Estimating the history of a random recursive tree*

Simon Briend
Université Paris Saclay, CNRS,
Laboratoire de Mathématiques d'Orsay,
91404, Orsay, France,

Christophe Giraud
Université Paris Saclay, CNRS,
Laboratoire de Mathématiques d'Orsay,
91404, Orsay, France,

Gábor Lugosi
Department of Economics and Business,
Pompeu Fabra University, Barcelona, Spain
ICREA, Pg. Lluís Companys 23, 08010 Barcelona,
Spain Barcelona School of Economics

Déborah Sulem
Barcelona School of Economics
08005, Barcelona, Spain

Abstract

This paper studies the problem of estimating the order of arrival of the vertices in a random recursive tree. Specifically, we study two fundamental models: the uniform attachment model and the linear preferential attachment

*Simon Briend acknowledges the support of Région Ile de France. Christophe Giraud acknowledges partial support by grant ANR-19-CHIA-0021-01 (BiSCottE, ANR) and by grant ANR-21-CE23-0035 (ASCAI, ANR and DFG). Gábor Lugosi acknowledges the support of Ayudas Fundación BBVA a Proyectos de Investigación Científica 2021 and the Spanish Ministry of Economy and Competitiveness grant PID2022-138268NB-I00, financed by MCIN/AEI/10.13039/501100011033, FSE+MTM2015-67304-P, and FEDER, EU.).

model. We propose an order estimator based on the Jordan centrality measure and define a family of risk measures to quantify the quality of the ordering procedure. Moreover, we establish a minimax lower bound for this problem, and prove that the proposed estimator is nearly optimal. Finally, we numerically demonstrate that the proposed estimator outperforms degree-based and spectral ordering procedures.

1 Introduction

Recursive random graphs are often used to model the evolution of relational data over time, with applications ranging from epidemiology to information technology. When the history of the network is not observed, the task of inferring past states of the network is often termed *network archaeology* (Navlakha and Kingsford [23]). Recovering hidden states of the growing network informs on the underlying spreading process and can explain the current network structure. One of the most studied problems in network archaeology is finding the root of the recursive network, that is, the vertex that first appeared and "started" the growing process. This task is closely related to the rumor source detection problem. It has been extensively analyzed for random recursive tree models, see Devroye and Reddad [12], Lugosi and Pereira [20], Haigh [16], Shah and Zaman [28, 27], Bubeck et al. [7, 6], Brandenberger et al. [4], Jog and Loh [18, 19], Banerjee and Bhamidi [2], Contat et al. [9] .

In this paper, we consider the problem of estimating the entire history of the network, that is, the arrival times of all the vertices in a random recursive tree. One may consider this as a question of latent variable estimation. A related statistical problem is the so-called *seriation*. Seriation is the problem of inferring an ordering of points, based on pairwise similarity or on the adjacency information between two points. This similarity measure is assumed to statistically decrease with the distance in a latent space and informs on the latent global order of the points. The seriation problem has been studied in various fields, such as in archaeology (Robinson [26]), bioinformatics (Recanati et al. [24]), and matchmaking (Bradley and Terry [3]). It has been theoretically analyzed in random graph models such as geometric graphs and graphons (Giraud et al. [15], Janssen and Smith [17]). In recursive trees, the pairwise affinity between nodes is encoded in the adjacency matrix, and the latent space and latent positions are respectively the temporal line and the arrival times of the vertices. Estimating the temporal order of the vertices in a recursive tree can therefore be interpreted as an instance of the seriation problem.

To estimate the vertices' order, we propose a procedure based on a centrality measure, specifically on the *Jordan centrality*. We prove that this procedure is nearly optimal in two random recursive tree models, namely, the uniform random recursive tree (URRT) and the preferential attachment (PA) model. In these models, a tree of $n \geq 1$ vertices is grown by adding and connecting one vertex at each time step. To describe the growing process, we assume that the vertices have intrinsic *labels* from 1 to n. At each step $t = 1, \ldots, n$ of the growth, a new vertex, say of label j_t, is picked arbitrarily among the set of nodes not yet in the tree, and added to the tree with the *rank* t. At $t = 1$, the first sampled vertex is the root of the tree. We denote by $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ the ordering (or, ranking) map of vertices such that $\sigma(j_t) = t$. In other words, σ is a permutation.

The URRT and PA models differ by the attachment rule used to connect a new vertex at each step $t = 2, \ldots, n$ of the growth process. In the URRT model, the vertex j_t is connected by an undirected edge to a vertex sampled uniformly among the vertices of the current tree. In the PA model, the vertex of the tree is sampled with a probability proportional to its degree. We denote by $T = T_n$ the obtained tree *structure*, that is, the set of nodes with labels in $\{1, \ldots, n\}$ and the undirected edges between them. In the statistical problem considered in this paper, after the tree is grown, the rank or arrival time $\sigma(i)$ of each node i is not observed on T. The random growing process defines a probability distribution on trees. We denote by $\mathbb{P}$ the corresponding probability distribution and $\mathbb{E}$ the associated expectation.

We note that in these random models, the sampling process is independent of the labels chosen to identify the vertices, here $\{1, \ldots, n\}$. Therefore, any coherent ordering procedure should be *label invariant*, that is, independent of these labels. Saying that $\widehat{\sigma}$ is label invariant means that for any fixed T and ranking σ,

$$\widehat{\sigma}(T, \sigma) \stackrel{\mathcal{L}}{=} \widehat{\sigma}(T^{\sigma'}, \sigma \circ \sigma') \,, \tag{1}$$

for a permutation σ', where $T^{\sigma'}$ denotes the tree with label i replaced by $\sigma'(i)$. Note that the equality in distribution is a simple equality if the ordering procedure $\widehat{\sigma}$ is deterministic. Let us also remark that an easy way to transform any ordering procedure into a label invariant ordering procedure is by applying a random permutation to the labels of the tree before feeding it to the ordering procedure.

In order to measure the quality of an estimator of the history, we introduce a family of risk measures that takes into account the error in the estimated arrival time of each vertex, weighted by a function of the arrival time. We define the following family of risk measures

$$R_\alpha(\widehat{\sigma}) \stackrel{\text{def}}{=} \mathbb{E}\left[\sum_{i=1}^{n} \frac{|\widehat{\sigma}(i) - \sigma(i)|}{\sigma(i)^\alpha}\right] , \tag{2}$$

where $\alpha > 0$. The parameter α tunes the importance given to vertices with small true rank $\sigma(i)$: the higher α, the more weight is given to vertices with low rank. Perhaps the most natural choice is $\alpha = 1$. In that case the risk corresponds to normalizing the error on the estimation of the arrival time of a vertex by its true arrival time. We note that it is often the early stages of a propagation phenomenon that are more relevant, for example, for designing prevention strategies. Additionally, in random growing trees, it is harder to accurately order the high-rank vertices, due to the inherent model symmetries (Sreedharan et al. [29]).

One way to construct an estimator $\widehat{\sigma}$ of the ranking map is to choose a score function on the set of vertices, and order vertices by increasing (or decreasing) values. Such a score function could be based on the likelihood under the tree model. However, the latter is generally difficult to compute, see Bubeck et al. [6]. Instead, score functions based on the degree (Navlakha and Kingsford [23]) or the so-called rumor centrality (Cantwell et al. [8]) can be computed in polynomial time.

Another approach are iterative algorithms that recursively infer previous states of the tree such as the history sampling algorithms (Crane and Xu [10], Cantwell et al. [8]) and the Peeling procedure (Sreedharan et al. [29]), which is related to the depth centrality score. These methods are guaranteed to recover a recursive ordering of the vertices. Moreover, Crane and Xu [10] show that the history sampling algorithm outputs confidence sets for the arrival time of a single vertex with valid frequentist coverage. Besides, Sreedharan et al. [29] demonstrate that the partial ordering retrieved by the Peeling procedure has good properties in settings where the root of the tree can be unambiguously identified. Nonetheless, there are not yet guarantees on the quality of the global ordering provided by these methods.

The ordering procedure we propose is based on the Jordan centrality, defined, for a vertex $u \in T$ belonging to a tree T, as

$$\psi_T(u) = \max_{v \in V(T),\ v \sim u} |(T,u)_v| \,. \tag{3}$$

where (T,u) denotes the tree T rooted at u, where $u \sim v$ means that u and v are neighbors in T, and where $(T,u)_v$ denotes the subtree of T containing all vertices w such that v lies on the path connecting w to u (see Figure 1). Somewhat informally, we call $(T,u)_v$ the subtree hanging from v in the rooted tree (T,u). The maximum in (3) is taken over vertices v of the tree that are connected to vertex u by an edge. Intuitively, if a vertex is *central*, then none of the subtrees hanging from it can be too large. Therefore, the lower $\psi_T(i)$, the more central is vertex i. It is straightforward to see that $(\psi_T(u))_{u\in T}$ only depends on the structure of the tree and not on the labels of its vertices. We then define $\widehat{\sigma_J} : \{1,\dots,n\} \to \{1,\dots,n\}$ the ordering obtained by ranking the vertices by increasing value of Jordan centrality–breaking ties at random. This estimator is label invariant. An equivalent formulation of this algorithm is to estimate the position of vertex 1 by the Jordan centroid, rooting the tree at this vertex and then ordering vertices by the size of their hanging subtree in the rooted tree. Thus, if the exact position of vertex 1 was known, we would be ordering vertices by the number of their descendants.

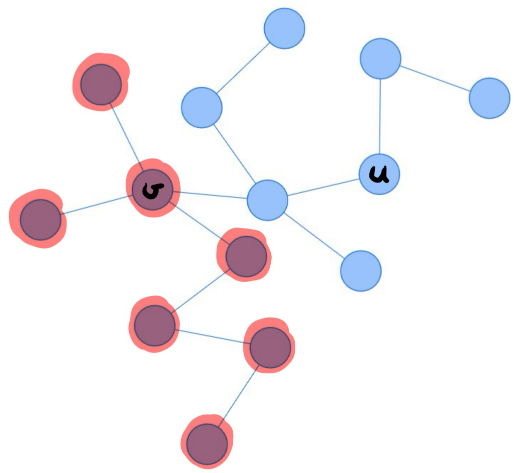

Figure 1: An illustration of the subtree $(T, u)_v$, corresponding to nodes highlighted in red.

While the risk defined in (2) can be computed for any value $\alpha > 0$, we restrict α to a range of values which are relevant for our ordering problem. Specifically, we only consider $\alpha \geq 1$, since for $\alpha < 1$ the problem becomes trivial, since even a random permutation has a risk which is minimax optimal up to constant factor (see Appendix A.1). In Theorems 1 and 7, we provide minimax lower bounds for the risk $R_\alpha(\widehat{\sigma})$ in the URRT and PA model, for any label-invariant estimator $\widehat{\sigma}$. Then, in Theorems 4 and 8 upper bounds for the risk of the Jordan ordering are obtained. Finally, in Corollaries 5 and 9 we prove that our proposed estimator is minimax optimal up to constant factors, in a non-trivial range of parameters α. In the following table, we summarise our findings. For $\alpha \geq 1$, we denote by R_α^* the optimal risk, and $R_\alpha(\widehat{\sigma}_J)$ the risk of the Jordan ordering.

	URRT	PA
R_α^*	$\geq n^{2-\alpha}/65 \vee 1/2$	$\geq n^{2-\alpha}/70 \vee 1/2$
$R_\alpha(\widehat{\sigma}_J)$	$= \mathcal{O}\Big(n^{2-\alpha} + \log^4(n)\Big)$	$= \mathcal{O}\Big(n^{2-\alpha} + n^{3/4}\Big)$

We also compare numerically the performance of the Jordan estimator with other ordering procedures in a simulation study.

In the rest of this section, we review previous works and introduce some notation. Then, in Section 2, we analyze the Jordan ordering in the URRT model. Next, we consider the PA model and prove analogous results in Section 3. Finally, in Section 4, we report the results of our simulation study and compare the empirical performance of the Jordan estimator to alternative methods based on the degree centrality, a peeling method (Navlakha and Kingsford [23, Section 2.3]) and a spectral method commonly used in seriation problems (Recanati et al. [25]).

1.1 Related work

Most methods for ranking the vertices of a random recursive tree have been introduced for the root-finding problem, that is, recovering a vertex (or a set of vertices) that is (contains) the root. For this problem, maximum likelihood estimators (Bubeck et al. [6], Haigh [16], Brandenberger et al. [4]) and estimators based on rumor centrality (Shah and Zaman [28, 27]) have been proposed and analyzed. Jordan centrality is another measure of centrality used by Bubeck et al. [7, 6] to construct confidence sets. Jog and Loh [18, 19], Banerjee and Bhamidi [2] study the persistence of the most central nodes in random recursive trees. Furthermore, while the vertex with maximum degree is generally not a good estimator of the root in the URRT model, in the PA model pairs degree centrality is useful for retrieving the first vertex (Banerjee and Bhamidi [1], Contat et al. [9]). Some recent work studies root-finding in Galton-Walton trees (Brandenberger et al. [4]) and more general graphs (Crane and Xu [11], Briend et al. [5]).

Crane and Xu [10] propose a general history-sampling procedure for network archaeology, which can be applied to the problem of estimating arrival times. The history sampling algorithm outputs a confidence set of rankings that contains the true one with high probability. However, there is no known bound of the size nor the average global error of an ordering in this confidence set.

The vertex arrival-time estimation problem bears some similarity to the seriation problem, though in the former, the dependence is intrinsically related to the tree structure. For example, in a random geometric graph (Gilbert [14]), the seriation problem is to estimate the position of the random points. Since there is no time structure in seriation, different metrics for the error are used, such as the maximum distance between the true and estimated latent position. Examples of methods are provided by Giraud et al. [15]. Another widely studied seriation method consists in ordering latent points by a spectral method on the graph Laplacian (see Section 4 and Recanati et al. [25] for details). They give guarantees for the quality of their method when the observed adjacency matrix is a perturbed Robinson matrix. The expected adjacency matrices of both URRT and PA trees are Robinson, and therefore the models studied here can be viewed as perturbed Robinson matrices. Nonetheless, none of the above-mentioned papers gives any insight about seriation in URRT and PA trees.

1.2 Notation

Let π_n be the set of permutations of $[n] := \{1, 2, \ldots, n\}$, and let $\mathbb{T}_n$ be the set of unlabelled trees of size n. We denote by $\textsc{urrt}(n)$ the distribution of a tree $\mathcal{T}_n$ of size $n \geq 1$, generated from the uniform attachment model. Similarly, we denote by $\textsc{pa}(n)$ the distribution of a tree sampled from the preferential attachment model. Moreover, we decompose the tree as $\mathcal{T}_n = (T_n, \sigma_n)$, where T_n is the shape of the tree and σ_n is the recursive ordering of the vertices in $\mathcal{T}_n$. For simplicity, we drop the subscript n when the size of the tree is fixed and clear from the context. We denote by $\mathbb{P}$ the probability distribution under the tree growing process and $\mathbb{E}$ the corresponding expectation.

Recall that for a tree T and a vertex $u \in T$, we denote by (T, u) the tree rooted at u. For a rooted tree (T, u) and a vertex v we denote by $(T, u)_v$ the subtree of T consisting of all vertices w such that v lies on the path connecting w to u (see Figure 1). For two vertices $u, v \in T$, $u \sim v$ means that u is a neighbor of v in T (and reciprocally). In a rooted tree (T, u), we say that w is a child of v if w is in $(T, u)_v$. We denote by $\mathrm{de}_n(u) = |(\mathcal{T}_n, 1)_u|$ the number of descendants of u in $\mathcal{T}_n$. For simplicity, we drop the subscript n and use $\mathrm{de}(u)$ when the size of the tree is fixed and clear from the context.

Recall the definition of the Jordan centrality; for a tree T and vertex $u \in T$,

$$\psi_T(u) = \max_{v \in T, v \sim u} |(T, u)_v| \,. \tag{4}$$

We denote by c a centroid of T, defined as $c = \arg\min_{u \in T} \psi_T(u)$. It is well-known that any tree has at least one and at most two centroids. Moreover, for a vertex $u \in T$ that is not a centroid, the subtree $(T, u)_v$, $v \sim u$ with maximum size, contains all centroids.

The Jordan ordering procedure consists in ordering points by increasing values of ψ (ties being broken randomly). Equivalently, it consists in rooting the tree at c and order vertices by $|(T, c)|_u$. We use $\widehat{\sigma_J}$ to refer to the Jordan ordering of T_n. As noted in the introduction, the Jordan centrality does not depend on the labelling of the tree (only on its shape), and so is a label invariant ordering.

2 The uniform attachment model

In this section, we focus on the uniform attachment model as the random growing process of the tree. We first present a lower bound for the risk $R_\alpha(\widehat{\sigma})$ of any label-invariant estimator $\widehat{\sigma}$ of the vertices order.

2.1 A lower bound

In the next proposition, we provide a lower bound for the risk $R_\alpha(\widehat{\sigma})$ for any label-invariant estimator of the recursive ordering in the URRT model. Define, for any $n \geq 1$, the optimal risk by

$$R_\alpha^* := \min_{\widehat{\sigma} \in \Pi_n} R_\alpha(\widehat{\sigma}) \, ,$$

where Π_n is the set of label-invariant recursive orderings.

Theorem 1. *In the* URRT *model, we have, for all $\alpha > 0$ and $n \geq 200$,*

$$R_\alpha^* \geq \frac{n^{2-\alpha}}{65} \, .$$

Proof. For a tree T and an ordering of its vertices σ, let $\tau = \sigma^{-1}$ (i.e, $\tau(i)$ is the label of the vertex that arrives at time i). We start by recalling that

$$R_\alpha(\widehat{\sigma}) = \mathbb{E}\left[\sum_{j=1}^{n} \frac{|\widehat{\sigma}(j) - \sigma(j)|}{\sigma(j)^\alpha}\right] = \mathbb{E}\left[\sum_{j=1}^{n} \frac{|\widehat{\sigma} \circ \tau(j) - j)|}{j^\alpha}\right] ,$$

which is lower bounded as follows

$$R_\alpha(\widehat{\sigma}) \geq \sum_{j=\lfloor n/2 \rfloor+1}^{\lfloor 3n/4 \rfloor} \frac{|\widehat{\sigma} \circ \tau(j) - j)|}{j^\alpha} + \sum_{j=\lfloor 3n/4 \rfloor+1}^{n} \frac{|\widehat{\sigma} \circ \tau(j) - j)|}{j^\alpha} \tag{5}$$

$$\geq \frac{1}{n^\alpha} \sum_{j=\lfloor n/2 \rfloor+1}^{\lfloor 3n/4 \rfloor} \mathbb{E}\left[\left|\widehat{\sigma} \circ \tau(j) - j\right| + \left|\widehat{\sigma} \circ \tau\left(\left\lfloor \frac{n}{4} \right\rfloor + j\right) - \left\lfloor \frac{n}{4} \right\rfloor - j\right|\right] . \tag{6}$$

The problem is reduced to a control of each term of the summand. For a labelled tree T and a permutation γ, we denote by T^γ the tree with γ applied to its labels. For $j \geq \lceil n/2 \rceil$, fix $\gamma = (\tau(j), \tau(\lfloor n/4 \rfloor + j))$, that is, the permutation sending j to $\lfloor n/4 \rfloor + j$ and vice versa, while keeping all other elements of $[n]$ in place. Introduce the event

$$\Omega_j := \{\tau(j) \text{ and } \tau(\lfloor n/4 \rfloor + j) \text{ are leaves, connected to vertices of rank } \leq n/2\} \, .$$

First, we check that Ω_j is an event whose probability is bounded away from 0. We note that for Ω_j to occur, it suffices that

- vertex j connects to a vertex of rank at most $n/2$. This happens with probability $\lfloor n/2 \rfloor/(j-1)$.

- For times ranging from $j+1$ to $\lfloor n/4 \rfloor + j - 1$ new vertices connect to vertices different from j. This happens with probability

$$\prod_{k=j+1}^{\lfloor n/4 \rfloor + j - 1} \frac{k-2}{k-1} .$$

- Vertex $\lfloor n/4 \rfloor + j$ connects to a vertex of rank at most $n/2$. This happens with probability $\lfloor n/2 \rfloor / (\lfloor n/4 \rfloor + j - 1)$.

- For times ranging from $\lfloor n/4 \rfloor + j + 1$ to n new vertices connect to vertices different from j and $\lfloor n/4 \rfloor + j$. This happens with probability

$$\prod_{k=\lfloor n/4 \rfloor + j + 1}^{n} \frac{k-3}{k-1} .$$

Finally, note that from the definition of the URRT model, the four events corresponding to the four items above are independent. Thus

$$\begin{aligned}
\mathbb{P}\left\{\Omega_j\right\} &= \frac{\lfloor n/2 \rfloor}{j-1} \cdot \frac{\lfloor n/2 \rfloor}{\lfloor n/4 \rfloor + j} \cdot \prod_{k=j+1}^{\lfloor n/4 \rfloor + j - 1} \frac{k-2}{k-1} \cdot \prod_{k=\lfloor n/4 \rfloor + j + 1}^{n} \frac{k-3}{k-1} \\
&= \frac{\lfloor n/2 \rfloor}{j-1} \cdot \frac{\lfloor n/2 \rfloor}{\lfloor n/4 \rfloor + j} \cdot \frac{j-2}{\lfloor n/4 \rfloor + j - 2} \cdot \frac{(\lfloor n/4 \rfloor + j - 2)(\lfloor n/4 \rfloor + j - 1)}{(n-1)n} ,
\end{aligned}$$

which simplifies to

$$\mathbb{P}\left\{\Omega_j\right\} \geq \frac{1}{4}\left(1 - \frac{1}{n-2}\right)^3 . \tag{7}$$

The first step is to use (7) to control one of the summands in (5) by conditioning on Ω_j.

$$\begin{aligned}
&\mathbb{E}\left[|\widehat{\sigma} \circ \tau(j) - j| + |\widehat{\sigma} \circ \tau(\lfloor n/4 \rfloor + j) - (\lfloor n/4 \rfloor + j)|\right] \\
&\geq \frac{1}{4}\left(1 - \frac{1}{n-2}\right)^3 \mathbb{E}\left[\left|\widehat{\sigma} \circ \tau(j) - j\right| + \left|\widehat{\sigma} \circ \tau\left(\left\lfloor \frac{n}{4} \right\rfloor + j\right) - \left(\left\lfloor \frac{n}{4} \right\rfloor + j\right)\right| \;\middle|\; \Omega_j \right] .
\end{aligned}$$

We then decompose on each possible realization of a recursive tree

$$\begin{aligned}
&\mathbb{E}\left[\left|\widehat{\sigma} \circ \tau(j) - j\right| + \left|\widehat{\sigma} \circ \tau\left(\left\lfloor \frac{n}{4} \right\rfloor + j\right) - \left(\left\lfloor \frac{n}{4} \right\rfloor + j\right)\right| \;\middle|\; \Omega_j \right] \\
&= \sum_{t \in \mathbb{T}} \mathbb{P}\left\{T = t \mid \Omega_j\right\} \mathbb{E}\left[|\widehat{\sigma} \circ \tau(j) - j| \mid \Omega_j,\, T = t\right] \\
&\quad + \sum_{t \in \mathbb{T}} \mathbb{P}\left\{T = t^\gamma \mid \Omega_j\right\} \mathbb{E}\left[\left|\widehat{\sigma} \circ \tau\left(\left\lfloor \frac{n}{4} \right\rfloor + j\right) - \left(\left\lfloor \frac{n}{4} \right\rfloor + j\right)\right| \;\middle|\; \Omega_j,\, T = t^\gamma \right] ,
\end{aligned}$$

which is a valid decomposition since $t \mapsto t^\gamma$ is a bijection from $\mathbb{T}$ to itself. Theorem 4 of Crane and Xu [10] states that, in the URRT model, two trees having the same shape but different recursive orders have the same probability. Since on the event Ω_j, t is recursive if and only if t^γ is recursive, then

$$\mathbb{P}\left\{T = t \mid \Omega_j\right\} = \mathbb{P}\left\{T = t^\gamma \mid \Omega_j\right\} .$$

As a consequence, the above expression factorizes to

$$\mathbb{E}\left[|\widehat{\sigma}\circ\tau(j) - j| + \left|\widehat{\sigma}\circ\tau\left(\left\lfloor\frac{n}{4}\right\rfloor + j\right) - \left\lfloor\frac{n}{4}\right\rfloor - j\right| \;\Big|\; \Omega_j\right] = \sum_{t\in\mathbb{T}} \mathbb{P}\left\{T = t \mid \Omega_j\right\} \times$$
$$\left(\mathbb{E}\left[|\widehat{\sigma}\circ\tau(j) - j| \mid \Omega_j,\ T = t\right] + \mathbb{E}\left[\left|\widehat{\sigma}\circ\tau\left(\left\lfloor\frac{n}{4}\right\rfloor + j\right) - \left(\left\lfloor\frac{n}{4}\right\rfloor + j\right)\right| \;\Big|\; \Omega_j,\ T = t^\gamma\right]\right). \tag{8}$$

The label invariant condition implies that

$$\widehat{\sigma}[T^\gamma]\circ\gamma \stackrel{\mathcal{L}}{=} \widehat{\sigma}[T] ,$$

and in particular,

$$\left(\widehat{\sigma}(j) \mid \Omega_j, T = t\right) \stackrel{\mathcal{L}}{=} \left(\widehat{\sigma}(\lfloor n/4\rfloor + j) \mid \Omega_j, T = t^\gamma\right) ,$$

which directly implies that

$$\mathbb{E}\left[|\widehat{\sigma}\circ\tau(j) - j| \mid \Omega_j,\ T = t\right] + \mathbb{E}\left[\left|\widehat{\sigma}\circ\tau\left(\left\lfloor\frac{n}{4}\right\rfloor + j\right) - \left\lfloor\frac{n}{4}\right\rfloor - j\right| \;\Big|\; \Omega_j,\ T = t^\gamma\right] \geq \frac{n}{4} .$$

By plugging the above inequality in (8)

$$\mathbb{E}\left[|\widehat{\sigma}\circ\tau(j) - j| + \left|\widehat{\sigma}\circ\tau\left(\left\lfloor\frac{n}{4}\right\rfloor + j\right) - \left\lfloor\frac{n}{4}\right\rfloor - j\right|\right] \geq \frac{n}{16}\left(1 - \frac{1}{n-2}\right)^3 .$$

Now, plugging the above inequality in (5) yields

$$R_\alpha(\widehat{\sigma}) \geq \frac{n^{2-\alpha}}{65} ,$$

for all $n \geq 200$. □

Remark. In the URRT, vertices 1 and 2 are indistinguishable. Indeed, when the tree has size 2, vertices 1 and 2 have exactly the same properties. Thus, no label invariant ordering procedure can assign order 1 to vertex 1 with probability higher than 1/2. As a result, we obtain, for any α, the trivial lower bound

$$R_\alpha^* \geq \frac{1}{2} ,$$

which improves the bound of Theorem 1 for $\alpha \geq 2$, and therefore Theorem 1 is non-trivial when $\alpha < 2$.

2.2 An auxiliary "descendant-ordering" procedure

In the sequel, we estabish upper bounds for the risk of Jordan ordering. Since this is a label-invariant procedure, we may assume, without loss of generality, that $\sigma = \mathrm{Id}$ is the identity permutation. In other words, the arrival time of a vertex and its label are the same. When the context is clear, vertex labels and arrival times are used interchangeably.

In order to analyze the Jordan ordering, we introduce an auxiliary centrality measure and the corresponding estimator of vertex arrival times. As observed in the introduction, the Jordan ordering procedure consists in estimating the position of vertex 1 by the Jordan centroid c and ordering vertices according to the values of $|(T,c)_u|$. If c was replaced by vertex 1, this measure would correspond to the number of descendants of u. Thus, a natural ordering is to order vertices by the number of their descendants, that is, the ordering according to the values of $|(T,1)_u|$. We call this *descendant ordering*, noting that, as before, ties are broken at random. Note that descendant ordering is not a valid procedure, since the location of the root vertex is not known. On the other hand, the number of descendants is easily analyzed by Pólya urns, and our approach is based on comparing Jordan ordering to this auxiliary procedure. In this section we prove an upper bound for difference of the risk of both procedures. For a tree T and for each $u \in T$, we define the descendant centrality

$$\psi'_T(u) = n - \mathrm{de}(u) ,$$

where $\mathrm{de}(u) = |(T,1)_u|$ is the number of descendants of u, as defined in Section 1.2. We denote by $\widehat{\sigma}'$ the ordering of the vertices induced by sorting the values of ψ'_T in increasing order.

In the following lemma, we first prove that for the URRT model, the Jordan centrality ψ_T, defined in (3), and the descendant centrality ψ'_T coincide for most vertices. Furthermore, we prove bounds on both the number of nodes for which ψ_T may differ from ψ'_T and the estimated rank of vertex 1. We recall that 1 and c denote respectively the root and the rank of a centroid of the tree.

Lemma 2. *Let $T \sim$ URRT, let $c \in [n]$ be a centroid of T and let $\{1 \to c\}$ be the set of vertices on the path connecting 1 to c in T. Then*

- *for any $v \in [n] \backslash \{1 \to c\}$, we have*

$$\psi_T(v) = \psi'_T(v) ;$$

- *there exists a universal constant K such that c is stochastically dominated by an exponential random variable with mean K;*

- *for $\epsilon \le 0.2$, with probability at least $1 - 5\epsilon$*

$$\widehat{\sigma_J}(1) \le 2.5 \frac{\log(1/\epsilon)}{\epsilon} .$$

Proof. Let $T \sim$ URRT. First, we decompose the vertices of T in four sets as shown in Figure 2: case 1 corresponds to the set $\{1 \to c\}$ of nodes connecting the root to the centroid, case 2 to the vertices of $(T,1)_c \setminus \{c\}$, case 3 to the vertices of $(T,c)_1 \setminus \{1\}$ and finally case 4 to the vertices of $(T,1)_i \setminus \{i\}$ for $i \in \{1 \to c\} \setminus \{1,c\}$. As we mentioned before, it is well known that for a non-centroid vertex u, its neighbor maximizing $|(T,u)_v|$ is such that $|(T,u)_v|$ contains any centroid. Note that for each vertex u in cases 2, 3 and 4, for $v \sim u$ such that the subtree $(T,u)_v$ contains c, $(T,u)_v$ also contains vertex 1. As a consequence, $\psi_T(u) = |(T,u)_{\mathrm{pa}(u)}|$, where $\mathrm{pa}(u)$ is the "parent" of u. But by definition, $|(T,u)_{\mathrm{pa}(u)}| = n - \mathrm{de}(u) = \psi'_T(u)$, concluding the proof of the first part of the lemma.

Moon [22] showed that the rank of the centroid c is dominated by an exponential random variable of mean K, for a universal constant. The third statement follows from Theorem 3 of Bubeck et al. [6]. □

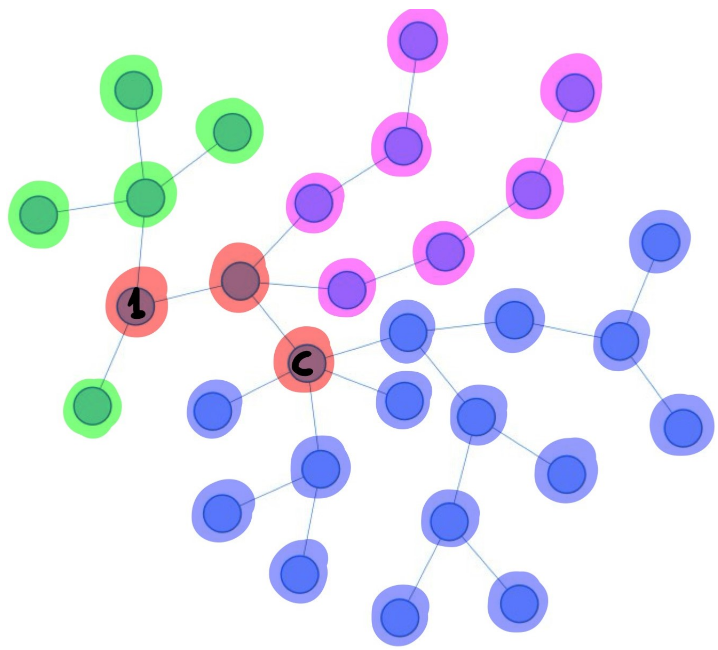

Figure 2: Sketch of a tree and its centroid. Circled in red are the vertices of the path $\{1 \to c\}$ (case 1). Blue vertices correspond to case 2, green to case 3 and purple vertices to case 4.

In the following lemma, we bound the risk of $\widehat{\sigma}_J$ by that of the descendant ordering $\widehat{\sigma}'$.

Lemma 3. *Let $T \sim$ URRT. For $\alpha > 0$*

$$R_\alpha(\widehat{\sigma_J}) \le R_\alpha(\widehat{\sigma}') + K\sum_{i=1}^{n}\frac{1}{i^\alpha} + C\log^4(n)\,,$$

where $C > 0$ is a constant (not depending on α).

Proof. Recall that $\sigma = \mathrm{Id}$, that is, we use the same integer to denote the label of a vertex and its arrival time. We first decompose the global risk $R_\alpha(\widehat{\sigma_J})$ into

$$R_\alpha(\widehat{\sigma_J}) = \mathbb{E}\left[\sum_{i\in\{1\to c\}}\frac{\left|\widehat{\sigma_J}(i)-i\right|}{i^\alpha}\right] + \mathbb{E}\left[\sum_{i\notin\{1\to c\}}\frac{\left|\widehat{\sigma_J}(i)-i\right|}{i^\alpha}\right].$$

Since by Lemma 2 vertices outside of the path $\{1 \to c\}$ are put in the same order by the Jordan ordering and the descendant ordering, for $i \notin \{1 \to c\}$, $|\widehat{\sigma_J}(i) - \widehat{\sigma}(i)| \le D+1$, where D is the distance between 1 and c. Thus, we can control the second term of the right-hand side as follows:

$$\begin{aligned}\mathbb{E}\left[\sum_{i\notin\{1\to c\}}\frac{\left|\widehat{\sigma_J}(i)-i\right|}{i^\alpha}\right] &= \mathbb{E}\left[\sum_{i\notin\{1\to c\}}\frac{\left|\widehat{\sigma}'(i)-i+\widehat{\sigma_J}(i)-\widehat{\sigma}'(i)\right|}{i^\alpha}\right]\\ &\le \mathbb{E}[D]\sum_{i=1}^{n}\frac{1}{i^\alpha} + \mathbb{E}\left[\sum_{i\notin\{1\to c\}}\frac{|\widehat{\sigma}'(i)-i|}{i^\alpha}\right].\end{aligned}$$

Since D is at most the arrival time of the centroid, Lemma 2 implies that $\mathbb{E}[D] \le \mathbb{E}[c] \le K$. On the other hand,

$$\mathbb{E}\left[\sum_{i\in\{1\to c\}}\frac{\left|\widehat{\sigma_J}(i)-i\right|}{i^\alpha}\right] \le \mathbb{E}\left[\sum_{i\in\{1\to c\}} i\right] + \mathbb{E}\left[\sum_{i\in\{1\to c\}}\widehat{\sigma_J}(i)\right].$$

Clearly,

$$\mathbb{E}\left[\sum_{i\in\{1\to c\}} i\right] \le \mathbb{E}[cD]\,.$$

Since $D \le c$ and since c is dominated by an exponential random variable of mean K, by Lemma 2,

$$\mathbb{E}[cD] \le \mathbb{E}\left[c^2\right] \le 2K^2\,.$$

In addition, since on the path $\{1 \to c\}$, $\widehat{\sigma_J}$ is decreasing, we have

$$\mathbb{E}\left[\sum_{i\in\{1\to c\}}\widehat{\sigma_J}(i)\right] \le \mathbb{E}\left[D\widehat{\sigma_J}(1)\right].$$

Since D and $\widehat{\sigma_J}(1)$ are bounded by n, they have finite moments. Using Hölder's inequality, for any $\gamma > 0$,

$$\mathbb{E}\left[D\widehat{\sigma_J}(1)\right] \le \left(\mathbb{E}\left[D^{\frac{1+\gamma}{\gamma}}\right]\right)^{\frac{\gamma}{1+\gamma}} \left(\mathbb{E}\left[\widehat{\sigma_J}(1)^{1+\gamma}\right]\right)^{\frac{1}{1+\gamma}} . \tag{9}$$

Since $D \le c$ which is dominated by an exponential random variable,

$$\left(\mathbb{E}\left[D^{\frac{1+\gamma}{\gamma}}\right]\right)^{\frac{\gamma}{1+\gamma}} \le C\frac{1+\gamma}{\gamma} , \tag{10}$$

for some positive constant C. Next, using Lemma 2,

$$\mathbb{P}\left\{\widehat{\sigma_J}(1) \ge f(\epsilon)\right\} \le 5\epsilon ,$$

where $f(\epsilon) = 2.5\frac{\log(1/\epsilon)}{\epsilon}$. f is a non-increasing function and therefore $f\left(5\log^2(k)/k\right) \le k$ for all $k \ge 1$. Therefore,

$$\mathbb{P}\left\{\widehat{\sigma_J}(1) \ge k\right\} \le 25\frac{\log^2(k)}{k} , \quad \text{for all } k \ge 1,$$

so for any $\gamma > 0$,

$$\mathbb{P}\left\{\widehat{\sigma_J}(1)^{1+\gamma} \ge k\right\} = \mathbb{P}\left\{\widehat{\sigma_J}(1) \ge k^{\frac{1}{1+\gamma}}\right\} \le 25\frac{\frac{1}{(1+\gamma)^2}\log^2(k)}{k^{\frac{1}{1+\gamma}}} .$$

It follows that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\sigma_J}(1)^{1+\gamma}\right] &= 1 + \int_{k=1}^{n^{1+\gamma}} \mathbb{P}\left\{\widehat{\sigma_J}(1)^{1+\gamma} \ge k\right\} dk \le 1 + \int_{k=1}^{n^{1+\gamma}} \frac{\frac{25}{(1+\gamma)^2}\log^2(k)}{k^{\frac{1}{1+\gamma}}} \, dk \\
&\le 1 + \frac{25}{(1+\gamma)^2}(1+\gamma)^2\log^2(n)\frac{\gamma+1}{\gamma}\left(n^{1+\gamma}\right)^{\frac{\gamma}{1+\gamma}} \\
&= 1 + 25\frac{(1+\gamma)\log^2(n)}{\gamma}n^{\gamma} .
\end{aligned}$$

Plugging the obtained inequality in (9) and recalling (10), we obtain

$$\mathbb{E}\left[D\widehat{\sigma_J}(1)\right] \le C\frac{1+\gamma}{\gamma}\left(1 + 25\frac{1+\gamma}{\gamma}\log^2(n)n^{\gamma}\right) .$$

Choosing $\gamma = 1/\log(n)$ we get

$$\mathbb{E}\left[D\widehat{\sigma_J}(1)\right] \le C'\log^4(n) .$$

This concludes the proof of the lemma. $\square$

2.3 Performance of Jordan ordering in the URRT model

In this section, we prove upper bounds for the risk $R_\alpha(\widehat{\sigma_J})$. In particular, we prove that for $\alpha \in [1,2)$, the risk $R_\alpha(\widehat{\sigma_J})$ has the same order as the optimal risk R^*_α, defined in Section 2.1.

Theorem 4. *Let $T \sim$ URRT. Then there exist positive constants C, K such that for $1 \le \alpha < 2$*

$$R_\alpha(\widehat{\sigma_J}) \le K(\alpha)n^{2-\alpha} + K\sum_{i=1}^{n}\frac{1}{i^\alpha} + C\log^4(n)\,,$$

where $K(\alpha) = \left(\frac{2}{2-\alpha} + \frac{2e^2}{(2-\alpha)^2} + \frac{2}{(2-\alpha)^3}\right)$. Moreover, for $\alpha \ge 2$

$$R_\alpha(\widehat{\sigma_J}) \le C\log^4(n)\,.$$

Before proving Theorem 4, we state a corollary that is a direct consequence of Theorems 1 and 4. This corollary notably implies that the Jordan ordering has a risk of optimal order for $\alpha \in [1,2)$. Note that one cannot hope to match the established lower bounds for the optimal risk in a broader range of α for this method. Indeed, in a uniform random recursive tree of size n, the probability that vertex 1 is a leaf is $1/n$. Since leaves are ordered last by $\widehat{\sigma_J}$, and that there are roughly $n/2$ leaves, $\mathbb{P}\left\{\widehat{\sigma_J}(1) \ge n/2\right\} \approx 1/n$. This implies that $\mathbb{E}[\widehat{\sigma_J}(1)] \gtrsim \log(n)/2$, so $\widehat{\sigma_J}$ has a risk of order at least $\log(n)$, while the lower boud is of constant order for $\alpha \ge 2$. We discuss in Appendix A.2 the possibility of estimating the position of vertex 1 better, namely using the rumor centroid. We conjecture this alternative method has a risk of optimal order for any $\alpha \ge 1$.

Corollary 5. *Let $T \sim$ URRT. For $\alpha = 1$*

$$R_\alpha(\widehat{\sigma_J}) \le (1+o(1))\,1170R^*_1$$

and for $\alpha \in (1,2)$,

$$R_\alpha(\widehat{\sigma_J}) \le (1+o(1))\left(\frac{1}{2-\alpha} + \frac{3}{(2-\alpha)^2} + \frac{1}{(2-\alpha)^3}\right)65R^*_\alpha\,.$$

Proof of Theorem 4. By the triangle inequality,

$$R_\alpha(\widehat{\sigma}') \le \sum_{i=1}^{n}\mathbb{E}\left[\frac{\widehat{\sigma}'(i)}{i^\alpha}\right] + \sum_{i=1}^{n}\frac{i}{i^\alpha}\,,$$

Let $i \in [n]$ be a vertex of T. We first note that

$$\mathbb{E}\left[\widehat{\sigma}'(i)\right] \le \mathbb{E}\left[\sum_{j:\ j\neq i} \mathbb{1}_{\mathrm{de}(j)\ge \mathrm{de}(i)}\right] . \tag{11}$$

Moreover, for any $\tau_{i,j} \in \mathbb{R}$,

$$\mathbb{P}\{\mathrm{de}(j) \ge \mathrm{de}(i)\} \le \mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \ge \tau_{i,j}\right\} + \mathbb{P}\left\{\frac{\mathrm{de}(i)}{n} \le \tau_{i,j}\right\} .$$

Therefore, we may upper bound (11) by

$$\mathbb{E}\left[\widehat{\sigma}'(i)\right] \le i+1+\sum_{j>i+1} \mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \ge \tau_{i,j}\right\} + \mathbb{P}\left\{\frac{\mathrm{de}(i)}{n} \le \tau_{i,j}\right\} . \tag{12}$$

Let $j > i+1$ be a vertex of T. Note that the distributions of $\mathrm{de}(i)$ and $\mathrm{de}(j)$ in a URRT model follow a Pólya urn model. In particular, for any vertex $k \in [n]$,

$$\mathbb{P}\{k \in (T,1)_v\} = \frac{de_{k-1}(j)+1}{k-1} ,$$

and each connection of a new vertex to a descendant of j is independent of the previous ones, conditionally on $\mathrm{de}(j)$. Let $N_n := n-j$ and let $\widetilde{W_N} = \mathrm{de}(j)$ be the number of descendants of j at time n. We thus have that $\widetilde{W_N} = \mathrm{de}(j)$ follows a Pólya urn distribution, where the Pólya urn process has balls of two colours, it is started when j is added to T, and it is run for a maximum number of steps N_n. From Mahmoud [21, Section 3.2], we have that

$$\mathbb{E}[\mathrm{de}(j)+1] \;=\; \frac{n}{j} ,$$

and also that

$$\mathbb{P}\{\mathrm{de}(j)=k\} \;=\; \frac{k!(j-1)(j)\cdots(n-k-2)}{j(j+1)\cdots(n-1)}\binom{n-j}{k}. \tag{13}$$

In Appendix A.4, we derive from these formulas the following upper-bounds.

Lemma 6. *$i \ge 2$ and $j > i+1$, by choosing $\tau_{i,j} = \frac{1}{j}\log\frac{j}{i}$,*

$$\mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \ge \tau_{i,j}\right\} + \mathbb{P}\left\{\frac{\mathrm{de}(i)}{n} \le \tau_{i,j}\right\} \le 2e^2 e^{-\log\frac{j}{i}} + \frac{i}{j}\log\frac{j}{i} \le \frac{i}{j}\left(2e^2+\log\frac{j}{i}\right),$$

and that for $i=1$, $j>i+1$, choosing $\tau_{1,j} = \frac{1}{j}\log(j)$,

$$\mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \ge \tau_{1,j}\right\} + \mathbb{P}\left\{\frac{\mathrm{de}(1)}{n} \le \tau_{1,j}\right\} \le \frac{1}{j}\left(2e^2+\log(j)\right) + \frac{1}{n-1} .$$

Once plugged into the expression of R_α, for $n \geq 60$, this leads to (details in Appendix A.4)

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i}\right] \leq 18n \,.$$

For $1 < \alpha < 2$, a similar computation yields (details in Appendix A.4)

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i^\alpha}\right] \leq \left(\frac{2}{2-\alpha} + \frac{2e^2}{(2-\alpha)^2} + \frac{2}{(2-\alpha)^3}\right) n^{2-\alpha} \,.$$

Lemma 3 concludes the proof.

3 Preferential attachment tree

In this section, we consider the preferential attachment model and investigate the performance of the Jordan ordering procedure. Since the arguments have a similar structure to the URRT model analyzed in Section 2, we omit some details of the proofs and report them to the Appendices. Similarly to the previous section, we first prove a minimax lower bound for the risk of any label-invariant estimator.

3.1 A lower bound

Theorem 7. *In the PA model, we have, for $\alpha = 1$ and $n \geq 300$*

$$R_\alpha^* \geq \frac{n^{2-\alpha}}{70} \,.$$

The proof is deferred to Appendix A.5.

Remark. In the same way as in the case of the URRT model, we have

$$R_\alpha^* \geq \frac{1}{2} \,,$$

which is better than the result of Theorem 7 for $\alpha > 2$.

3.2 Performance of the Jordan ordering in the PA model

Similarly to Section 2.3, we establish upper bounds for $R_\alpha(\widehat{\sigma_J})$. In a subsequent corollary, we bound the risk $R_\alpha(\widehat{\sigma_J})$ in terms of the optimal risk R_α^*.

Theorem 8. *Let $T \sim \mathrm{PA}$. Then, there exist positive constants C, K, such that for $\alpha \in [1, 5/4)$*

$$R_\alpha(\widehat{\sigma_J}) \le \left(\frac{2}{2-\alpha} + \frac{1}{(\alpha - 5/4)(\alpha-2)}\right) n^{2-\alpha} + K \sum_{i=1}^{n} \frac{1}{i^\alpha} + C \log^2(n)\sqrt{n} \,.$$

For $\alpha \ge 5/4$,

$$R_\alpha(\widehat{\sigma_J}) \le \frac{2}{2-\alpha} n^{2-\alpha} + \frac{32}{3}\zeta\left(\alpha - \frac{1}{4}\right) n^{3/4} \,,$$

where ζ denotes the Riemann zeta function.

Corollary 9 is a direct consequence of Theorems 7 and 8. It states that the Jordan ordering has a risk of optimal order for $\alpha \in [1, 5/4)$. Let us remark that, here, the boundary value 5/4 does not appear for the same reason as in the URRT case. In the URRT, the optimality result is limited to $\alpha < 2$ because of the error originating from the estimation of vertex 1. Here, the limitation to $\alpha < 5/4$ has a different origin than in the URRT model. Indeed, the descendant ordering is only order optimal for $\alpha < 5/4$, meaning that even if the position of vertex 1 was known, ordering vertices by the number of their descendants would not result in a risk bound that matches the lower bound for $\alpha \ge 5/4$.

Corollary 9. *Let $T \sim \mathrm{PA}$. For $\alpha \in [1, 5/4)$*

$$R_\alpha(\widehat{\sigma_J}) \le (1 + o(1))\, 70 \left(\frac{2}{2-\alpha} + \frac{1}{(\alpha - 5/4)(\alpha-2)}\right) R^*_\alpha \,.$$

The proof of Theorem 8 is reported to Appendix A.7.

4 Simulations

In this section, we first report a numerical illustration of our theoretical results on trees generated from the URRT and PA models. Then, we compare the performance of the Jordan ordering to other ordering procedures. For computational reasons, we display results for the descendant ordering procedure. The descendant ordering can be computed in time $\mathcal{O}(n \log(n))$. Also, one can find the Jordan centroid in linear time.

Note that the bounds of Lemmas 3 and 11 show that the risk of the descendant ordering is a good approximation of the risk of Jordan ordering.

In the first experiment, we compute the risk $R_\alpha(\widehat{\sigma}')$ (see (2)) of the descendant ordering and display the theoretical upper bound and minimax lower bound from Theorems 1 and 4 (Theorems 7 and 8 in the PA model).

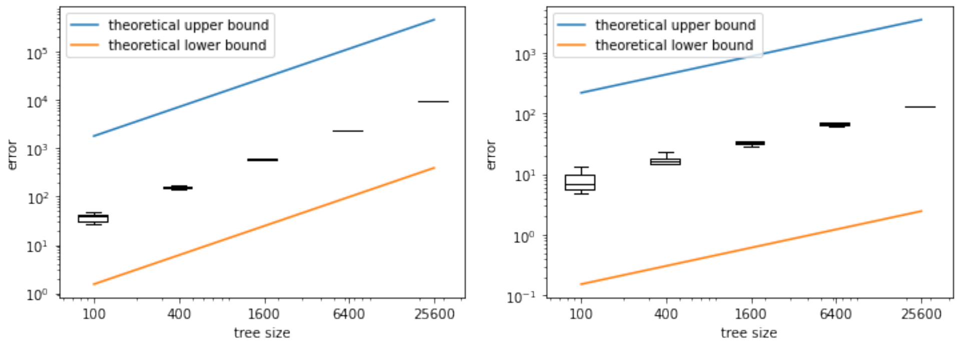

Figure 3: Risk R_α of the descendant ordering versus the tree size n in logarithmic scales, for $\alpha = 1$ (left panel) and for $\alpha = 1.5$ (right panel), and for trees simulated from the URRT model. Here, we sample 10 trees for each size, and report a boxplot with the median, first, and last quartiles, for each tree size.

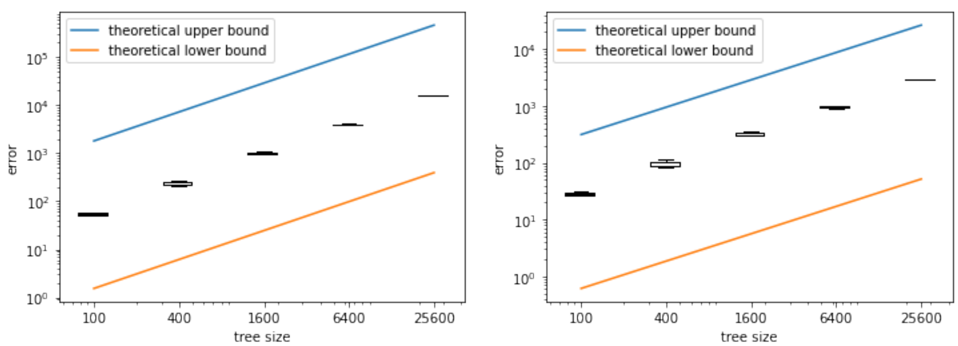

Figure 4: Risk R_α of the descendant ordering versus the tree size n in logarithmic scales, for $\alpha = 1$ (left panel) and for $\alpha = 1.2$ (right panel), and for trees simulated from the PA model. Here, we sample 10 trees for each size, and report a boxplot with the median, first, and last quartiles, for each tree size.

In the second experiment, we perform an empirical comparison of Jordan ordering with the three following ordering methods:

- **Degree** ordering, which orders the vertices by decreasing degree. Again, we break ties at random. Degree ordering is justified by the fact that the lower the rank of a vertex, the higher its expected degree is. Note, however, that ordering vertices by degree does not necessarily produce a recursive ordering.

- **Spectral** method by Recanati et al. [25]. This method is widely used in seriation problems and consists of finding the eigenvector associated to the second smallest eigenvalue of the Laplacian of the graph. Then, considering the entries of this eigenvector as a score function, the estimated ordering is derived by sorting these entries by increasing values.

- **Reverse DMC** algorithm, proposed by Navlakha and Kingsford [23]. This algorithm is analogous to a pruning method, which consists of ordering the vertices by sequentially removing all leaves from the tree and ordering the leaves removed at each step. In Reverse DMC, a score is computed for each leaf and the algorithm sequentially removes the leaf with the highest score. This score function corresponds to the likelihood of the leaf being the last vertex in the current tree, therefore, at each step, the leaf which is the most likely to be the last vertex arrived in the tree is removed.

Remark. We note that the spectral method is a reasonable method to compare with in our setting of recursive trees since (i) spectral methods recovers the order of a Robinson matrix Recanati et al. [25], and (ii) in the URRT and PA models, the expected value of the adjacency matrix is a Robinson matrix.

Similarly to the previous experiment, we compute the risk R_α for the four methods, on trees simulated from the URRT or PA models, in multiple settings. From Figures 5 and 6 we see that the Jordan estimator has the lowest risk, for all values of the trees sizes, and that the degree method is the second best one. In fact, it is not surprising that the degree method performs well for PA trees, since, in this model, the degree has a power law distribution and the order by degree correlates well with the arrival times of the vertices. However, this results is more surprising for the URRT model, where degree-centrality is known to be sub-optimal in the root-finding problem, see Bubeck et al. [6]. This is discussed further in Appendix A.3. Moreover, the spectral method has the poorest performance in both models. A possible explanation for this is that the random fluctuations of the adjacency matrix in the considered recursive tree models are large, leading to a large difference between the expected and empirical adjacency matrices in spectral norm. This absence of concentration, which is generally required in spectral ordering methods, could explain why this method poorly performs in our setting. Finally, the Reverse DMC algorithm performs similarly poorly as the spectral method in the PA model.

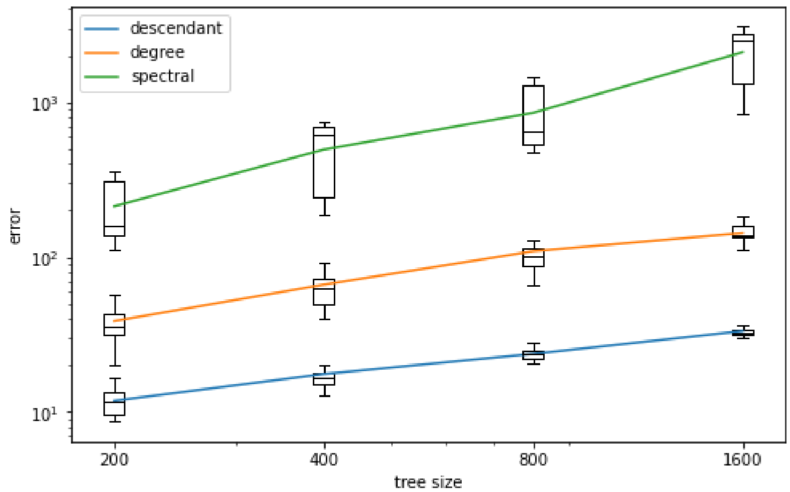

Figure 5: Risk R_α versus the tree size n in logarithmic scales, for $\alpha = 1.5$, and for trees simulated from the URRT model. Here, we sample 10 trees for each size. We compare the risk of descendant (blue), degree (orange), and spectral methods (green), and report a boxplot with the median, first, and last quartiles, for each tree size. In all settings, the descendant ordering largely outperforms the other methods.

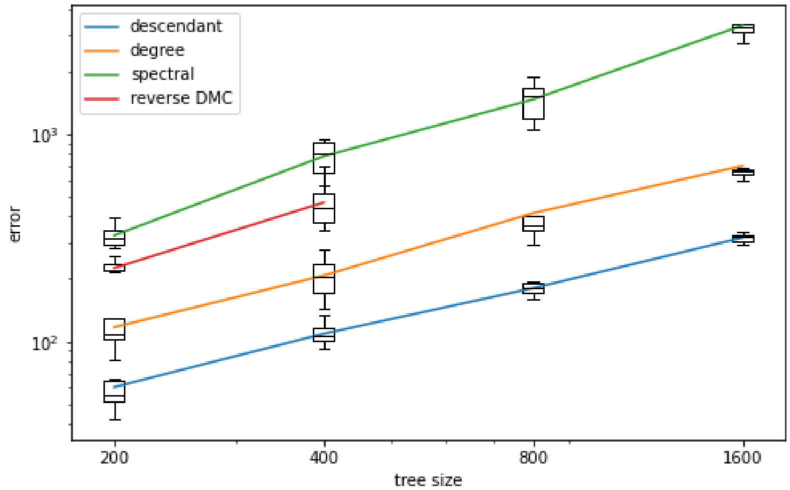

Figure 6: Risk R_α versus the tree size n in logarithmic scales, for $\alpha = 1.2$, and for trees simulated from the PA model. Here, we sample 10 trees for each size, only considering small trees for the reverse DMC method due to its high computational cost. We compare the risk of descendant (blue), degree (orange), spectral (green), and reverse DMC (red) methods, and report a boxplot with the median, first, and last quartiles, for each tree size. Just like in the URRT model, the descendant ordering outperforms the other methods.

A Appendix

In this appendix we discuss some issues concerning the choice of the parameter α in the definition of the risk and ordering according to degrees. Some elements of the proofs for the URRT model are also reported here. The final sections contain the proofs of all results in the PA model.

A.1 A remark on the choice of α

The risk R_α defined in (2) leads to a meaningful performance measure in the URRT and PA models only for some values of α. In particular, for $\alpha < 1$, it is easy to see that the risk of a random permutation is of the same order as the established lower bound, both in the URRT and PA models. More precisely, let Σ be a permutation chosen uniformly at random. Simple computation for $\alpha < 1$ leads to

$$R_\alpha(\Sigma) \le c_\alpha n^{2-\alpha} ,$$

for some positive constant c_α. On the other hand, Theorems 1 and 7 imply that for $\alpha < 1$, $R^*_\alpha \ge c'_\alpha n^{2-\alpha}$. Therefore, with $\alpha < 1$, the a random ordering has a risk of the same order as that of the optimal one. This is why we restrict our analysis of the risk to $\alpha \ge 1$. Our analysis of the Jordan ordering proves that this method has a risk of optimal order for $\alpha \in [1, 2)$ in the URRT case and $\alpha \in [1, 5/4)$ in the PA tree.

A.2 A remark on rumor centrality

We conjecture that in the URRT model, there exists an ordering procedure whose risk is of the order of $n^{2-\alpha}$ for any $\alpha \ge 1$, matching that of the minimax lower bound. Indeed, in our analysis, the risk is decomposed in two parts. First, a part coming from the difference between the Jordan and the descendant ordering (i.e, the error made by estimating the position of vertex 1 by the Jordan centre), second the risk of the descendant ordering. A possible way to improve our bound on the risk is to estimate the position of vertex 1 more precisely. To do so, using the rumor centrality appears to be a promising option. Indeed, due to recent results from Crane and Xu [10], in the URRT model, the rumor centrality orders vertices by their likelihood of being vertex 1. In particular, using the rumor centrality is optimal for minimizing the size of a conficence set containing the root, outperforming Jordan centrality (Bubeck et al. [6]). However, one step in the analysis is missing. Copying the proof of Lemma 3, with the rumor center instead of the Jordan center, one needs to bound the moment of order $1 + \gamma$ of the arrival time of the rumor centre. Bounding it by a constant (for any value of γ) would be sufficient to prove that this new ordering procedure has a risk of optimal order for any $\alpha \ge 1$.

A.3 A remark on ordering by degree

As discussed in Section 4, a simple ordering procedure is by the degrees. Simulations suggest that it does not perform as well as Jordan ordering, and it may produces non-recursive ordering. Nonetheless, it is a simple procedure worth mentioning. Since in a PA tree the degree of a given node follows a Pólya urn distribution, analysing the performance of the degree ordering is similar to the analysis carried out for Jordan centrality. However, the simulations results displayed in Figure 7 suggest that for any $\alpha \in [1, 5/4)$ the risk of the degree ordering grows at a faster rate than $n^{2-\alpha}$. Both in the URRT and PA model, for sizes of trees $\{1000, 2000, 4000, 8000\}$, we sample 10 trees at each size and compute the risk of the descendant and degree ordering for different values of α. Then, for each value of α, we perform a linear regression on the log-plot of the risk, to estimate the exponent of the polynomial.

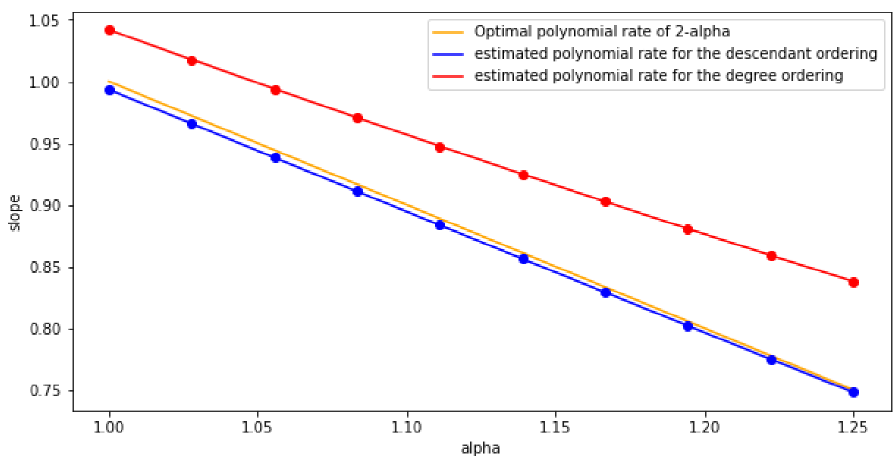

Figure 7: An estimation of the rate at which the risk increases with the size of the tree for different values of α in the PA tree. Here, we compare the case of the descendant and degree ordering. Proposition 7 shows that the optimal risk grows as $n^{2-\alpha}$, and that the the risk of descendant ordering grows at the same rate (see Proposition 8). This experiment confirms these results. For the degree ordering, this plot suggests that the risk grows faster than $n^{2-\alpha}$.

On the other hand, ordering vertices by their degree in the URRT is known to be suboptimal for finding the root, as there are many vertices with much higher degree (Eslava [13]). Simulation results displayed in Figure 8 suggest that, for most values of α, the risk of ordering by degree in the URRT model grows at a faster rate than $n^{2-\alpha}$ for any $\alpha \in (1, 2)$. On the other hand, observing Figure 8, it seems like, for $\alpha = 1$, the degree ordering may have a risk growing at the optimal rate of n.

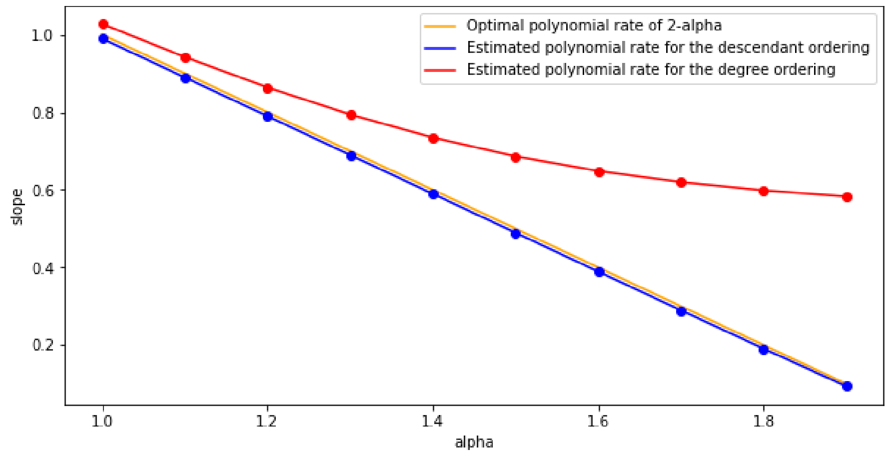

Figure 8: Estimation of the rate at which the risk increases with the size of the tree for different values of α in the URRT. Here we compare the case of descendant and degree ordering. Proposition 1 shows that the optimal risk grows as $n^{2-\alpha}$, and that the Jordan and descendant ordering's risks grow at the same rate (see Proposition 4). This experiment is in accordance with these results. For the degree ordering, this plot suggests that the risk grows at a rate faster than $n^{2-\alpha}$.

In Section 4, where the empirical performance of different ordering procedures are compared, the degree ordering is the method with second best performance. The above simulations suggest that all the other tested methods have risks growing at a faster rate than $n^{2-\alpha}$.

A.4 Proof of Theorem 4

Here, we present the arguments to complete the proof of Theorem 4. We recall that we need to upper bound $\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i)-i|}{i^\alpha}\right]$, and that we reduced in (12) the problem to upper bounding $\mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \geq \tau_{i,j}\right\} + \mathbb{P}\left\{\frac{\mathrm{de}(i)}{n} \leq \tau_{i,j}\right\}$, which is done in Lemma 6.

Proof of Lemma 6. From (13), we have

$$\mathbb{P}\{\mathrm{de}(j)=k\} = \frac{k!(j-1)(j)\cdots(n-k-2)}{j(j+1)\cdots(n-1)}\binom{n-j}{k}.$$

Re-arranging the factors,

$$\begin{aligned}\mathbb{P}\{\mathrm{de}(j)=k\} &= k!\frac{(n-k-2)!}{(j-2)!}\frac{(j-1)!}{(n-1)!}\frac{(n-j)!}{k!(n-j-k)!} \\ &= (j-1)\frac{(n-k-2)!}{(n-j-k)!}\frac{(n-j)!}{(n-1)!}\end{aligned}$$

Since $j \geq 3$,

$$\begin{aligned}
\mathbb{P}\{\mathrm{de}(j)=k\} &= (j-1)\frac{(n-j-k+1)\cdots(n-k-2)}{(n-j+1)\cdots(n-1)} \\
&= \frac{j-1}{n-1}\frac{n-j-k+1}{n-j+1}\cdots\frac{n-k-2}{n-2} \\
&\leq \frac{j-1}{n-1}\left(1-\frac{k}{n-2}\right)^{j-2} .
\end{aligned}$$

Therefore, for $n, j \geq 3$, we can upper bound the second term of (12) by

$$\begin{aligned}
\mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \geq \tau_{i,j}\right\} &\leq \sum_{(\tau_{i,j}n)\leq k\leq n} \frac{j-1}{n-1}\left(1-\frac{k}{n}\right)^{j-2} \\
&\leq \frac{(j-1)n}{n-1}\int_{\tau-\frac{1}{n}}^{1}(1-t)^{j-2}dt = \frac{n}{n-1}\left(1-\tau_{i,j}+\frac{1}{n}\right)^{j-1} \leq 2e^2e^{-j\tau_{i,j}},
\end{aligned}$$

using $\log(1+x) \leq x$. Moreover, for $i \geq 2$, we bound the third term of (12) by

$$\begin{aligned}
\mathbb{P}\left\{\frac{\mathrm{de}(i)}{n} \leq \tau_{i,j}\right\} &\leq \sum_{k\in[1,\tau_{i,j}n]} \frac{i-1}{n-1}\left(1-\frac{k}{n-2}\right)^{i-2} \\
&\leq \tau_{i,j}i\left(1-\frac{1}{n-2}\right)^{i-2} \leq \tau_{i,j}i .
\end{aligned}$$

Since for $i=1$, $\mathbb{P}\{\mathrm{de}(1)=k\} = 1/(n-1)$,

$$\mathbb{P}\left\{\frac{\mathrm{de}(1)}{n} \leq \tau_{i,j}\right\} \leq \sum_{k\in[1,\tau_{i,j}n]} \frac{1}{n-1} \leq \tau_{i,j} + \frac{1}{n-1} .$$

Therefore, for $i \geq 2$ and $j > i+1$, by choosing $\tau_{i,j} = \frac{1}{j}\log\frac{j}{i}$, we obtain that

$$\mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \geq \tau_{i,j}\right\} + \mathbb{P}\left\{\frac{\mathrm{de}(i)}{n} \leq \tau_{i,j}\right\} \leq 2e^2e^{-\log\frac{j}{i}} + \frac{i}{j}\log\frac{j}{i} \leq \frac{i}{j}\left(2e^2+\log\frac{j}{i}\right),$$

and for $i=1$, $j > i+1$, choosing $\tau_{1,j} = \frac{1}{j}\log(j)$ we obtain that

$$\mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \geq \tau_{1,j}\right\} + \mathbb{P}\left\{\frac{\mathrm{de}(1)}{n} \leq \tau_{1,j}\right\} \leq \frac{1}{j}\left(2e^2+\log(j)\right) + \frac{1}{n-1} .$$

□

Plugging the upper-bounds of Lemma 6 in (12), for $\alpha = 1$, we get

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i}\right] \leq n + \sum_{i=1}^{n} \mathbb{E}\left[\frac{\widehat{\sigma}'(i)}{i}\right] \\
&\leq n + \sum_{i=2}^{n} \frac{1}{i}\left(i + 1 + \sum_{j=i+2}^{n} \frac{i}{j}\left(2e^2 + \log\frac{j}{i}\right)\right) + 2 + \sum_{j=3}^{n}\left(\frac{1}{j}\left(2e^2 + \log(j)\right) + \frac{1}{n-1}\right) \\
&\leq 2n + 3 + \log(n) + \sum_{j=3}^{n} \frac{1}{j} \sum_{i=1}^{j-2}\left(2e^2 + \log\frac{j}{i}\right) . \qquad (14)
\end{aligned}$$

Since

$$\sum_{i=1}^{j-2} \log\frac{j}{i} \leq \log\left(\frac{j^j}{j!}\right) ,$$

and that the Stirling formula implies that $j! \geq (j/3)^j$,

$$\sum_{i=1}^{j-2} \log\frac{j}{i} \leq j\log(3) .$$

Plugging this in (14) yields

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i}\right] \leq 3 + \log(n) + (2 + 2e^2 + \log 3)n ,$$

which in turn proves that for $n \geq 60$,

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i}\right] \leq 18n .$$

For $1 < \alpha < 2$, a similar calculation yields

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i^\alpha}\right] &\leq \frac{1}{2-\alpha} n^{2-\alpha} + \sum_{i=1}^{n} \mathbb{E}\left[\frac{\widehat{\sigma}'(i)}{i^\alpha}\right] \\
&\leq \frac{1}{2-\alpha} n^{2-\alpha} + 1 + \sum_{i=1}^{n} \frac{1}{i^\alpha}\left(i + 1 \sum_{j=i+2}^{n} \frac{i}{j}\left(2e^2 + \log\frac{j}{i}\right)\right) \\
&\leq \frac{2}{2-\alpha} n^{2-\alpha} + 1 + \zeta(\alpha) + \sum_{j=3}^{n} \frac{1}{j} \sum_{i=1}^{j-2} \frac{2e^2}{i^{\alpha-1}} + \frac{1}{i^{\alpha-1}} \log\frac{j}{i} \\
&\leq \frac{2}{2-\alpha} n^{2-\alpha} + 1 + \zeta(\alpha) + \frac{2e^2}{(2-\alpha)^2} n^{2-\alpha} + \sum_{j=3}^{n} \frac{1}{j} \sum_{i=1}^{j-2} \frac{1}{i^{\alpha-1}} \log\frac{j}{i} .
\end{aligned}$$

Recall that ζ denotes the Riemann zeta function. We may upper bound

$$\sum_{i=1}^{j-2} \frac{1}{i^{\alpha-1}} \log \frac{j}{i} \le 2 \int_1^j \frac{1}{t^{\alpha-1}} \log\left(\frac{j}{t}\right) ,$$

which in turn can be evaluated by integration by parts, leading to

$$\sum_{i=1}^{j-2} \frac{1}{i^{\alpha-1}} \log \frac{j}{i} \le \frac{2}{(2-\alpha)^2} j^{2-\alpha} .$$

Finally

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i)-i|}{i^\alpha}\right] \le \left(\frac{2}{2-\alpha} + \frac{2e^2}{(2-\alpha)^2} + \frac{2}{(2-\alpha)^3}\right) n^{2-\alpha} .$$

For $\alpha \ge 2$, we similarly get

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i)-i|}{i^\alpha}\right] \le C ,$$

for some positive constant C.

A.5 Proof of the minimax lower bound in the PA model

Here we prove Theorem 7

Proof. The proof follows the same argument as that of Theorem 1. It suffices to check that the event

$$\Omega_j := \{\tau(j) \text{ and } \tau(\lfloor n/4 \rfloor + j) \text{ are leaves, connected to vertices of rank in } [n/2]\}$$

has a probability bounded away from 0. Proceeding as in the proof of Theorem 1, we get

$$\begin{aligned}
\mathbb{P}\left\{\Omega_j\right\} &= \frac{2(\lfloor n/2 \rfloor - 1)}{2(j-2)} \prod_{k=j+1}^{\lfloor n/4 \rfloor + j - 1} \frac{2k-3}{2(k-1)} \frac{2(\lfloor n/2 \rfloor - 1)}{2(\lfloor n/4 \rfloor + j - 2)} \prod_{k=\lfloor n/4 \rfloor + j + 1}^{n} \frac{2k-4}{2(k-1)} \\
&= \frac{(\lfloor n/2 \rfloor - 1)}{(j-2)} \frac{2j-1}{2(\lfloor n/4 \rfloor + j - 2)} \frac{(\lfloor n/2 \rfloor - 1)}{(\lfloor n/4 \rfloor + j - 2)} \frac{(\lfloor n/4 \rfloor + j - 1)(\lfloor n/4 \rfloor + j)}{(n-1)(n-2)} \\
&= \frac{(\lfloor n/2 \rfloor - 1)^2}{(n-1)(n-2)} \frac{2j-1}{2j-4} \frac{(\lfloor n/4 \rfloor + j - 1)(\lfloor n/4 \rfloor + j)}{(\lfloor n/4 \rfloor + j - 2)^2} \\
&\ge \frac{1}{4}\left(1 - \frac{5}{n}\right)^5 .
\end{aligned}$$

□

A.6 Descendant ordering in the PA model

Here, we analyze the descendant ordering in the PA model. Recall the notation introduced in Section 2.2: the centrality measure $\psi'(u) = n - \mathrm{de}(u)$, and the corresponding ordering $\widehat{\sigma}'$. In the next lemma we prove that ψ' and ψ coincide for most vertices and provide a control both on the number of vertices for which they differ and the estimated arrival time of vertex 1.

Lemma 10. *Let c be the rank of a Jordan's centroid, and let* $\{1 \to c\}$ *be the set of vertices on the path from the root to the centroid. Then*

1. $\forall v \in [n] \backslash \{1 \to c\}$, $\psi_T(v) = \psi'_T(v)$;
2. *there exists an universal constant K such that c is stochastically dominated by an exponential random variable with parameter K;*
3. *for any* $\epsilon > 0$, *with probability at least* $1 - \epsilon$

$$\widehat{\sigma_J}(1) \leq \frac{C}{\epsilon^2} \exp\left(\sqrt{C \log\left(\frac{1}{\epsilon}\right)}\right) .$$

Proof. The first part of the proof is identical to the proof of Lemma 2. First, we use Theorem 6 of Wagner and Durant [30], which extend the result of Moon [22] from uniform random recursive trees to preferential attachment trees. Using their result, we obtain that

$$\mathbb{P}\{c \geq k\} \leq \sum_{j=k}^{\infty} \frac{(-\log(2)/2)^j}{j!} ,$$

so there exists an exponential random variable of parameter K such that $c \leq \mathcal{E}(K)$. Using Corollary 3.3.b of Banerjee and Bhamidi [2], we have that the event

$$\widehat{\sigma_J}(1) \leq \frac{C}{\epsilon^2} \exp\left(\sqrt{C \log\left(\frac{1}{\epsilon}\right)}\right) ,$$

holds with probability at least $1 - \epsilon$. This conclude the proof of the lemma. $\square$

The next lemma allows us to compare the risk of Jordan and descendant ordering.

Lemma 11. *Let* $T \sim PA$. *Then, there exist positive constants* C, K, *such that, for* $\alpha > 0$

$$R_\alpha(\widehat{\sigma_J}) \leq R_\alpha(\widehat{\sigma}') + K \sum_{i=1}^{n} \frac{1}{i^\alpha} + C \log^2(n)\sqrt{n} .$$

Proof. The proof is similar to the one of Lemma 3. Recalling that D is the distance between vertices 1 and c, we have

$$\mathbb{E}\left[\sum_{i\in\{1\to c\}}\frac{\left|\widehat{\sigma_J}(i)-i\right|}{i^\alpha}\right]\le\mathbb{E}\left[\sum_{i\in\{1\to c\}}i\right]+\mathbb{E}\left[\sum_{i\in\{1\to c\}}\widehat{\sigma_J}(i)\right]$$

$$\le\frac{1}{2}\mathbb{E}\left[D^2\right]+\mathbb{E}\left[D\widehat{\sigma_J}(1)\right].$$

As in Lemma 3, we use the fact that $D\le c$ and the domination of c by an exponential random variable (see Lemma 10) to get that

$$\frac{1}{2}\mathbb{E}\left[D^2\right]\le K^2 .$$

Then, it follows from Hölder's inequality that

$$\mathbb{E}\left[D\widehat{\sigma_J}(1)\right]\le\left(\mathbb{E}\left[D^{\frac{1+\gamma}{\gamma}}\right]\right)^{\frac{\gamma}{1+\gamma}}\left(\mathbb{E}\left[\widehat{\sigma_J}(1)^{1+\gamma}\right]\right)^{\frac{1}{1+\gamma}} . \tag{15}$$

Using once again the domination of D by an exponential random variable,

$$\left(\mathbb{E}\left[D^{\frac{1+\gamma}{\gamma}}\right]\right)^{\frac{\gamma}{1+\gamma}}\le C\frac{1}{1+\gamma}\gamma ,$$

for some positive constant C. Next, using Lemma 10,

$$\mathbb{P}\left\{\widehat{\sigma_J}(1)\ge f(\epsilon)\right\}\le\epsilon ,$$

where $f(\epsilon)=\frac{C}{\epsilon^2}\exp\left(\sqrt{C\log\left(\frac{1}{\epsilon}\right)}\right)$. The function f is a non-increasing, therefore $f\left(\frac{C}{\sqrt{k}}\exp\left(\sqrt{C\log(k)}\right)\right)\le k$. So

$$\mathbb{P}\left\{\widehat{\sigma_J}(1)\ge k\right\}\le\frac{C}{\sqrt{k}}\exp\left(\sqrt{C\log(k)}\right) .$$

Following the same steps as in Lemma 3, and choosing $\gamma=1/\log(n)$, yields

$$\mathbb{E}\left[D\widehat{\sigma_J}(1)\right]\le C\log^2(n)\sqrt{n} ,$$

which concludes the proof of the lemma. □

A.7 Performance of Jordan ordering in the PA model

In this section we prove Theorem 8.

Proof. Similarly to the URRT case, in the PA model, the number of descendants of a vertex is distributed as a Pólya urn. This well-know fact is easily seen since in the PA model, sampling a vertex with a probability proportional to its degree is the same as sampling an edge uniformly at random and picking one of its endpoints at random. In turn, it is the same as picking a half edge uniformly at random. Therefore, the resulting Pólya urn has slightly different initial conditions than in the URRT. Such Pólya urns are well understood. In particular, by Mahmoud [21, Section 3.2], for a vertex $i \in [n]$, the distribution of $\mathrm{de}(i)$ is given by

$$\mathbb{P}\{\mathrm{de}(i) = k\} = \frac{(13\cdots(2k-1))((2i-3)(2i-1)\cdots(2n-2k-5))}{(2i-2)2i\cdots(2n-4)}\binom{n-i}{k} .$$

Re-arranging the terms in the above expression,

$$\begin{aligned}\mathbb{P}\{\mathrm{de}(i) = k\} \quad = \quad & \underbrace{\frac{1\cdot 3\cdots(2k-1)}{k!}}_{\stackrel{\text{def}}{=}A} \cdot \underbrace{\frac{(2i-3)(2i-1)\cdots(2n-5)}{(2i-2)2i\cdots(2n-4)}}_{\stackrel{\text{def}}{=}B} \\ & \cdot \underbrace{\frac{(n-i)!/(n-i-k)!}{(2n-2k-3)(2n-2k-1)\cdots(2n-5)}}_{\stackrel{\text{def}}{=}C} . \end{aligned} \tag{16}$$

We bound each term on the right-hand side of (16). First, for $k \geq 1$,

$$A = \frac{1}{k}\prod_{j=1}^{k-1}\frac{2j+1}{j} = \frac{2^{k-1}}{k}\prod_{j=1}^{k-1}\left(1+\frac{1}{2j}\right) .$$

Since

$$\prod_{j=1}^{k-1}\left(1+\frac{1}{2j}\right) = \exp\left(\sum_{j=1}^{k-1}\log\left(1+\frac{1}{2j}\right)\right) \leq \exp\left(\sum_{j=1}^{k-1}\frac{1}{2j}\right) \leq \sqrt{k} ,$$

then,

$$A \leq \frac{2^{k-1}}{\sqrt{k}} .$$

Second, we have

$$
\begin{aligned}
B &= \prod_{j=i}^{n-1} \frac{2j-3}{2j-2} = \prod_{j=i}^{n-1} \left(1 - \frac{1}{2j-2}\right) \\
&= \exp\left(\sum_{j=i}^{n-1} \log\left(1 - \frac{1}{2j-2}\right)\right) \le \exp\left(-\sum_{j=i}^{n-1} \frac{1}{2j-2}\right) \le 2\sqrt{\frac{i}{n}}
\end{aligned}
$$

Finally, we have that

$$
C = \prod_{j=n-k-2}^{n-3} \frac{j-i-3}{2j+1} = \frac{1}{2^{k-1}} \prod_{j=n-k-2}^{n-3} \left(1 - \frac{i+2.5}{j+0.5}\right) \le \frac{1}{2^{k-2}} \left(1 - \frac{k}{n}\right)^i .
$$

Plugging these bounds into (16), we get

$$
\mathbb{P}\{\mathrm{de}(i) = k\} \le 4\sqrt{\frac{i}{kn}} \left(1 - \frac{k}{n}\right)^i .
$$

Thus, for any $\tau > 0$,

$$
\mathbb{P}\left\{\frac{\mathrm{de}(i)}{n} \le \tau\right\} \le \sum_{k=1}^{n\tau} 4\sqrt{\frac{i}{kn}} \le 4\sqrt{i\tau} . \tag{17}
$$

Now, for $j \in [n]$, we have

$$
\begin{aligned}
\mathbb{P}\left\{\frac{\mathrm{de}(j)}{n} \ge \tau\right\} &\le \sum_{k=n\tau}^{n} 4\sqrt{\frac{j}{kn}} \left(1 - \frac{k}{n}\right)^j \le 4\sqrt{\frac{j}{\tau}} \sum_{k=n\tau}^{n} \frac{1}{n} \left(1 - \frac{k}{n}\right)^j \\
&\le \frac{4}{\sqrt{j\tau}} . \qquad (18)
\end{aligned}
$$

Combining (17) and (18) with $\tau = 1/\sqrt{ij}$,

$$
\mathbb{P}\{\mathrm{de}(i) \le \mathrm{de}(j)\} \le 8\left(\frac{i}{j}\right)^{1/4} .
$$

Following similar calculations as in Section 2.3,

$$
\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i^\alpha}\right] &\le \frac{1}{2-\alpha} n^{2-\alpha} \sum_{i=1}^{n} \frac{1}{i^\alpha} \left(i + \sum_{j=i+1}^{n} 8\left(\frac{i}{j}\right)^{1/4}\right) \\
&\le \frac{2}{2-\alpha} n^{2-\alpha} + 8 \sum_{j=1}^{n} \left(\frac{1}{j^{1/4}} \sum_{i=1}^{j-1} \frac{1}{i^{\alpha-1/4}}\right) .
\end{aligned}
$$

For $\alpha \in [1, 5/4)$ we obtain

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i^\alpha}\right] \leq \frac{2}{2-\alpha} n^{2-\alpha} + \frac{8}{\alpha - 5/4} \sum_{j=1}^{n} \frac{1}{j^{\alpha-1}}$$

$$\leq \left(\frac{2}{2-\alpha} + \frac{1}{(\alpha - 5/4)(\alpha - 2)}\right) n^{2-\alpha} ,$$

while for $\alpha \geq 5/4$

$$\sum_{i=1}^{n} \mathbb{E}\left[\frac{|\widehat{\sigma}'(i) - i|}{i^\alpha}\right] \leq \frac{2}{2-\alpha} n^{2-\alpha} + 8\,\zeta\left(\alpha - \frac{1}{4}\right) \sum_{j=1}^{n} \frac{1}{j^{1/4}}$$

$$\leq \frac{2}{2-\alpha} n^{2-\alpha} + \frac{32}{3} \zeta\left(\alpha - \frac{1}{4}\right) n^{3/4} .$$

which concludes the proof of Theorem 8. □

References

[1] Sayan Banerjee and Shankar Bhamidi. Persistence of hubs in growing random networks. *Probability Theory and Related Fields*, 180(3-4):891–953, 2021.

[2] Sayan Banerjee and Shankar Bhamidi. Root finding algorithms and persistence of Jordan centrality in growing random trees. *The Annals of Applied Probability*, 32(3): 2180–2210, 2022.

[3] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

[4] Anna M Brandenberger, Luc Devroye, and Marcel K Goh. Root estimation in Galton–Watson trees. *Random Structures & Algorithms*, 61(3):520–542, 2022.

[5] Simon Briend, Francisco Calvillo, and Gábor Lugosi. Archaeology of random recursive dags and Cooper-Frieze random networks. *Combinatorics, Probability and Computing*, pages 1–15, 2023.

[6] Sébastien Bubeck, Luc Devroye, and Gábor Lugosi. Finding Adam in random growing trees. *Random Structures & Algorithms*, 50(2):158–172, 2017.

[7] Sébastien Bubeck, Ronen Eldan, Elchanan Mossel, and Miklós Rácz. From trees to seeds: on the inference of the seed from large trees in the uniform attachment model. *Bernoulli*, 23(4A):2887–2916, 2017.

[8] George T Cantwell, Guillaume St-Onge, and Jean-Gabriel Young. Recovering the past states of growing trees. *arXiv preprint arXiv:1910.04788*, 2019.

[9] Alice Contat, Nicolas Curien, Perrine Lacroix, Etienne Lasalle, and Vincent Rivoirard. Eve, adam and the preferential attachment tree. *Probability Theory and Related Fields*, pages 1–16, 2024.

[10] Harry Crane and Min Xu. Inference on the history of a randomly growing tree. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(4):639–668, 2021.

[11] Harry Crane and Min Xu. Root and community inference on latent network growth processes using noisy attachment models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkad102, 09 2023. ISSN 1369-7412. doi: 10.1093/jrsssb/qkad102.

[12] Luc Devroye and Tommy Reddad. On the discovery of the seed in uniform attachment trees. *Internet Mathematics*, pages 75–93, 2019.

[13] Laura Eslava. Depth of vertices with high degree in random recursive trees. *ALEA*, 2022.

[14] Edward N Gilbert. Random plane networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543, 1961.

[15] Christophe Giraud, Yann Issartel, and Nicolas Verzelen. Localization in 1d non-parametric latent space models from pairwise affinities. *Electronic Journal of Statistics*, 17(1):1587–1662, 2023.

[16] John Haigh. The recovery of the root of a tree. *Journal of Applied Probability*, 7(1): 79–88, 1970.

[17] Jeannette Janssen and Aaron Smith. Reconstruction of line-embeddings of graphons. *Electronic Journal of Statistics*, 16(1):331 – 407, 2022. doi: 10.1214/21-EJS1940.

[18] Varun Jog and Po-Ling Loh. Analysis of centrality in sublinear preferential attachment trees via the crump-mode-jagers branching process. *IEEE Transactions on Network Science and Engineering*, 4(1):1–12, 2016.

[19] Varun Jog and Po-Ling Loh. Persistence of centrality in random growing trees. *Random Structures and Algorithms*, 52(1):136–157, 2018.

[20] Gábor Lugosi and Alan S. Pereira. Finding the seed of uniform attachment trees. *Electronic Journal of Probability*, 24:1–15, 2019.

[21] Hosam Mahmoud. *Pólya urn models*. Chapman and Hall/CRC, 2008. doi: 10.1201/9781420059847.

[22] John W Moon. On the centroid of recursive trees. *Australasian Journal of Combinatorics*, 25:211–220, 2002.

[23] Saket Navlakha and Carl Kingsford. Network archaeology: Uncovering ancient networks from present-day interactions. *PLOS Computational Biology*, 7(4):1–16, 04 2011. doi: 10.1371/journal.pcbi.1001119.

[24] Antoine Recanati, Thomas Brüls, and Alexandre d'Aspremont. A spectral algorithm for fast de novo layout of uncorrected long nanopore reads. *Bioinformatics*, 33(20): 3188–3194, 2017.

[25] Antoine Recanati, Thomas Kerdreux, and Alexandre d'Aspremont. Reconstructing latent orderings by spectral clustering. *arXiv preprint arXiv:1807.07122*, 2018.

[26] William S Robinson. A method for chronologically ordering archaeological deposits. *American antiquity*, 16(4):293–301, 1951.

[27] Devavrat Shah and Tauhid Zaman. Finding rumor sources on random trees. *Operations Research*, 64(3):736–755, 2016.

[28] Devavrat Shah and Tauhid R. Zaman. Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.

[29] Jithin Kazuthuveettil Sreedharan, Abram Magner, Ananth Y. Grama, and Wojtek Szpankowski. Inferring temporal information from a snapshot of a dynamic network. *Scientific Reports*, 9, 2019.

[30] Stephan Wagner and Kevin Durant. On the centroid of increasing trees. *Discrete Mathematics & Theoretical Computer Science*, 21, 2019.