follows: if there is only one rumor center, then we choose this node; if there are two, we either ignore them or choose one of them uniformly at random.

The appropriateness of the rumor source estimator as defined above depends on the random model. In the definitions above, we considered ordered trees. This, however, might not be always appropriate, for instance if the underlying tree model has not a fixed but dynamic structure (e.g., if a node can spread the rumor to an arbitrary large number of neighbors; see the definition of recursive trees below). Then, considering unordered trees might be advantageous. For such trees, the above definition of $R(v, \Gamma)$ has to be suitable modified. Unfortunately, the resulting characterization of nodes v which maximize the score becomes messier; see the appendix of this paper for details.

(Grown) Simple Families of Increasing Trees. In this paragraph, we are going to explain the random tree models which will be used in this paper. First, consider the set of all rooted ordered increasing trees. A *simple family of increasing trees* consists of this set together with a sequence of weights $(\phi_i)_{i \geq 0}$ with $\phi_0 > 0$ and $\phi_i > 0$ for some $i \geq 2$. For every tree T, we define its weight as

$$w(T) = \prod_{v \in V(t)} \phi_{d(v)},$$

where $d(v)$ is the out-degree of v (= the number of edges of v which point away from the root). Moreover, set

$$\tau_n := \sum_{V(T)=n} w(T).$$

Then, a probability space on trees of size n is defined as follows: a tree T of size n has probability $w(T)/\tau_n$. The resulting family of random trees is called a *simple family of random increasing trees*.

We give some prominent examples.

- d-ary trees: $\phi_i = \binom{d}{i}, 0 \leq i \leq d$ and $\phi_i = 0$ for all $i > d$ (here, $d \in \{2, 3, \ldots\}$).

- Recursive trees: $\phi_i = 1/i!$ for all $i \geq 0$.

- Generalized plane-oriented recursive trees: $\phi_i = \binom{r+i-2}{i}$ for all $i \geq 0$ (here, $r > 1$ is a real number).

These three families contain, e.g., random binary trees (d-ary trees with $d = 2$) which are equivalent to random binary search trees from computer science and plane-oriented recursive trees (PORTs for short;

these are generalized PORTs with $r = 2$); see the introduction and [2] for more explanation concerning the relevance of these two random tree models.

The above three families of random increasing trees are very special; see Panholzer and Prodinger [8]. More precisely, it was shown in [8] that out of all families of random increasing trees they are the only ones for which the random model alternatively can also be obtained from a (natural) tree evolution process. Consequently, they have been nicknamed *grown simple families of random increasing trees*; see, e.g., Kuba and Panholzer [6].

We briefly describe the tree evolution process for the above three families.

- d-ary trees: the first node is the root and d empty leaves are attached; for the second node, one leaf is chosen uniformly at random and the node together with d empty leaves is placed there; for the third node, again one of the leafs is chosen uniformly at random, etc.

- Recursive trees: assume that a tree with $n - 1$ nodes was already constructed; for the next node, choose one of the nodes uniformly at random and add the next node as child. (Note that the tree here is unordered).

- Generalized plane-oriented recursive trees: again assume that a tree with $n - 1$ nodes was already constructed; for the next node, choose an existing node v with probability proportional to $d(v) + r - 1$ ($d(v)$ is the out-degree of v) and add the next node as child. (The tree is again unordered; however, for $r = 2$, this random model is equivalent to the uniform model on rooted ordered increasing trees).

From these descriptions, it is obvious that the random model of d-ary trees is the uniform model on rooted ordered increasing d-ary trees and the random model for PORTs (generalized PORTs with $r = 2$) is the uniform model on rooted ordered increasing trees (as already mentioned above). Moreover, the random model for recursive trees is the uniform model on rooted unordered increasing trees. Thus, the rumor source estimator described in the previous paragraph is a ML estimator only for the former two families of random increasing trees but not for the latter (and also not for generalized PORTS with $r \neq 2$).

For later purpose, we need some more properties of the above three families of random increasing trees. Therefore, set

$$\phi(z) = \sum_{i \geq 0} \phi_i z^i, \qquad \tau(z) = \sum_{n \geq 1} \tau_n \frac{z^n}{n!}.$$

Then, it is straightforward to show that

$$\tau'(z) = \phi(\tau(z)).$$

Solving this differential equation for the above families gives the following:

- d-ary trees: $\tau(z) = -1 + (1 - (d - 1)z)^{-1/(d-1)}$.

- Recursive trees: $\tau(z) = \log(1/(1 - z))$.

- Generalized plane-oriented recursive trees: $\tau(z) = 1 - (1 - rz)^{1/r}$.

From this τ_n is easy to derive by standard Taylor series expansion.

Results. In this paragraph, we explain our results. Consider a random increasing tree with n nodes (as random model, we choose one of the three random models from the previous paragraph). We denote by C_n the probability that the node obtained from the rumor source estimator is indeed the rumor source, where we use here the strategy that ties are ignored (since ties anyway occur only with asymptotic probability zero; see below). Then, we have the following result for grown simple families of random increasing trees.

4

Theorem 3.   *(a) (d-ary Trees) We have,*

$$\lim_{n\to\infty} P(C_n) = k_{d\text{-ary}} = 1 - d + d2^{-1/(d-1)}$$

is increasing in d and

$$\lim_{d\to\infty} k_{d\text{-ary}} = 1 - \ln 2.$$

Thus, for $d \geq 3$,

$$0.12132\cdots = \frac{3}{2}\sqrt{2} - 2 \leq k_{d\text{-ary}} < 1 - \ln 2.$$

  (b) (Recursive Trees) We have, $\lim_{n\to\infty} P(C_n) = 1 - \ln 2$.

  (c) (Generalized PORTs) We have,

$$\lim_{n\to\infty} P(C_n) = k_r = r - (r-1)2^{1/r}$$

is decreasing in r with

$$\lim_{r\to 1} k_r = 1 \qquad and \qquad \lim_{r\to\infty} k_r = 1 - \ln 2.$$

Thus, for $r > 1$,

$$1 - \ln 2 < k_r < 1.$$

Remark 1. Due to part (b), recursive trees can be seen as the limiting case of d-ary trees as d tends to infinity. Moreover, note that the detection probability increases from 0 (for d-ary trees with $d = 2$) all the way to 1 as one goes from d-ary trees to recursive trees to generalized PORTs.

Part (a) of Theorem 3 will follow from a result on a more general family of random trees: the subtree of the root has d_1 subtrees and all other subtrees have d_2 subtrees (subtrees are possibly empty). The random model of this family of trees is as follows: the first node is the root and d_1 empty leafs are attached; for the next node, one leaf is chosen uniformly at random and the next node together with d_2 empty leafs is placed there; for the third node, again one leaf is chosen uniformly at random, etc. Then, we have the following result.

Theorem 4. *We have,*

$$\lim_{n\to\infty} P(C_n) = 1 - d_1 + d_1 I_{1/2}\left(\frac{1}{d_2 - 1}, \frac{d_1 - 1}{d_2 - 1}\right),$$

where $I_x(a, b)$ is the regularized incomplete beta function.

Note the for $d_1 = d_2 = d$, we obtain the above result for d-ary trees. Moreover, this result also contains one of the main results from [10], namely, $d_1 = d$ and $d_2 = d - 1$ which are d-regular trees.

Theorem 5 (d-regular Trees; see also [11])**.** *For d-regular trees, we have $\lim_{n\to\infty} P(C_n) = k_{d\text{-reg}}$, where*

$$k_{d\text{-reg}} = 1 - \frac{d}{2} + \frac{d\Gamma(\frac{d}{d-2})}{2^{\frac{d}{d-2}}\Gamma(\frac{d-1}{d-2})^2}.$$

Remark 2. As observed in [11], Stirling's formula implies that

$$\lim_{n\to\infty} k_{d\text{-reg}} = 1 - \ln 2.$$

Hence, recursive trees are also the limiting case of d-regular trees as d tends to infinity (this is of course not surprising).

We conclude the introduction with a brief sketch of the paper. In the next section, we prove Theorem 4. In contrast to [11] this will be done by using tools from Analytic Combinatorics (in [11] the authors used Pólya urn models and tools from the theory of stochastic processes). As a consequence, we will obtain part (a) of Theorem 3 and Theorem 5. In Section 3, we will prove part (b) of Theorem 3. In Section 4, we will prove part (c) of Theorem 4. Finally, in an appendix, we will give a brief discussion of the rumor center for rooted unordered trees.

2 Generalized d-ary Trees

In this section, we will prove Theorem 4.

We start by fixing some notation. First, recall the definition of the trees from Theorem 4 (see the paragraph preceding the theorem). The number of these trees with n nodes will be denoted by $\tilde{\tau}_n$. Moreover, we will denote by τ_n the number of d_2-ary trees with n nodes . Then, observe that

$$\tilde{\tau}_n = \sum_{j_1+\cdots+j_{d_1}=n-1} \binom{n-1}{j_1,\ldots,j_{d_1}} \tau_{j_1}\cdots\tau_{j_{d_1}},$$

where $j_1,\ldots,j_{d_1} \geq 0$ are the sizes of the d_1 subtrees of the root and $\tau_0 := 1$. Consequently,

$$\tilde{\tau}'(z) = (1+\tau(z))^{d_1}, \tag{1}$$

where $\tau(z)$ is as in the introduction and

$$\tilde{\tau}(z) = \sum_{n\geq 1} \tilde{\tau}_n \frac{z^n}{n!}.$$

Recall that

$$\tau(z) = -1 + (1-(d_2-1)z)^{-1/(d_2-1)}. \tag{2}$$

Now, we turn to the probability of C_n. By Theorem 2, we have

$$P(C_n) = 1 - d_1 P(\text{size of leftist subtree} \geq n/2). \tag{3}$$

Denote by I the size of the leftist subtree. Then,

$$\begin{aligned}
P(I=j) &= \frac{1}{\tilde{\tau}_n} \sum_{j+j_2+\cdots+j_{d_1}=n-1} \binom{n-1}{j,j_2,\ldots,j_{d_1}} \tau_j \tau_{j_2}\cdots\tau_{j_{d_1}} \\
&= \frac{(n-1)!\tau_j}{j!\tilde{\tau}_n} \sum_{j_2+\cdots+j_{d_1}=n-1-j} \frac{\tau_{j_2}}{j_2!}\cdots\frac{\tau_{j_{d_1}}}{j_{d_1}!} \\
&= \frac{(n-1)!\tau_j}{j!\tilde{\tau}_n} [z^{n-1-j}](1+\tau(z))^{d_1-1} \\
&= \frac{(n-1)!\tau_j}{j!\tilde{\tau}_n} [z^{n-1-j}](1-(d_2-1)z)^{-\frac{d_1-1}{d_2-1}} \\
&= \frac{(n-1)!\tau_j}{j!\tilde{\tau}_n} (d_2-1)^{n-1-j}[z^{n-1-j}](1-z)^{-\frac{d_1-1}{d_2-1}}. \tag{4}
\end{aligned}$$

In the sequel, we need the following standard lemma from analytic combinatorics.

6

Theorem 6 (Theorem VI.1 in [5]). *For $\alpha \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$ set*

$$f(z) := (1 - z)^{-\alpha}.$$

Then, as $n \to \infty$,

$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1 + \sum_{k=1}^{\infty} \frac{e_k(\alpha)}{n^k}\right),$$

where $e_k(\alpha)$ is a polynomial of degree $2k$.

Applying this result to (2) gives

$$\tau_n = n![z^n]\tau(z) \sim n!(d_2 - 1)^n \frac{n^{\frac{1}{d_2-1}-1}}{\Gamma(\frac{1}{d_2-1})}.$$

Similarly, applying the result to (1) yields

$$\tilde{\tau}_n = (n-1)![z^{n-1}]\tilde{\tau}'(z) \sim (n-1)!(d_2-1)^{n-1}\frac{n^{\frac{d_1}{d_2-1}-1}}{\Gamma(\frac{d_1}{d_2-1})}.$$

By (3), we need to compute

$$\sum_{n/2 \leq j \leq n-1} P(I = j),$$

where $P(I = j)$ is given by (4). To accomplish this task, we again use Theorem 1 and the expansions for τ_n and $\tilde{\tau}_n$ from above. This gives

$$\sum_{n/2 \leq j \leq n-1} P(I = j) \sim \frac{(n-1)!}{\tilde{\tau}_n} \sum_{n/2 \leq j \leq n-1} \frac{\tau_j}{j!}(d_2-1)^{n-1-j}\frac{(n-1-j)^{\frac{d_1-1}{d_2-1}-1}}{\Gamma(\frac{d_1-1}{d_2-1})}$$

$$\sim \frac{\Gamma(\frac{d_1}{d_2-1})}{\Gamma(\frac{1}{d_2-1})\Gamma(\frac{d_1-1}{d_2-1})} \cdot \frac{1}{n^{\frac{d_1}{d_2-1}-1}} \sum_{n/2 \leq j \leq n-1} j^{\frac{1}{d_2-1}-1}(n-1-j)^{\frac{d_1-1}{d_2-1}-1}$$

$$\sim \frac{\Gamma(\frac{d_1}{d_2-1})}{\Gamma(\frac{1}{d_2-1})\Gamma(\frac{d_1-1}{d_2-1})} \cdot \frac{1}{n} \sum_{n/2 \leq j \leq n-1} \left(\frac{j}{n}\right)^{\frac{1}{d_2-1}-1}\left(\frac{n-1-j}{n}\right)^{\frac{d_1-1}{d_2-1}-1}$$

$$\sim \frac{\Gamma(\frac{d_1}{d_2-1})}{\Gamma(\frac{1}{d_2-1})\Gamma(\frac{d_1-1}{d_2-1})} \int_{1/2}^{1} x^{\frac{1}{d_2-1}-1}(1-x)^{\frac{d_1-1}{d_2-1}-1}\mathrm{d}x. \tag{5}$$

Observe that

$$\int_{1/2}^{1} x^{\frac{1}{d_2-1}-1}(1-x)^{\frac{d_1-1}{d_2-1}-1}\mathrm{d}x = \frac{\Gamma(\frac{1}{d_2-1})\Gamma(\frac{d_1-1}{d_2-1})}{\Gamma(\frac{d_1-1}{d_2-1})} - B\left(1/2; \frac{1}{d_2-1}, \frac{d_1-1}{d_2-1}\right),$$

where $B(x; a, b)$ denotes the incomplete beta function. Plugging this into (5) and (5) in turn into (3) yields Theorem 4.

Proof of Theorem 3, part (a). Setting $d_1 = d_2 = d$ and evaluating the expression obtained in Theorem 4 yields the claimed result for $k_{d\text{-ary}}$. Moreover, the claims for $k_{d\text{-ary}}$ follow by simple calculus. ∎

Next, we consider the case of d-regular trees, where we set $d_1 = d$ and $d_2 = d - 1$. We need the following lemma.

Lemma 1. *For $\alpha > 0$,*

$$\int_{1/2}^1 x^{\alpha-1}(1-x)^\alpha \mathrm{d}x = \frac{1}{2}\left(B(\alpha, \alpha+1) - \frac{1}{\alpha 2^{2\alpha}}\right),$$

where $B(a, b)$ denotes the beta function.

Proof. First, observe that

$$B(\alpha, \alpha+1) = \int_0^1 x^{\alpha-1}(1-x)^\alpha \mathrm{d}x = \int_0^{1/2} x^{\alpha-1}(1-x)^\alpha \mathrm{d}x + \int_{1/2}^1 x^{\alpha-1}(1-x)^\alpha \mathrm{d}x.$$

Now, call the first and second integral on the right hand side L and R, respectively. By integration by parts and substitution, we have

$$L = \frac{1}{\alpha}x^\alpha(1-x)^\alpha \Big|_0^{1/2} + R.$$

Thus,

$$R = \frac{1}{2}\left(B(\alpha, \alpha+1) - \frac{1}{\alpha 2^{2\alpha}}\right)$$

which is the claimed result. ∎

This lemma can be used to evaluate the integral in (5). Plugging the result then in turn into (3) yields Theorem 5.

3 Recursive Trees

We consider now recursive trees and will prove part (b) of Theorem 3. Recall that

$$\tau(z) = \ln\left(\frac{1}{1-z}\right)$$

and thus $\tau_n = (n-1)!$. Similar to d-ary trees, we have that

$$P(C_n) = 1 - P(\text{one subtree of the root has size} \geq n/2).$$

In order to find the latter probability observe that at most one subtree of the root has size at least $n/2$. Consequently, since recursive trees are unordered trees, we can arrange the subtrees such that this subtree is the leftist one. Then,

$$P(\text{one subtree of the root has size} = j) = \frac{1}{\tau_n} \sum_{\ell \geq 1} \frac{1}{(\ell-1)!} \sum_{j+j_2+\ldots j_\ell = n-1} \binom{n-1}{j, j_2, \ldots, j_\ell} \tau_j \tau_{j_2} \ldots \tau_{j_\ell}$$

$$= \frac{(n-1)!\tau_j}{j!\tau_n}[z^{n-1-j}] \sum_{\ell \geq 1} \frac{\tau(z)^{\ell-1}}{(\ell-1)!}$$

$$= \frac{(n-1)!\tau_j}{j!\tau_n}[z^{n-1-j}]\frac{1}{1-z} = \frac{1}{j},$$

where $j \geq n/2$. Plugging this into the expression above gives

$$P(C_n) = 1 - \sum_{n/2 \leq j \leq n-1} \frac{1}{j} = 1 - H_{n-1} + H_{\lceil n/2 \rceil - 1},$$

where H_n denotes the n-th harmonic number. We summarize this in a result.

Theorem 7 (Recursive Trees). *We have, $P(C_n) = 1 - H_{n-1} + H_{\lceil n/2 \rceil - 1}$.*

From this part (b) of Theorem 3 follows from standard expansions for harmonic numbers.

8

4 Generalized Plane-oriented Recursive Trees

Finally, we consider generalized plane-oriented recursive trees and prove part (c) of Theorem 5. Recall that

$$\tau(z) = 1 - (1 - rz)^{1/r}. \tag{6}$$

Then, as in the last section

$$P(C_n) = 1 - P(\text{one subtree of the root has size} \geq n/2).$$

Since now the subtrees are ordered, we obtain

$$
\begin{aligned}
P(\text{one subtree of the root has size} = j) &= \frac{1}{\tau_n} \sum_{\ell \geq 1} \ell \phi_\ell \sum_{j+j_2+\ldots j_\ell = n-1} \binom{n-1}{j, j_2, \ldots, j_\ell} \tau_j \tau_{j_2} \cdots \tau_{j_\ell} \\
&= \frac{(n-1)! \tau_j}{j! \tau_n} [z^{n-1-j}] \sum_{\ell \geq 1} \ell \phi_\ell \tau(z)^{\ell-1} \\
&= \frac{(n-1)! \tau_j}{j! \tau_n} (r-1) [z^{n-1-j}] (1 - \tau(z))^{-r} \\
&= \frac{(n-1)! \tau_j}{j! \tau_n} r^{n-1-j} (r-1), \tag{7}
\end{aligned}
$$

where $j \geq n/2$.

We now turn to asymptotic expansions. First, applying Theorem 6 to (6) gives

$$\tau_n \sim -n! r^n \frac{n^{-1/r-1}}{\Gamma(-1/r)}.$$

Plugging this into (7) yields

$$
\begin{aligned}
P(\text{one subtree of the root has size} \geq n/2) &\sim \frac{r-1}{r} \cdot \frac{1}{n} \sum_{n/2 \leq j \leq n-1} \left(\frac{j}{n}\right)^{-1/r-1} \\
&\sim \frac{r-1}{r} \int_{1/2}^1 x^{-1/r-1} dx \\
&= (r-1)(2^{1/r} - 1).
\end{aligned}
$$

This proves the claimed limit result. The claimed properties for k_r follow by simple calculus.

Remark 3. Alternatively to the above asymptotic derivation, one can also derive an exact expression (similar as in the last section). To give more details, note that from (6), one obtains that

$$\tau_n = n! (-1)^{n+1} r^n \binom{1/r}{n}.$$

Consequently, from (7),

$$P(\text{one subtree of the root has size} \geq n/2) = \frac{r-1}{r} \cdot \frac{(-1)^n}{n \binom{1/r}{n}} \sum_{n/2 \leq j \leq n-1} (-1)^j \binom{1/r}{j}.$$

Note that

$$\sum_{n/2 \leq j \leq n-1} (-1)^j \binom{1/r}{j} = nr(-1)^{n+1} \binom{1/r}{n} + \lceil n/2 \rceil r (-1)^{\lceil n/2 \rceil} \binom{1/r}{\lceil n/2 \rceil}.$$

Plugging this into the above formula gives the following result.

9

Theorem 8 (Generalized PORTs). *We have,*

$$P(C_n) = r - (r-1)(-1)^{n+\lceil n/2 \rceil} \frac{\lceil n/2 \rceil \binom{1/r}{\lceil n/2 \rceil}}{n \binom{1/r}{n}}.$$

References

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

[2] F. Bergeron, P. Flajolet and B. Salvy (1992). Varieties of increasing trees, *Lectures Notes in Computer Science*, **581**, 24–48.

[3] Q. Feng and H. M. Mahmoud (2010). On the variety of shapes on the fringe of a random recursive tree, *J. Appl. Probab.*, **47:1**, 191-200.

[4] J. A. Fill (1996). On the distribution of binary search trees under the random permutation model, *Random Structures Algorithms*, **8:1**, 1-25.

[5] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*, Cambridge University Press, 2009.

[6] M. Kuba and A. Panholzer (2007). On the degree distribution of the nodes in increasing trees, *Journal of Combinatorial Theory, Series A*, **114**, 597–618.

[7] J. Moon (1974). The distance between nodes in recursive trees, *London Math. Soc. Lecture Notes*, **13**, Cambridge Univ. Press, London, 125–132.

[8] A. Panholzer and H. Prodinger (2007). The level of nodes in increasing trees revisited, *Random Structures and Algorithms*, **31**, 203–226.

[9] D. Shah and T. Zaman (2010). Finding sources of computer viruses in networks: Theory and experiment, in *Proc. ACM Sigmetrics*, **15**, 5249–5262.

[10] D. Shah and T. Zaman (2011). Rumors in a network: who is the culprit?, *IEEE Trans. Inform. Theory*, **57:8**, 5163–5181.

[11] D. Shah and T. Zaman. Finding rumor sources on random graphs, submitted.

Appendix: Rumor Center for Unordered Trees

In this appendix, we will use the same notation as in Section 1. Moreover, recall that $R(v, \Gamma)$ gives the number of rooted ordered increasing trees which are isomorphic to Γ^v.

If trees are now considered to be unordered instead of ordered, $R(v, \Gamma)$ has to be replaced by the shape functional for unordered trees, i.e., by the number of rooted unordered increasing trees which are isomorphic to Γ^v. This shape functional has been introduced and studied by Feng and Mahmoud in [3]. Following this paper, for $v \in V(\Gamma)$, we define now the score by

$$S(v, \Gamma) = (n-1)! \prod_{u \in V(\Gamma)} \frac{w(\Gamma_u^v)}{|\Gamma_u^v|},$$

10

where

$$w(\Gamma_u^v) = \prod_{i=1}^{r} \frac{1}{m_i!}$$

with $(m_1, \ldots, m_r)$ the multiplicities of the subtrees of u in Γ_u^v.

With a slight abuse of notation, we give the following definition.

Definition 3. Let Γ be a tree. A node $v \in V(\Gamma)$ is called a *local rumor center* if $S(v, \Gamma) \geq S(u, \Gamma)$ for all $u \in V(\Gamma)$ with $\{u, v\} \in E(\Gamma)$.

Then, as for ordered trees, we have the following theorem.

Theorem 9. *Let Γ be a tree. Then, every local rumor center is a rumor center.*

Proof. Denote by v the local rumor center and consider Γ^v. Let $u \in V(\Gamma)$ with $\{u, v\} \in E(\Gamma)$ be fixed. In the sequel, we will use the following notation: by m_u^v we denote the multiplicity of Γ_u^v amongst the subtrees of v in Γ^v.

First consider

$$\frac{S(v, \Gamma)}{S(u, \Gamma)} = \frac{m_v^u |\Gamma_v^u|}{m_u^v |\Gamma_u^v|} = \frac{m_v^u (n - |\Gamma_u^v|)}{m_u^v |\Gamma_u^v|},$$

where we used the (trivial) fact that $|\Gamma_v^u| + |\Gamma_u^v| = n$. Since, due to the assumptions, the above ratio must be at least one, we have

$$|\Gamma_u^v| \leq \frac{m_v^u n}{m_v^u + m_u^v}. \tag{8}$$

We will fix now an $\tilde{u} \in V(\Gamma)$ with $\tilde{u} \neq v$ and $\{\tilde{u}, u\} \in E(\Gamma)$; see the tree on the left in Figure 1.
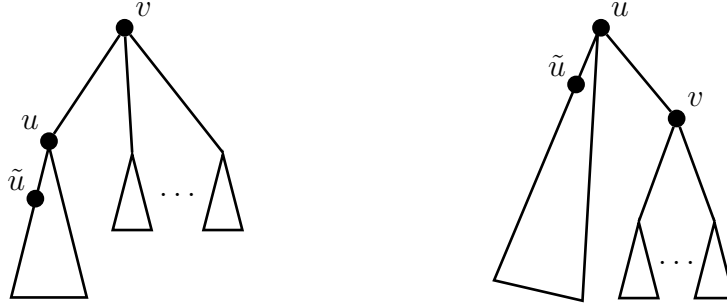

Figure 1: The two trees from the proof of Theorem 9

Observe that

$$\frac{S(v, \Gamma)}{S(\tilde{u}, \Gamma)} = \frac{S(v, \Gamma)}{S(u, \Gamma)} \cdot \frac{S(u, \Gamma)}{S(\tilde{u}, \Gamma)} \geq \frac{m_u^{\tilde{u}} (n - |\Gamma_{\tilde{u}}^u|)}{m_{\tilde{u}}^u |\Gamma_{\tilde{u}}^u|} \tag{9}$$

and we have to show that this is at least one. For this, we will consider two cases.

In the first case, we will assume that $|\Gamma_u^v| \leq n/2$. Then, $|\Gamma_v^u| \geq n/2$ and hence $m_v^u = 1$. Similarly, $|\Gamma_u^{\tilde{u}}| \geq n/2$ and $m_u^{\tilde{u}} = 1$. Moreover, we have

$$m_{\tilde{u}}^u |\Gamma_{\tilde{u}}^u| \leq |\Gamma_u^v| \leq n/2.$$

This implies that $|\Gamma_{\tilde{u}}^u| \leq n/(2m_{\tilde{u}}^u)$ which in turn implies that $|\Gamma_{\tilde{u}}^u| \leq n/(1 + m_{\tilde{u}}^u)$. Thus, (9) is indeed at least one and hence $S(v, \Gamma) \geq S(\tilde{u}, \Gamma)$.

11

It should be clear that the above argument can be repeated. Consequently, if we choose a path from v via $\tilde{u}$ to a leaf, then the S-value of the nodes is non-increasing as required.

Next, we consider the second case, where we assume that $|\Gamma_u^v| > n/2$. Consider the tree on the right in Figure 1 which is just the left one rooted at u (i.e., Γ^u). Due to (8), we have that

$$|\Gamma_v^u| \geq \frac{n}{m_v^u + m_u^v}. \tag{10}$$

If $\Gamma_{\tilde{u}}^u$ is isomorphic to Γ_v^u, then the S-value of $\tilde{u}$ and v are the same and nothing has to be proved. So, assume that $\Gamma_{\tilde{u}}^u$ is not isomorphic to Γ_v^u. Then, a simple counting argument shows that

$$m_{\tilde{u}}^u |\Gamma_{\tilde{u}}^u| + m_v^u |\Gamma_v^u| \leq n.$$

Using (10), we obtain

$$|\Gamma_{\tilde{u}}^u| \leq \frac{n}{m_{\tilde{u}}^u (m_v^u + m_u^v)} \leq \frac{n}{1 + m_{\tilde{u}}^u}.$$

This in particular implies that $|\Gamma_{\tilde{u}}^u| \leq n/2$ and hence $m_u^{\tilde{u}} = 1$. Consequently, (9) is again shown to be at least one. Finally, similar to the first case, this argument can be iterated such that again the S-value of nodes along paths from v to a leaf (via $\tilde{u}$) are non-increasing. ∎

Remark 4. From the proof, we find the following sufficient and necessary condition for $v \in V(\Gamma)$ to be a rumor center (compare with Theorem 2 from the introduction).

Theorem 10. *Let Γ be a tree with n nodes. Then, $v \in V(\Gamma)$ is a rumor center of Γ if and only if the following holds for all $u \in V(\Gamma)$ with $\{u, v\} \in E(\Gamma)$:*

$$|\Gamma_u^v| \leq \frac{m_v^u n}{m_v^u + m_u^v}.$$

Remark 5. In contrast to the ordered case, here Γ can have more than two rumor centers; also, rumor centers are not necessary adjacent; see Figure 2 for examples.
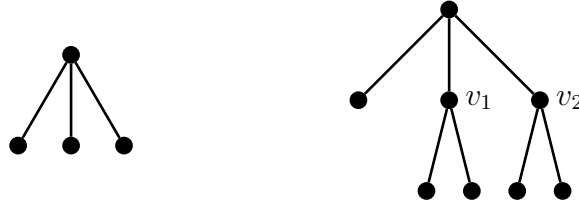

Figure 2: Every node of the tree on the left is a rumor center; the nodes v_1 and v_2 of the tree on the right are (non-adjacent) rumor centers.

Remark 6. As mentioned in Section 1, the random model of recursive trees is the uniform model on rooted unordered increasing trees. Thus, the rumor center defined in this appendix is the ML estimator for the rumor source. It would be interesting to compute the detection probability for this estimator (which will be at least $1 - \ln 2$ as follows from Theorem 3). However, the more complicated characterization of the rumor center from Remark 4 makes this a seemingly complicated task.