

Received January 4, 2022, accepted February 7, 2022, date of publication February 17, 2022, date of current version March 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3152541

# Design and Analysis of Distributed Tree Growing Algorithms

MARCIO NUNES DE MIRANDA<sup>1</sup>, DANIEL RATTON FIGUEIREDO<sup>2</sup>, DANIEL FRAZÃO LUIZ<sup>3</sup>, CELSO BARBOSA CARVALHO<sup>4</sup>, AND DANIEL SADOC MENASCHE<sup>5</sup>, (Member, IEEE)

<sup>1</sup>Departamento de Computação, Instituto de Ciências Exatas BR, Universidade Federal Rural do Rio de Janeiro (UFRRJ), Seropédica 23897-000, Brazil

<sup>2</sup>COPPE/Programa de Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro 21941-972, Brazil

<sup>3</sup>Programa de Pós-Graduação em Engenharia Elétrica (PPGEE), Faculdade de Tecnologia, Universidade Federal do Amazonas (UFAM), Manaus 69067-005, Brazil

<sup>4</sup>Programa de Pós-Graduação em Engenharia Elétrica, Faculdade de Tecnologia e Centro de P&D em Tecnologia Eletrônica e da Informação (CETELI), Universidade Federal do Amazonas (UFAM), Manaus 69067-005, Brazil

<sup>5</sup>Departamento de Ciência da Computação, Instituto de Computação, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro 21941-590, Brazil

Corresponding author: Celso Barbosa Carvalho (ccarvalho\_@ufam.edu.br)

This work was supported in part by Samsung Electronics of Amazonia Ltd., according to Article 48 of Decree no 6.008/2006, under the terms of Federal Law no 8.387/1991, through Agreement 004, signed with the Center for Research and Development in Electronics and Information from the Federal University of Amazonas—Centro de P&D em Tecnologia Eletrônica da Informação (CETELI)/Universidade Federal do Amazonas (UFAM); and in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) under Grant E-26/203.215/2017, Grant E-26/211.144/2019, and Grant E-26/201.376/2021.

**ABSTRACT** Tree-based systems rely on real-time dissemination trees to deliver information to nodes. In order to offer good services, two fundamental aspects should guide the real-time growth process: low node degree and short distances to the server. In this paper, we propose a growth process to construct trees and make a detailed study on modeling and performance analysis of these tree-based systems. Our generative mechanism is based on the preferential attachment principle, where preference is given in terms of node quality. The proposed growth mechanism has a single parameter to weigh the relative importance of node degree and node distance on assessing node quality. We aim at understanding this mechanism when considering the local aspect of the node's degree and the global aspect of the distance to a source. With this goal, we investigate our model through simulations and compare it to other growth processes. Our results indicate that the proposed model is capable of self-organizing nodes into good trees under six metrics of interest.

**INDEX TERMS** Preferential attachment, real-time growth process, tree-based systems, node quality, recursive tree, random tree, random recursive tree, power of two choices.

## I. INTRODUCTION

Large-scale content distribution on the Internet has received much attention over the last two decades. The number of users of applications like Whatsapp, Facebook, Instagram, YouTube and Twitter has increased enormously and so the traffic at the network. Most of the difficulties arise from the large number of resources required (e.g., network bandwidth, memory) by the applications and by the traffic itself (for example, video traffic) when serving thousands (sometimes millions) of users. Moreover, most of these users are mobile, using devices that, usually, have restricted resources. In the near future, most the population in the world will have mobile connectivity and a substantial amount of devices and connections will be through 5G networks; this results in a

The associate editor coordinating the review of this manuscript and approving it for publication was Mingjun Dai<sup>1</sup>.

significant demand for high-scale real-time dissemination schemes. Other examples of applications that require new dissemination schemes or access to content appear in Industry 4.0, Sensor Networks, Internet of Things (IoT) and smart cities, where cyber-physical systems, smart objects or smart vehicles will be connected.

This huge demand for resources forces large-scale applications, such those involving content distribution, to scale more efficiently with the increasing number of users. An approach to achieve this goal, under delay and bandwidth constraints, is to construct a real-time dissemination tree, where the server corresponds to the root of the tree and the users to its internal nodes [1], [2]. As users arrive at the system, they can receive some information either directly from the server or from another node already in the tree. Thus, nodes receiving information also forward content to arriving nodes (from now on called newcomers).

A major challenge within this approach relates to the growth process of the real-time dissemination tree, as different trees will offer different qualities to users (or nodes). The main difficulty is that neither the server nor any other node has a global view of the quality of the tree, since the tree is constructed online. Even if all nodes that would form the tree were known in advance, arranging them in the best tree (the tree with the best quality) is not a simple task, e.g., as information about bandwidth and delay is usually not readily available. Thus, most tree-based systems rely on probing, side information and randomization in their growing processes.

Two issues are central when considering the quality of a tree:

- **Node degree.** The degree of a node in the tree (i.e., number of children) corresponds to the number of users being served by this node. Since nodes have finite resources (e.g., network bandwidth, memory and others) and each user served consumes resources, node degree directly affects the quality (e.g. *throughput*) of the information passed by a node.
- **Node distance.** The node distance is the number of hops between the node and the root of the tree (equivalently, it is the level of the node on the tree). Since information is forwarded from node to node down the tree, node distance directly affects the quality of the information received, as it is likely to experience larger *delays and losses*.

Thus, mechanisms to construct efficient trees usually consider these two characteristics. However, an important issue to assess the quality of a tree is the relative impact of these two characteristics. For example, if bandwidth is widely abundant, then node distance can affect the quality (e.g. video delay) relatively more than node degree.

In this paper, we are interested in understanding the quality and topological properties of online constructed trees when a simple mechanism is used. In particular, we model the growing process of the tree using a simple probabilistic process that considers node degree, node distance and a single parameter that captures the relative importance of these properties when assessing the node quality. Our generative model is based on the idea of “preferential attachment” [3], where preference is given to nodes with higher *utility*, which is a measure for the quality of service offered by a given node in the tree. To assess the quality of the proposed model, its topological properties are compared with the properties of the other five different models (two *offline* and three *online* mechanisms).

For the purpose mentioned above, we evaluate the models numerically through simulations and report the topological properties and quality of the trees constructed. Our findings indicate that our preferential attachment model generates relatively good quality trees when compared to *offline trees*, that are carefully organized (e.g., complete  $k$ -ary trees) and with other *online trees*, e.g., power of two choices (P2C) trees. Moreover, we find that the topological properties among the

online trees are not extremely different even when comparing opposite ends of the balance between node degree and node distance. Intuitively, the probabilistic approach captured by the model avoids trees with extreme topological structures, such as a star or a line tree.

Note that our goal is not to model any specific protocol, but to understand and characterize trees constructed through a simple, self-organizing mechanism based on the “preferential attachment” principle. We abstract all system-level details, such as bandwidth capacity and node location, and consider just the fundamental aspects that determine quality.

In summary, our key contributions are:

- **A configurable utility function.** The proposed node utility function balances the importance of the degree of each node and its distance to the root of the tree (server) through a single parameter ( $\alpha$ ). This allows the analysis of the topological properties of the generated trees (online and offline) for different networks through typical graph metrics, e.g., maximum and average node degrees, root degree, maximum and average node distances to the root (Section III).
- **New model for distributed tree design.** We propose a preferential attachment model which leverages the proposed utility function to assess the quality of the trees, given by the average quality of its nodes. For each node arrival, probabilistic preferential utility attachment (PPUA) relies on the quality of the nodes in the tree at the time of the arrival to determine the parent of the newcomer, in an online fashion (Sections III and IV).
- **Comparison of offline and online tree growing algorithms.** We numerically investigate how the quality of the proposed online algorithm compares against other online and offline algorithms, indicating that PPUA is competitive against counterparts (Section V). In addition, we show how previous models considered in the literature [4]–[6] can be framed as special cases of the tree growing algorithms studied in this work (Appendices B and C).

The remainder of this paper is organized as follows. In Section II we discuss related work. Section III describes the proposed model in detail and introduces the metrics of interest. Section IV describes the tree growing mechanisms to be investigated. Section V presents a numerical evaluation of different tree growing mechanisms, accounting for the graph metrics of interest, under the offline and online settings. Finally, Section VI concludes the paper with final remarks. Appendix A provides expressions for the quality of  $k$ -ary trees, Appendices B and C report specialized results and directions for future work accounting for cases where node quality is a function only of distance to the source and degree, respectively, and Appendix D briefly discusses optimal topologies.

## II. RELATED WORK

Within the last years, several solutions have been adopted to address large-scale information systems. Such solutions

can be classified roughly into tree-based systems or mesh-based systems [7], [8]. In tree-based systems, nodes are organized into a tree structure and information flows down one or more trees [7], [9], [10]. In mesh-based systems, there is no particular structure and nodes exchange information directly with one another, dynamically changing their neighbors over time, similar to epidemic dissemination [8], [11]. Magharei *et al.* [12] and Goh *et al.* [8] present a detailed comparative study between the tree and mesh-based systems. Usually, tree-based systems use push protocols, where nodes proactively send the information to other nodes in order to disseminate it through the system. On the other hand, mesh-based systems use pull protocols, where nodes explicitly request the missing information to other nodes. Cigno *et al.* [13] presents a hybrid push/pull protocol to take advantage of each topology.

### A. TREE-BASED SOLUTIONS

Small *et al.* [14] formulates a topology optimization problem as a minimization of server bandwidth cost, which leads to the scalability of the system concerning the number of nodes participating in the session. Liu *et al.* [15] derive performance bounds and present optimal tree-construction algorithms that service providers can use to provide scalable, node-assisted streaming services. Sayit *et al.* [16] propose a dynamic tree construction and maintenance method for streaming applications. Qiu *et al.* [17] proposes a tree-based self-organizing protocol for sensor networks, where nodes determine how to join the network based on a self-organizing process, using metrics such as the number of child nodes and communication distance.

Maccari *et al.* [18] propose a cross-layer optimization scheme to minimize the impact of the streaming overlay on the underlay wireless distributed network, exploiting information on the topology and routing of the underlay network. Telerius and Johansson [19] propose an algorithm that builds an overlay topology using graphs in which each node prefers to connect to nodes with a higher utility function value, aiming to improve streaming performance. The expected convergence time for the construction of the overlay topology is evaluated.

### B. TRAFFIC OFFLOADING AND HYBRID TREE-MESH SOLUTIONS

Some hybrid solutions involving tree models were proposed in the literature to deal with the scalability limitation of mesh-based models [20], [21]. Hasimoto-Beltran *et al.* [20] propose a hierarchical hybrid architecture that uses time-proximity for grouping the nodes in clusters following a hierarchical interconnected n-ary tree. The information is disseminated between neighbor clusters and, inside each cluster, in a top-bottom way. Budhkar and Tamara-palli [21] propose an overlay strategy for content delivery networks (CDN) using the serviceability of nodes to improve QoS and to reduce the load of CDN servers. The topology is arranged as tree-mesh based on serviceability. Peer-assisted

CDNs [22], [23] combine the benefits of traditional CDNs with the scalability properties of P2P networks in order to reduce the load on the CDN servers and provide lower latency. In such systems, a P2P content distribution network is constructed forming a tree topology, with the CDN server as the root, and used as backup by the peers.

Zhang and Hassanein [24] and AlTuhafi [25] present a survey of video streaming topologies. Huang *et al.* [26] applies a top-down construction of spanning trees to a secure message distribution in reliable communication networks. They propose an algorithm for constructing spanning trees using breadth-first search (BFS) to connect the nodes to the network. Hameed *et al.* [27] proposed a decision tree-based model that predicts the perceptual quality of the video transmitted over wireless networks using FEC algorithms to improve the QoE (Quality of Experience) of the transmitted video.

In another research thread, authors have studied the costs to maintain tree and mesh topologies. In [28], the authors aim at modeling and controlling dynamic networks through distributed consensual control. In [29] the authors propose a spanning tree coverage algorithm for distributed path planning of flying robots, where each robot constructs its spanning tree that grows towards uncovered areas.

The search for an optimal content dissemination topology may involve different sorts of heuristics. As an example, evolutionary algorithms can be used for clustering the nodes [30]. Alternatively, multi-objective optimization [31] can serve to account for delay and bandwidth minimization as conflicting goals. In this work, in contrast, we focus on simple real-time mechanisms, such as preferential attachment.

### C. GRAPH BASED ANALYSIS OF VIDEO STREAMING

Pandey *et al.* [32] implements and compares different approaches of content cache on the internet. The approaches are evaluated using different network topologies, cache size, content popularity, and a number of requests. For the implemented algorithms, the number of hops between the generating and consuming nodes is evaluated. This number is proportional to the delay. As we will soon describe, our proposed model can be adapted and applied to different scenarios of internet caching, despite the simplicity. We consider both the number of hops and the out-degree of nodes in the growing process of the tree.

A graph-based model to capture the dynamics of dissemination trees is discussed in [33]. The latter work considers a scenario close to the one we investigate in this paper. However, there are fundamental differences. We are concerned with understanding fundamental properties of a real-time dissemination tree constructed through a very simple mechanism based on preferential attachment, where all system-level parameters are abstracted. For example, differently from prior work, our model has no notion of bandwidth capacity or any explicit limits on the maximum number of children a node can have. These constraints are inherently captured by the

self-organizing nature of the proposed model, as we will soon discuss.

Since video transmission over wireless and wired networks demands high bandwidth and low delay, the objective function used in our work is suited for this kind of systems. In such scenarios, our model can be applied to reduce the transmission delay and balance the traffic and energy expenditure between network nodes. P2P and sensor networks can also be modeled as nodes arriving in a tree, like our model.

#### D. MATHEMATICAL MODELS AND PREFERENTIAL ATTACHMENT

Mathematical models have been used to understand the fundamental limits of information systems, as well as to understand design tradeoffs. Small *et al.* [34] presents a tradeoff study of system-level parameters in the scaling of peer-to-peer systems. Kumar *et al.* [35] and Carra *et al.* [36] propose a stochastic fluid-based modeling framework to evaluate the performance of general streaming systems. Bonald *et al.* [37] propose a model for epidemic-style dissemination to evaluate different dissemination strategies.

As an additional illustration of the applicability of epidemic models and its connection to growing trees, Li *et al.* [38] proposed a model for the spread of epidemics, in the framework of branching processes. Based on real data, they simulated the spread of an epidemic, studying the number of infected regions and the first arrival time of the contagion in each region. The paper shows that the first arrival time can be captured through a branching process [38], [39], that corresponds to a growing tree. Whereas the growing trees considered in branching processes account for multiple nodes being added to the tree as a result of infections caused by their corresponding parent node, in this work we account for individual nodes entering the system based on preferential attachment.

The principle of preferential attachment has been used to model the growth process of several different networks. In particular, preference by out-degree has been studied since the early 1990s [40], and was later on popularized by Barabasi and Albert [3], [41]. Then, authors generalized the idea to other preferences, including the notion of “fitness” of a node [42], as well as preference by nodes with old [43] or young ages [44].

Barabasi and Albert [3], [41] have applied the principle of preferential attachment to model networks that exhibit a power-law scaling behavior in their degree distribution. In their model, preference is proportional to node degree, such that nodes with a higher degree are more likely to receive edges from arriving nodes. They show that this simple growth process leads to graphs with a power-law degree distribution. Later, Karsai *et al.* [45] have discovered that this tree growth may be slow depending on the burstiness of nodes. This implies, for example, that infections by a computer virus can be reported years after its emergence or introduction.

Motivated by biological networks, Sevim and Rikvold [6] have also investigated a model where preference is inversely proportional to node out-degree. In their work, arriving nodes prefer to connect to nodes with lower out-degree. Sevim and Rikvold [6] show that their preferential attachment model entails a degree distribution which has tail probabilities that decay faster than an exponential.

*Remark 1:* In trees where nodes connect to parents uniformly at random, degree distribution has exponentially decreasing tail probabilities, i.e., the proportion of nodes with out-degree  $d$  converges, as  $n \rightarrow \infty$ , to  $2^{-(d+1)}$  almost surely [4], [46]. In trees where nodes connect to parents with preference towards nodes with larger degree, degree distribution follows a power-law [41]. Finally, if nodes connect to parents with preference towards nodes with smaller degree, degree distribution decays faster than an exponential [6].

In [47] the authors extended [6] and proposed a more general model, where preference is proportional to a positive power of the ratio of in-degree to out-degree. Although similar, the model we investigate in this paper is inherently different, as preference is inversely proportional to both node degree and node distance. As can be seen in the next section, our proposed model can degenerate to the model investigated by Sevim and Rikvold [6], when  $\alpha = 1$  (see also Appendix C).

The so called “power of choice” has also been considered in the realm of tree growth. The power of two choices, for instance, corresponds to sampling two nodes uniformly at random from the tree after each arrival, and selecting one among them as the parent of the newcomer. The degree and height distribution of a certain class of networks where nodes select their parents from a subset of nodes sampled uniformly at random have been studied in [5] and [48], respectively. In this work, we extend those studies to account for the joint role of heights and degrees in the choice of parents.

### III. TREE CONSTRUCTION MODEL

We consider a system composed of a single server and the sequential arrival of homogeneous nodes. Once it arrives, a node connects to a single node in the system to start receiving service. Nodes in the system offer service to a newcomer by simply forwarding it content. Note that the topology of the system is a tree, since every newcomer connects to a single node in the system. We assume that nodes always forward the content if they are chosen to be the parent of a newcomer (i.e., all nodes are altruistic). Finally, we also assume that nodes never leave the system nor move in the tree, thus, their position in the tree is determined at the time of their arrival. Figure 1 illustrates the construction of a tree.

A fundamental problem in this model is determining the parent node for an arriving node, addressed by a tree growing mechanism (algorithm). This mechanism is inherently an *online algorithm*, as it has no knowledge of the number of nodes that will join the tree. For example, this assumption is rather realistic when considering distributed, large-scale video streaming systems.

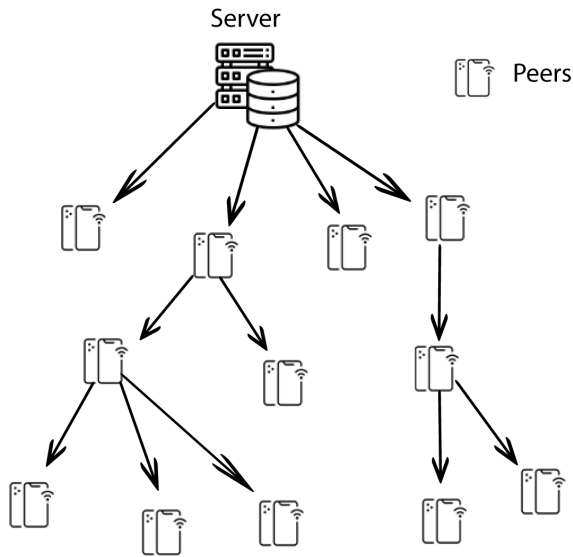


FIGURE 1. The real-time dissemination tree.

**A. OFFERED NODE QUALITY AND UTILITY FUNCTION**

An important consideration is the quality of service that a node on the tree can offer to a newcomer. Intuitively, the offered node quality is inversely proportional to its degree, as the finite resources must be shared among all its children. Moreover, offered node quality is also inversely proportional to the distance on the tree to the source (or server), as network characteristics that negatively affect the quality (e.g., delay and losses) are proportional to distances. Thus, the offered node quality degrades as either out-degree or distance increases. This motivates the use of the following utility function  $u(d, l)$  to assess offered node quality:

$$u(d, l) = \frac{1}{\alpha(d+1) + (1-\alpha)(l+1)}, \quad \text{for } 0 \leq \alpha \leq 1, \quad (1)$$

where  $d$  and  $l$  correspond to the out-degree of the node and its distance to the server (measured in hops), respectively.<sup>1</sup> The parameter  $\alpha$  is used to weigh the relative importance of the two properties  $d$  and  $l$ .

Note that when  $\alpha = 0$  offered node quality depends only on distance. This represents a system where (server or node) bandwidth is rather unlimited (i.e., nodes have virtually infinite bandwidth). On the other extreme, when  $\alpha = 1$  offered node quality depends only on degree. This represents a system that has severe bandwidth limitations. The extreme scenarios where  $\alpha = 0$  and  $\alpha = 1$  are studied in details in Appendices B and C, respectively.

Intuitively,  $\alpha$  is a parameter that determines the kind of system being considered. Of course,  $\alpha$  will have a fundamental influence when assessing node and tree quality.

Equation (1) will be used to determine the parent node of a newcomer, i.e., the node from which the arriving node will receive service. In particular, we consider a probabilistic

<sup>1</sup>From now on every time we write “node degree” we will be referring to “node out-degree”.

approach, using the idea of “preferential attachment”. Thus, a newcomer node randomly connects to the tree using a probability that is proportional to the utility it will receive from its parent node. Let  $p_v$  denote the probability that a newcomer chooses a parent node  $v$  already in the tree, where  $v$  can also be the server. Under probabilistic preferential utility attachment (PPUA),  $p_v$  is given as follows:

$$p_v = \frac{u(d_v, l_v)}{\sum_{s \in S} u(d_s, l_s)}, \quad (2)$$

where  $d_v$  and  $l_v$  correspond to the degree of node  $v$  and the distance between node  $v$  and the server, and  $S$  is the set of nodes already in the tree, including the server, at the time a new node arrives. Note that  $p_v$  varies with the number of nodes in the tree.

The above mechanism models distributed or centralized algorithms that perform an informed guess when determining the parent node for a newcomer. For example, in a centralized approach the server could provide the newcomer with this informed guess. In particular, the server can maintain and update the tree information for every node arrival, providing the newcomer with its randomly chosen parent node. Note that this random choice is much more efficient (in terms of computational cost) than determining the optimal parent in the tree for every newcomer (i.e., a node that would yield the highest utility). In what follows and for the purpose of comparison, we also consider a mechanism that computes for each newcomer its optimal parent in the tree, and refer to such an algorithm as deterministic preferential utility attachment (DPUA).

Last, the above model assumes there are no concurrent arrivals and that the system can update itself between consecutive arrivals. While this is not unrealistic given the computational requirements of processing an incoming node, we conjecture that as the number of nodes in the tree grows this assumption can be lifted with negligible consequences. In particular, the results concerning tree structure and node quality are likely to hold even if the system is updated only after a (small) batch of arrivals. However, a detailed analysis of this scenario is marginal to the current work and left for future investigation.

**B. TREE QUALITY AND OTHER METRICS OF INTEREST**

In order to characterize the topological properties of the trees constructed by our proposed model, from now on referred to as probabilistic preferential utility attachment trees (PPUA trees), we will use traditional graph-theoretical metrics, such as maximum and average node degree, server degree, node degree distribution, maximum and average node distance, node distance distribution and tree quality. These metrics will also be used to characterize the topological properties of many comparison trees. Table 1 summarizes these metrics and their acronyms.

To assess the quality of the tree constructed by the mechanism we consider a metric to capture the average quality of a node in the tree. Let  $q_v$  denote the quality received From

TABLE 1. Metrics to analyze topological properties.

Metric	Acronym
Maximum node degree	GMAX
Average node degree	GAVG
Server degree	GSER
Maximum node distance	DMAX
Average node distance	DAVG
Average tree quality	QAVG

now on every time we write “quality of a node” we will be referring to the “quality received by a node”. by node  $v$ . Thus, similarly to Equation (1), we have:

$$q_v = \frac{1}{\alpha d_{\mathcal{P}(v)} + (1 - \alpha)l_v}, \quad \text{for } 0 \leq \alpha \leq 1. \quad (3)$$

where  $\mathcal{P}(v)$  denotes the parent of node  $v$ ,

$$q_v = u(d_{\mathcal{P}(v)} - 1, l_v - 1). \quad (4)$$

whereas the values of parameters of Equation (1) are assigned before the newcomer joins the network, the parameters of Equation (3) are assigned after the newcomer joins. In particular, note that the quality received by newcomer  $v$  depends on the degree of its parent,  $d_{\mathcal{P}(v)}$ , whose level equals the level of  $v$ ,  $l_v$ , minus one. Moreover,  $\alpha$  is identical in Equations (1) and (3), as utility and quality must be assessed using the same relative importance between node degree and node distance, as indicated by Equation (4).

Using Equation (3), we define the tree quality  $\bar{q}$  as the average node quality of a given tree as follows:

$$\bar{q} = \frac{\sum_{s \in S \setminus \{r\}} q_s}{|S| - 1}, \quad (5)$$

where  $S$  is the set of nodes in the tree, including the root node,  $|S|$  the cardinality of this set, and  $r$  is the root node (server).

Table 2 contains a summary of the symbols used to define variables and parameters used throughout the remainder of this paper, along with their meanings.

TABLE 2. Table of notation.

Symbol	Meaning
$u$	Utility function used to assess the offered node quality
$d$	(Out-)degree of a node
$l$	Node distance, in hops, to the root (server); level
$\alpha$	Parameter used to weigh the relative importance of $d$ and $l$
$p_v$	Probability that a new node chooses a parent node $v$
$d_v$	Degree of a node $v$
$l_v$	Distance from node $v$ to the root; level of $v$
$S$	The set of nodes in the tree (including the root)
$s$	Node belonging to the set $S$ of nodes in the tree
$r$	Root of the tree (server)
$q_v$	Quality received by a node $v$ after joining the tree
$\mathcal{P}(v)$	Parent of a node $v$
$l_{\mathcal{P}(v)}$	Distance from the parent of node $v$ to the server
$d_{\mathcal{P}(v)}$	Degree of the parent of node $v$
$\bar{q}$	Average node quality of a given tree
$q_s$	Quality of a node $s$
$ S $	Cardinality of the set $S$

#### IV. COMPARING TREE FORMATION ALGORITHMS

We compare the behavior and quality of the trees generated by the PPUA model against other five tree-formation algorithms: two *offline algorithms* and three *online algorithms*. The offline algorithms receive as input the total number of nodes joining the tree. They output a final tree indicating how nodes should be organized. In particular, one mechanism considers the ensemble of best complete  $k$ -ary trees (K-C tree), while the other attempts to construct a uniform quality tree (UNIFQ tree). The other three algorithms build trees online. The trees are generated by simulating a growth process, as the tree generated by the proposed model (PPUA tree): uniform at random tree (URND tree), deterministic preferential utility attachment tree (DPUA tree) and power of two choices tree (P2C tree). Below we detail how these trees are constructed.

##### A. OFFLINE TREES

###### 1) COMPLETE $k$ -ARY TREES

Recall that a complete  $k$ -ary tree is a tree where all nodes, except leaves and parents of leaves, have exactly  $k$  children. Intuitively, complete  $k$ -ary trees should yield a good tree quality, under the metric defined above, as they can tradeoff distances and degree by varying  $k$ . In particular, note that if  $k = 1$  we have a line tree, and if  $k = |S| - 1$  (where  $|S| = n$  is the number of nodes in the tree) we have a star tree with the root connected to  $|S| - 1$  leaves.

Consider a complete  $k$ -ary tree with  $n$  nodes, for a given value of  $k$ . The complete tree is said to be full if at level  $l$  it has exactly  $k^l$  nodes, for  $l = 0, \dots, h$  where  $h$  is the tree height. The height  $h$  of the  $k$ -ary tree is given by:

$$h = \lceil \log_k(-(-1 - k)n + 1) - 1 \rceil. \quad (6)$$

Let  $q_{k,n}^{(c)}$  denote the average node quality of a K-C tree, as given by Equation (5). In what follows, we drop subscript  $n$  whenever it is clear from context, and let  $q^{(c)}$  denote the best average node quality over all possible  $k$ -ary trees with  $n$  nodes (best K-C tree). Thus, we have:

$$q^{(c)} = \max q_k^{(c)}, \quad \text{for } 1 \leq k \leq |S| - 1, \quad (7)$$

where an expression for  $q_k^{(c)}$  is derived in Appendix A.

Note that  $q_k^{(c)}$  depends on  $\alpha$ , as  $\alpha$  determines the relative importance between node degree and node distance. Figure 2 shows the average tree quality of complete  $k$ -ary trees with 60,000 nodes as a function of  $k$ . Each curve corresponds to an  $\alpha$  value.

Figure 2 shows that when  $\alpha = 0$  an average tree quality of  $q_k^{(c)} = 1$  is achieved by the star tree with  $k = |S| - 1 = 59,999$  wherein all distances between the nodes and the server are equal to one. When  $\alpha = 1$ , an average quality of  $q_k^{(c)} = 1$  is achievable with  $k = 1$ , that leads to a line tree where all node degrees are equal to one, except for a single leaf. With  $\alpha = 0.5$ , we observe that the best tree is obtained when  $k = 4$ , a relatively small value compared to  $|S|$ .

In what follows, we further discuss the role of  $\alpha$  while determining the best value of  $k$ . In general, we will be

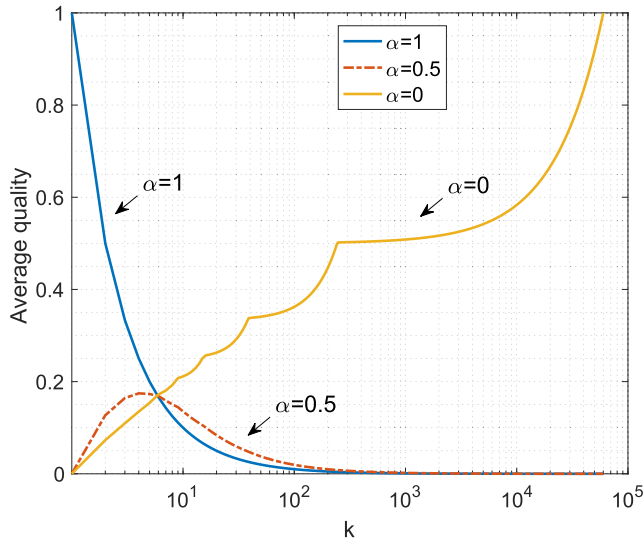


FIGURE 2. Average tree quality for complete  $k$ -ary trees with different  $\alpha$  values ( $n = 60,000$ ).

interested in determining the optimal value of  $k$  for each  $\alpha$ , i.e., to find the best K-C tree for any given  $\alpha$ , and to explain the interplay between  $\alpha$  and the optimal  $k$ .

2) OPTIMAL  $k$  IN COMPLETE  $k$ -ARY TREE

Next, we consider an approximation to determine the optimal value of  $k$  in K-C trees as a function of  $\alpha$ . Given  $\alpha$  and  $n$ , our goal is to determine how to set  $k$  to maximize average node quality. The approximation consists in focusing on leaf nodes. Indeed, we consider a tree where all nodes except leaves have degree  $k$ . Then, leaf nodes are at distance  $\log_k(n - 1)$  from the server, and their corresponding quality is given by:

$$q = \left( (1 - \alpha) \left( \frac{\ln(n - 1)}{\ln k} \right) + \alpha k \right)^{-1} \tag{8}$$

Taking the derivative of the quality of leaves given by Equation (8), with respect to  $k$ , and setting it to zero, we obtain Equation (9),

$$\frac{dq}{dk} = - \frac{\alpha + \frac{(\log_k(n-1))(\alpha-1)}{k(\ln k)}}{\left( (1 - \alpha) \left( \frac{\ln(n-1)}{\ln k} \right) + \alpha k \right)^2} = 0. \tag{9}$$

Note that for  $\alpha = 0$  (resp.,  $\alpha = 1$ ) we have  $\frac{dq}{dk} > 0$  (resp.,  $\frac{dq}{dk} < 0$ ), which corresponds to the fact that the utility monotonically increases (resp., decreases) with respect to  $k$  in these two extreme cases. Indeed, the star and line topologies are optimal in those two extremes, respectively (see Section III-A). For  $\alpha$  between 0 and 1, the value of  $k$  which maximizes  $q$  satisfies:

$$\frac{\log_k(n - 1)}{k(\ln k)} = \frac{\alpha}{1 - \alpha}. \tag{10}$$

Therefore,

$$k = e^{2W(\sqrt{\alpha}/2)} \tag{11}$$

where

$$\kappa = \left( \frac{1}{\alpha} - 1 \right) \ln(n - 1) \tag{12}$$

and  $W(x)$  is the principal branch of the Lambert function, i.e.,  $W(x)$  is the solution of  $we^w = x$ . For  $n = 60,000$  and  $\alpha = 0.5$ , the above approximation implies that the optimal value of  $k$  is between 4 and 5, which is in agreement with Figure 2, obtained empirically.

In Figure 3 we plot the floor of the solution of Equation (10), for  $n = 600$  and  $n = 60,000$ . As  $\alpha$  increases from 0.1 to 1.0, the optimal degree decreases from 14 to 1, corresponding to a transition from graphs favoring smaller distances to smaller degrees. Note also that the optimal degree either increases or remains unchanged as the population size grow from 600 to 60,000. This is because larger populations may require higher node degrees to maintain the distance from nodes to the server at optimal levels.

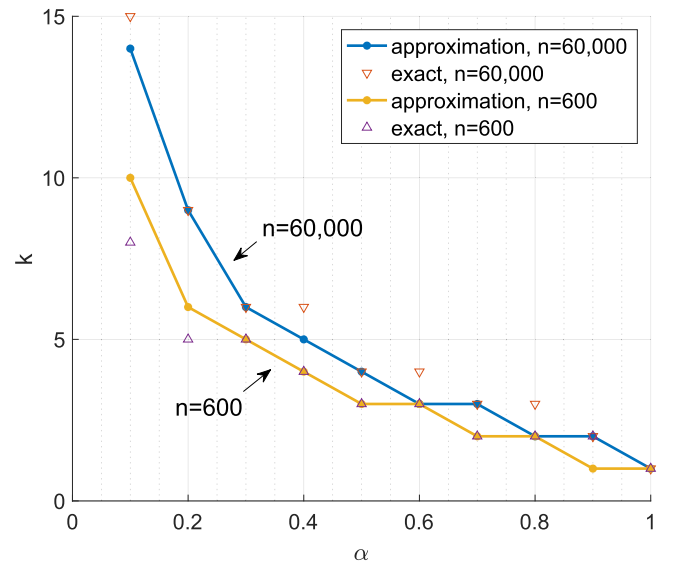


FIGURE 3. Approximate optimal degree ( $\lfloor k \rfloor$  which satisfies Equation (10)) and exact optimal degree of K-C trees for  $n = 600$  and  $n = 60,000$ .

The smaller the value of  $\alpha$ , the larger the optimal degree. In particular, for  $\alpha = 0$  (not shown in Figure 3) the optimal degree equals  $n - 1$ . As  $\alpha$  increases from 0 to 0.1, the optimal degree sharply decreases to values ranging between 10 and 14 for  $n = 600$  and 60,000, respectively. Finally, for  $\alpha = 1$  the optimal degree equals 1 independently of  $n$ . From now on, every time we say “K-C tree” we are referring to the best K-C tree for a given  $\alpha$ .

3) UNIFORM QUALITY TREE

The complete  $k$ -ary tree has the drawback that all nodes in the tree, except for leaves and parents of leaves, must have the same degree. This may not be optimal, as nodes at larger distances (further down the tree) will experience lower quality. However, this reduction of quality could be

compensated by smaller degrees. This is exactly the intuition behind the uniform quality tree (UNIFQ tree).

Consider a system with  $n = |S|$  nodes. In the UNIFQ tree, we assume all nodes should have roughly the same quality. Since all nodes have similar quality, all nodes in a given level of the tree, except leaves, must have parents with the same degree. Thus, all nodes at the same level of the tree have the same degree, except for parents of leaves. Let  $d_i$  denote the degree of nodes in level  $i$ , where  $i = 0, 1, 2, \dots$ . Note that  $d_0$  is the degree of the server. Thus, as given by Equation (3), the quality of a node in level  $i$  of a UNIFQ tree,  $q_i^{(u)}$ , is given by:

$$q_i^{(u)} = \frac{1}{\alpha d_{i-1} + (1 - \alpha)i}, \quad \text{for } 0 \leq \alpha \leq 1. \quad (13)$$

Whenever it is feasible to have  $q_i^{(u)}$  identical for all  $i$ , we have  $q_i^{(u)} = q_{i+1}^{(u)}$ . Thus, we find:  $d_i = d_{i-1} - (1 - \alpha)/\alpha$ . For  $i = 1$ , we have  $d_1 = d_0 - (1 - \alpha)/\alpha$ . By induction, we solve for  $d_i$ :

$$d_i = d_0 - \left(\frac{1}{\alpha} - 1\right)i, \quad \text{for } 0 < \alpha \leq 1 \quad (14)$$

where  $d_0$  is the degree of the server which is necessarily greater than 0.

In general, the degree sequence described in Equation (14) may yield non-integer degrees. Therefore, we relax the assumption that  $q_i^{(u)}$  must be identical for all  $i$ , and let  $d_i$  be given as follows:

$$d_i = \left\lfloor d_0 - \left(\frac{1}{\alpha} - 1\right)i \right\rfloor, \quad \text{for } 0 < \alpha \leq 1 \quad (15)$$

The server degree  $d_0$  determines the degree sequence of the tree, as given by Equation (15). The degree sequence  $d_0, d_1, d_2, \dots, d_{n-2}$  must be such that it can form a tree that can hold  $n$  nodes. Let  $m_i$  denote the number of nodes in level  $i$ , with  $i = 0, 1, 2, \dots, n - 1$ . In particular, the highest level  $n - 1$  is reached only by the line tree. We have that:

$$m_i \leq \prod_{j=0}^{i-1} d_j \quad i = 0, 1, 2, \dots, n - 1 \quad (16)$$

where  $m_0 = \prod_{j=0}^{-1} d_j = 1$ . Since the tree must hold all  $n$  nodes, we have the following additional condition:

$$\sum_{i=0}^h m_i \geq n \quad (17)$$

where  $h$  is the tree height. In particular, if the tree is a star we have  $h = 1$  and if the tree is a line we have  $h = n - 1$ .

For a given  $\alpha$  and target tree height  $h$ , the corresponding UNIFQ tree is obtained by determining the value of  $d_0$  satisfying Equation (17). Once  $d_0$  is determined, the whole degree sequence follows from Equation (15).

The optimal tree satisfying the above conditions is obtained through the solution of the following optimization problem, given  $0 \leq \alpha \leq 1$ :

$$\max_{d_0} \bar{q}(d_0, h(d_0)) \quad (18)$$

subject to:

$$d_i = \max \left( \left\lfloor d_0 - \left(\frac{1}{\alpha} - 1\right)i \right\rfloor, 0 \right), \quad (19)$$

$$i = 1, \dots, h - 1 \quad (20)$$

$$\sum_{i=0}^h \prod_{j=0}^{i-1} d_j \geq n \quad (21)$$

$$h \geq 1, \quad 0 \leq d_i \leq n - 1 \quad (22)$$

The above optimization problem is solved by iterating over  $d_0$  and determining the corresponding tree height  $h$  for each  $d_0$ . The optimal tree corresponds to parameters  $(d_0^*, h^*)$  maximizing average quality.

Whenever is feasible to build a UNIFQ tree wherein all node qualities are identical we have,

$$\bar{q} = q_v = q_i^{(u)} = \frac{1}{\alpha d_0 + 1 - \alpha}. \quad (23)$$

In this case, under the best UNIFQ tree  $d_0$  must be minimized. Motivated by this observation, in the remainder of this paper we consider UNIFQ trees wherein  $d_0$  must be minimized, which implies that the UNIFQ trees considered in this work are obtained solving the following optimization problem:

$$\min d_0 \quad (24)$$

subject to:

$$d_i = \max \left( \left\lfloor d_0 - \left(\frac{1}{\alpha} - 1\right)i \right\rfloor, 0 \right), \quad (25)$$

$$i = 1, \dots, h - 1 \quad (26)$$

$$\sum_{i=0}^h \prod_{j=0}^{i-1} d_j \geq n \quad (27)$$

$$h \geq 1, \quad 0 \leq d_i \leq n - 1 \quad (28)$$

For any given  $\alpha$  and  $n$ , the UNIFQ tree considered in this work corresponds to parameters  $(d_0^*, h^*)$  which solve the above optimization problem.

## B. ONLINE TREES

### 1) PROBABILISTIC PREFERENTIAL UTILITY ATTACHMENT TREE

Probabilistic preferential utility attachment tree (PPUA tree), as already stressed in the text, is obtained by a growth process where a node randomly connects to the tree based on a probability proportional to the utility it will receive from its parent node.

### 2) UNIFORM AT RANDOM TREE

Uniform tree (URND tree), also known as random recursive trees [4], is obtained by a growth process where the parent of a newcomer is selected uniformly at random among the nodes already present in the tree. The topology of this tree does not depend on the value of the parameter  $\alpha$ .



3) DETERMINISTIC PREFERENTIAL UTILITY ATTACHMENT TREE

Deterministic preferential utility attachment tree (DPUA tree) is obtained by a growth process where we compute the utility function (Equation (1)) for each node already in the tree. The node with the highest utility will be chosen as the parent of the arriving node. If we have a tie at the highest value, the tie is broken uniformly at random.

4) POWER OF TWO CHOICES TREE

Power of two choices tree (P2C tree) is obtained by a growth process that works as follows: first, we randomly select two nodes. Between these two nodes, the newcomer chooses the one with the highest utility. If we have a tie, we randomly select one of the nodes to be the parent of the newcomer.

Table 3 summarizes the tree formation policies and Table 4 contains a summary of the symbols used to define those policies.

TABLE 3. Tree formation policies.

	Policy	Description
offline	Best Complete $k$ -ary (K-C)	obtained through Equation (10) for a fixed $\alpha$
	Uniform quality (UNIFQ)	obtained iterating Equation (14) over $d_0$
online	Probabilistic preferential utility attachment (PPUA)	connects to $v$ with probability $p_v = \frac{u(d_v, l_v)}{\sum_{s \in S} u(d_s, l_s)}$
	Uniform at random (URND)	connects to $v$ chosen uniformly at random
	Determinist preferential utility attachment (DPUA)	connects to the $v$ with the greatest $u(d_v, l_v)$
	Power of two choices (P2C)	connects to vertex $v$ with the greatest $u(d_v, l_v)$ between two $v$ chosen uniformly at random

TABLE 4. Symbols used to define tree formation policies.

Symbol	Meaning
$n$	Number of nodes of the tree (including root)
$h$	Height of the tree
K-C tree	
$k$	Number of children nodes of non-leaf nodes of a K-C Tree
$q_k^{(c)}$	Average quality of a K-C tree with $k$ children per node
$q^{(c)}$	Average quality of best K-C tree
$W(x)$	Principal branch of the Lambert function
UNIFQ tree	
$m_i$	Number of nodes in level $i$
$d_i$	Degree of nodes in level $i$
$q_i^{(u)}$	Quality of a node in level $i$ of a UNIFQ tree

V. RESULTS

Next, we investigate the performance and topological features of the considered tree growing mechanisms. In particular,

we compare the online and offline solutions presented in the previous section. The simulation starts out with the server (i.e., root node). Nodes are sequentially added to the tree following the corresponding growth processes, described in Section IV, noting that PPUA, DPUA and P2C leverage Equation (2) while determining the parent of the new node to be added. After each node is added to the tree, the utilities and corresponding attachment probabilities of all nodes in the tree are recomputed.

*Simulation Setup:* For the results that follow, the simulation stops after exactly  $|S| = n = 60,000$  nodes are added to the tree (including the root).<sup>2</sup> Each simulation scenario is executed 20 times for each kind of tree and each  $\alpha$  value. We report the sample average of each metric with its confidence interval. Finally, we are interested in the behavior and performance as a function of its sole parameter  $\alpha$ , which determines the relative importance between node degree and distances when assessing the quality of nodes. Some curves will not show the value of the corresponding metric of interest for the extreme value of  $\alpha = 0.0$  (or for  $\alpha = 1.0$ ) in order to present more clearly the differences between the values of that metric for the other values of  $\alpha$  being considered; indeed, the values of some of the metrics of interest are outliers (too high) in those extremes, e.g., under DPUA, corresponding to the degree of a star with  $n - 1$  leaves and to the maximum distance in a line with  $n$  levels when  $\alpha = 0$  and  $\alpha = 1$ , respectively.

A. NODE DEGREE

We start investigating node degree, which measures the number of nodes for which a given node will forward content. For this purpose, we assess the maximum node degree, average node degree, server node degree and the complementary cumulative distribution function (CCDF) of node degrees.

1) MAXIMUM NODE DEGREE—GMAX

Figure 4 shows the maximum node degree for all the trees. In a K-C tree, as all nodes have the same degree (except for leafs and parents of leaves), maximum node degree is determined by the node degree of the K-C tree with the best quality (best K-C tree), for each  $\alpha$  value, as we have mentioned in Section 4. In UNIFQ tree, the closer the node is to the server, the greater its degree, to maintain “fairness” in node quality. Thus, the smaller the value of  $\alpha$ , the larger the value of maximum node degree (implying a greater maximum node degree compared to the K-C tree). For  $\alpha = 0$ , maximum node degree is  $n - 1$  ( $n = 60,000$ ). The DPUA tree has a behavior similar to that of the UNIFQ tree.

Under URND trees, the value of  $\alpha$  has no influence on the node selection. Thus, the maximum degree does not statistically vary as a function of  $\alpha$ . In addition, it is well known that under URND the maximum degree is highly concentrated

<sup>2</sup>In Appendices B and C we discuss the extent at which our results for  $n = 60,000$  also hold asymptotically.

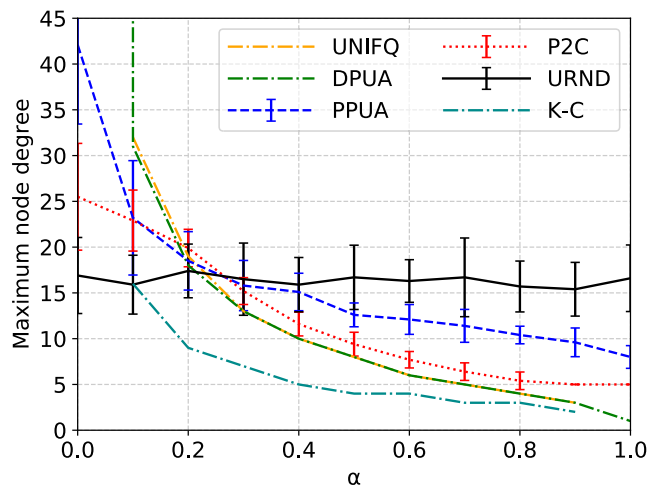


FIGURE 4. Maximum node degree as a function of  $\alpha$ .

around  $\log_2 n$  (Theorem 6.12 in [4]), which corresponds to 15.8 for  $n = 60,000$ .

In P2C trees each newcomer chooses its parent based on the best utility between two nodes selected uniformly at random. In this case, the probability of choosing a leaf increases with respect to  $\alpha$ . Nonetheless, even though the maximum node degree is sensitive to  $\alpha$  it varies in a smaller range of values compared to other alternatives, such as PPUA.

Next, we consider the PPUA tree. We can see that the PPUA tree does not present extreme values when  $\alpha = 0$  or  $\alpha = 1$  as is the case for DPUA, UNIFQ and K-C trees. Recall that as  $\alpha$  tends to zero, the contribution of node degree to the node utility (Equation (1)) goes to zero and only distances are important. Thus, one would expect to observe trees with much larger maximum degrees and shorter distances. In the limit of  $\alpha = 0$ , one would expect the server degree to be proportional to  $n$ , the number of nodes in the tree. However, this is not the case. In the PPUA tree the probability of choosing a specific node (Equation (2)) decreases with  $n$  when  $\alpha$  is very small. Thus, it is extremely unlikely that a single node will attract all newcomers, leading to a star topology. Figure 4 shows this characteristic. Notice that the maximum degree decreases monotonically with  $\alpha$  varying from over 45 (when  $\alpha = 0$ ) to 8 (when  $\alpha = 1$ ).

*Remark 2:* Under deterministic policies (DPUA, UNIFQ and K-C trees) the maximum node degree equals  $n-1$  when  $\alpha = 0$  (not shown in Figure 4) and decreases as  $\alpha$  grows, reaching 1 when  $\alpha = 1$ . Under probabilistic policies (PPUA and P2C), in contrast, the maximum node degree is also a decreasing function of  $\alpha$  but it varies in a much more restricted range of values as randomness in the choice of parents tends to balance node degrees preventing extreme topologies such as the star or line.

## 2) AVERAGE NODE DEGREE—GAVG

In this section, we explicitly distinguish between total degree, in-degree and out-degree. This is in contrast to the rest of this work, where degree is assumed to refer to the out-degree.

In any tree, the average total degree (out-degree plus in-degree) equals

$$\bar{d}^{(t)} = \sum_{v=1}^n d_v^{(t)} / n = 2(n-1) / n = 2 \left(1 - \frac{1}{n}\right), \quad (29)$$

where  $d_v^{(t)}$  is the total degree of vertex  $v$ . Therefore, the average out-degree,  $\bar{d}^{(o)}$ , of any of the trees considered in this work equals

$$\bar{d}^{(o)} = \frac{\left(\sum_{v=1}^n d_v^{(t)}\right) - (n-1)}{n} = 1 - \frac{1}{n}. \quad (30)$$

Next, we consider the average total degree of internal vertices, i.e., all vertices except leaves. Let  $P(L_n = \ell)$  be the probability that a tree with  $n$  vertices has  $\ell$  leaves. Then,

$$\bar{d}^{(i,t)} = \sum_{\ell=1}^{n-1} \sum_{v=1}^{n-\ell} P(L_n = \ell) d_v^{(t)} / (n-\ell) \quad (31)$$

$$= \sum_{\ell=1}^{n-1} P(L_n = \ell) (2(n-\ell-1) + \ell) / (n-\ell) \quad (32)$$

Similarly, we consider the average out-degree of internal vertices,

$$\bar{d}^{(i,o)} = \sum_{\ell=1}^{n-1} \sum_{v=1}^{n-\ell} P(L_n = \ell) d_v^{(o)} / (n-\ell) \quad (33)$$

$$= \sum_{\ell=1}^{n-1} P(L_n = \ell) (n-1) / (n-\ell) \quad (34)$$

In URND trees, for instance, it is known that on average half the nodes are leaves. In addition, for URND trees there is a closed-form expression for  $P(L_n^{(URND)} = \ell)$  [4, equation (6.14)],

$$P(L_n^{(URND)} = \ell) = \frac{1}{(n-1)!} \binom{n-1}{\ell-1} \quad (35)$$

where  $\binom{n}{\ell}$  denotes the Eulerian numbers.

Figure 5 numerically evaluates the above equations for average node degrees, for  $n$  varying between 1 and 50. Note that whereas the average degrees in Equations (29) and (30) hold for any tree with  $n$  vertices, i.e., any instance of a tree with  $n$  vertices satisfies those equations, the average degrees in Equations (31) and (34) vary across tree growth strategies and across tree instances. Therefore, in this section we focus on the average out-degree of internal nodes, given by Equation (34), as illustrated in Figures 6 and 7.

Figure 6 shows the behavior of average out-degree for internal nodes under PPUA, P2C and URND trees as a function of  $\alpha$ . Figure 7 shows the behavior for K-C trees.

Recall that for URND trees the tree topology is insensitive to  $\alpha$ . Figure 6 shows that URND trees have average out-degree of internal nodes equal to 2, in agreement with Equation (34), for large  $n$ .

Next, we consider the behavior of the out-degree of internal nodes of P2C and PPUA trees. P2C trees are constructed from

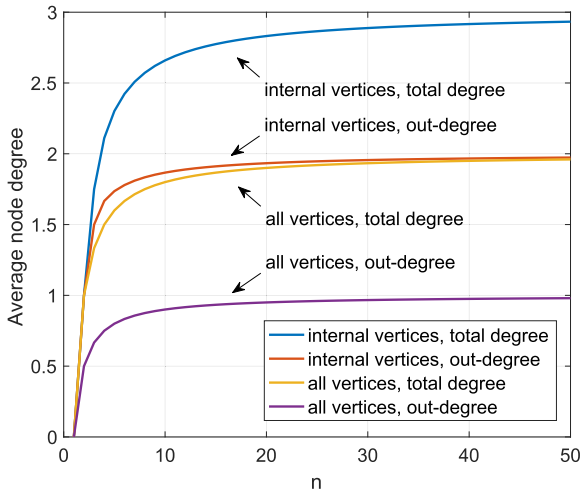


FIGURE 5. Average degrees under URND trees.

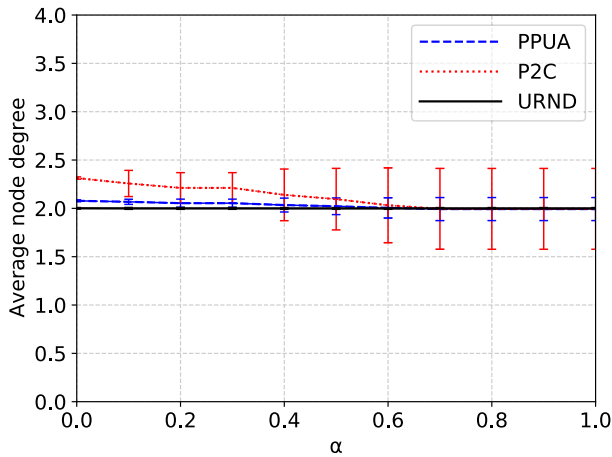


FIGURE 6. Average out-degree of internal nodes as a function of  $\alpha$  for URND, P2C and PPUA trees.

vertices uniformly drawn from the tree. Although its growth process depends on  $\alpha$ , causing an increase in the out-degree of internal nodes for small  $\alpha$ , for larger values of  $\alpha$  the behavior of P2C and URND are roughly the same with respect to average out-degree of internal nodes. Similarly, the average out-degree for internal nodes of PPUA trees will also not vary much as a function of  $\alpha$  because, as we have already pointed out, the PPUA tree does not have extreme values for the degrees of its nodes. As shown in Figure 6, P2C is slightly more sensitive to  $\alpha$  than PPUA.

In K-C trees the optimal degree sequence can be computed as a function of  $\alpha$ , and the behavior of the average out-degree of internal nodes is illustrated in Figure 7 (see optimal  $k$  in Figures 2 and 3). For  $\alpha = 0$ , the optimal K-C tree is a star, with a single internal node with out-degree  $n - 1$  (not shown in Figure 7).

The elements in Figure 7 are computed using the following equation, which is a special case of Equation (34):

$$\bar{d}^{(i,o)} = \frac{n - 1}{n - \ell_n^{(KC)}(\alpha)} \quad (36)$$

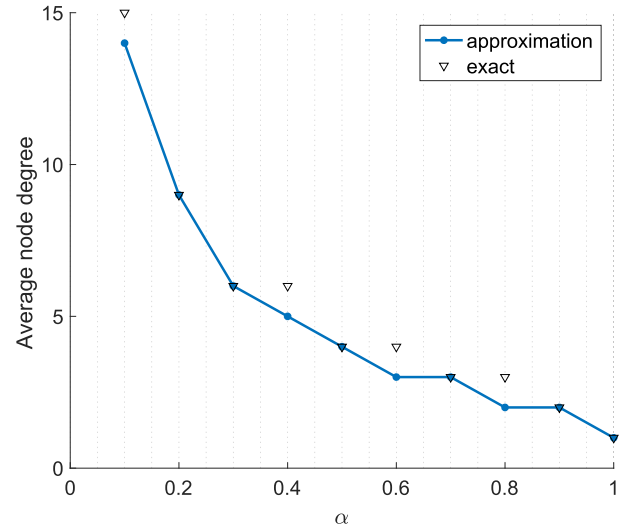


FIGURE 7. Average out-degree of internal nodes as a function of  $\alpha$  for the best K-C trees, with  $n = 60,000$ .

where  $\ell_n^{(KC)}(\alpha)$  is the number of leaves in the best K-C tree with  $n$  nodes (see Equation (40)). As almost all internal nodes in the K-C tree have degree  $k$ , the above expression produces results very close to  $k$ . This, in turn, explains why Figures 3 and 7 are very similar to each other, noting that the optimal degrees in Figure 3 are integers whereas the average degrees in Figure 7 are real numbers.

*Remark 3:* The average total degree (out-degree plus in-degree) of any tree equals  $2(1 - 1/n)$ . Under probabilistic policies (URND, PPUA and P2C) the average out-degree of internal nodes is also close to 2 for all  $\alpha$ , when  $n = 60,000$ . Under deterministic policies (DPUA, UNIFQ and K-C trees) the average out-degree of internal nodes equals  $n - 1$  and 1 when  $\alpha = 0$  and  $\alpha = 1$ , respectively, and varies between those extremes for  $0 < \alpha < 1$ .

### 3) SERVER NODE DEGREE—GSER

Next, we consider the server degree. Figure 8 shows the server node degree as a function of  $\alpha$  for the online trees. One can note that URND and P2C trees show similar behavior, by the same reasons that we have mentioned while discussing the maximum node degree (Section V-A1).

In P2C trees, when both the number of nodes and  $\alpha$  are small, the probability of the server being chosen by a newcomer is high. As the number of nodes increases, this probability decreases. For larger values of  $\alpha$ , as the tree grows the utility of the server tends to decrease, further reducing the probability that the server is selected by a newcomer. Thus, the server degree decreases, albeit not very significantly, as  $\alpha$  increases from 0 to 1. Interestingly, for  $\alpha = 0$  the average server degree under P2C can be shown to be asymptotically equal to  $2 \ln n$  which is twice the average server degree under URND (see [5] and [4], respectively). For  $n = 60,000$ , the average server degree equals 22 and 11, for P2C and URND, respectively.

In the DPUA tree the server has degree equal to  $n - 1$  for  $\alpha = 0$  and degree equal to 1 for  $\alpha = 1$ . As  $\alpha$  grows, the server degree decreases. In addition, note that in the DPUA tree the server degree achieves more extreme values than the ones obtained for PPUA trees.

In PPUA trees, we note a similar trend as the one obtained for the maximum node degree. The server degree decreases monotonically with  $\alpha$ . For  $n = 60,000$ , it ranges from 45 to 5.

*Remark 4:* Under the considered online policies (DPUA, PPUA, P2C and URND) the server degree roughly corresponds to the maximum node degree (compare Figures 4 and 8). In particular, it is well known that under URND the maximum degree is highly concentrated around  $\log_2 n$  (Theorem 6.12 in [4]) whereas the server degree is approximately  $\ln n$  (page 260 in [4]). For  $n = 60,000$ , the maximum degree and the server degree under URND are well approximated by 15.8 and 11, respectively. It is also worth noting that under UNIFQ trees degrees increase as node level grows, and in this case the server degree is typically strictly smaller than the maximum degree.

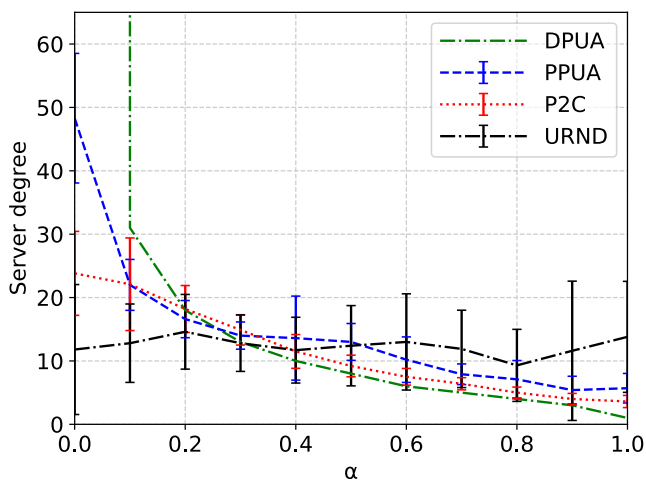


FIGURE 8. Server node degree as a function of  $\alpha$ .

4) CCDF OF NODE DEGREES

Figures 9(a), 9(b) and 9(c) show the CCDF of the node degrees<sup>3</sup> for PPUA, DPUA, URND and P2C trees, for  $\alpha = 0$ ,  $\alpha = 0.5$  and  $\alpha = 1.0$ , respectively. Whereas in Section V-A2 we considered the average out-degree of internal nodes, in this section we consider the CDDF of the out-degree of all nodes, including leaves. We note that the tail of the degree distribution increases as  $\alpha$  decreases, except for the URND tree which is insensitive to  $\alpha$ . For larger values of  $\alpha$ , i.e.,  $\alpha > 0.5$ , the degree distribution drops very sharply. In any case, even for smaller values of  $\alpha$ , the tail does not seem to follow a power law degree distribution (note the semi-log scale of the graph).

<sup>3</sup>Recall that, except otherwise noted, we refer to the out-degree simply as degree.

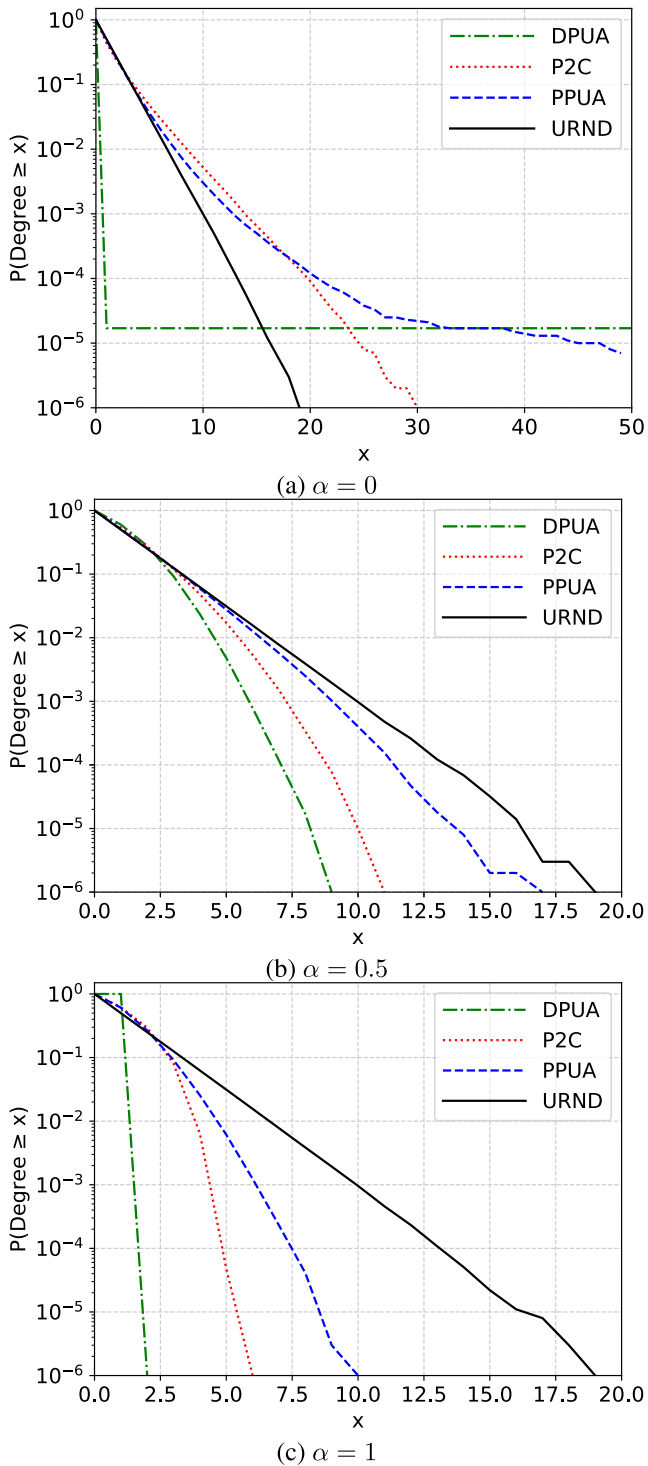


FIGURE 9. CCDF of node degree.

For  $\alpha = 0$  (Figure 9(a)) the tree nodes are expected to be concentrated near the server. Therefore, node degrees are expected to increase when  $n$  increases. For K-C, UNIFQ and DPUA trees, the server degree will be equal to  $n - 1$  and all other nodes will be leaves. Such trees are represent by the DPUA tree in Figure 9(a), noting that the degree of the server equals  $n - 1$  which corresponds to an horizontal line of height

$1.6 \cdot 10^{-5}$  from  $x = 1$  up until  $x = 59,999$ . The step down to zero at  $x = 59,999$  is not shown in Figure 9(a), whose horizontal axis is truncated at 50 to simplify visualization. P2C trees tend to have more nodes with higher degree values than URND trees, given that P2C is sensitive to  $\alpha$ . Indeed, under P2C once two nodes are selected uniformly at random the one with the highest utility is chosen. PPUA trees, in turn, tend to have yet more nodes with higher degrees than P2C. As an example, a fraction of up to  $10^{-5}$  nodes reach degree 49 under PPUA, but the corresponding fraction of nodes under P2C is negligible.

For  $\alpha = 0.5$  (Figure 9(b)), the URND tree has the same behavior as  $\alpha = 0$  whereas the P2C tree tends to have nodes with lower degrees than for  $\alpha = 0$ , as now the utility function depends on both the degree and the distance parameters. For the same reason, the PPUA tree with  $\alpha = 0.5$  also has a degree distribution with degrees that tend to have smaller values than those in the previous case with  $\alpha = 0$ . In addition, degrees of PPUA trees tend to be greater than those of P2C trees. The distribution of degrees for the DPUA tree produces the smallest degree values, as a consequence of the greedy optimization performed at each step of the algorithm.

For  $\alpha = 1.0$  (Figure 9(c)), K-C, UNIFQ and DPUA trees will have a line topology. Therefore,  $n - 1$  nodes will have degree equal to one and the last node (the single leaf) will have degree equal to zero. The URND tree will be statistically equivalent to the previous cases presented in Figures 9(a) and 9(b), as URND is insensitive to  $\alpha$ . P2C and PPUA trees will have a degree distribution with lower values than URND. In addition, as in the previous scenarios, P2C trees have a distribution with lower degree values than PPUA trees.

Under the K-C tree, node degrees will be all smaller than or equal to  $k$ . The optimal value of  $k$ , for any given  $\alpha$ , is discussed in Section IV-A2. In a UNIFQ tree, each level of the tree has nodes with the same degree value, to maintain “fairness” in node quality. The degree values are given in Section IV-A3, and the corresponding CCDF of out-degrees can be readily obtained.

*Remark 5:* In trees where nodes connect to parents uniformly at random (URND), degree distribution has exponentially decreasing tail probabilities [4], [46]. When  $\alpha \geq 0.5$ , the degree distributions of all the tree growing mechanisms considered in this work have tail probabilities that decay faster than an exponential. Turning to the other end of the spectrum, when  $\alpha = 0$ , DPUA and the best K-C and UNIFQ trees yield a star topology, wherein a single node has degree  $n - 1$  and all other nodes have degree 0. The degree distributions of PPUA and P2C, when  $\alpha = 0$ , in turn, have tail probabilities that decay slower than an exponential, as some nodes will have large degrees whereas a significant number of nodes will be leaves. A detailed analysis of the degree distribution under PPUA and P2C when  $\alpha = 0$  is instrumental to determine the asymptotic behavior of PPUA and P2C in this regime, and is left as subject for future work (see Appendix B).

## B. NODE DISTANCES

We now consider node distances, which measure the distance in hops from the node to the root of the tree (i.e., the server). For this purpose, we assess the maximum and average node distance and the CCDF of node distances.

### 1) MAXIMUM NODE DISTANCE—DMAX

Figure 10 shows the maximum distance as a function of  $\alpha$  for all the trees. The maximum distance measures the height of the trees. The distances for K-C, UNIFQ and DPUA trees have similar behavior. Nonetheless, for  $0.7 \leq \alpha \leq 0.9$  the maximum distance of the DPUA tree is significantly larger than that of K-C and UNIFQ trees. Indeed, the “greedy algorithm” used in the growth process of the DPUA tree explains why its maximum distance rapidly increases as a function of  $\alpha$ , given that the relevance of node degrees, captured through  $\alpha$ , grows. When  $\alpha = 1$ , DPUA, K-C and UNIFQ trees all have maximum distance equal  $n - 1$ , as in this case they all correspond to a line graph.

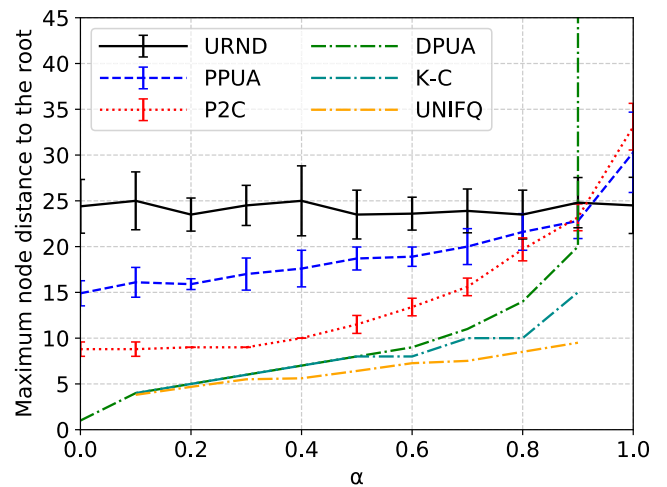


FIGURE 10. Maximum node distance to the root, as a function of  $\alpha$ .

UNIFQ tree has a lower maximum degree compared to K-C and DPUA trees because it performs a global optimization of distances and degrees to produce the tree topology. The K-C tree, in contrast, subsumes a constant degree for a significant fraction of the nodes. This, in turn, leads to higher distances.

Next, we consider probabilistic trees. When nodes select their parents uniformly at random (URND tree), distances are not sensitive to  $\alpha$ . Therefore, the fluctuations of node distance as a function of  $\alpha$  under URND are not statistically significant, and are fruit of the random nature of the uniform sampling of parents.

In the P2C and PPUA trees, as  $\alpha$  increases the probability of selecting a leaf as a parent to a newcomer increases and, consequently, maximum distance tends to increase. In particular, under PPUA, maximum distance ranges from around 15, when  $\alpha = 0$ , to 30, when  $\alpha = 1$ . The range of maximum

distance values, as  $\alpha$  varies from 0 to 1, is smaller under PPUA when compared against P2C or DPUA.

Recall that as  $\alpha$  approaches 1, the contribution of the node distance to the node's utility (Equation (1)) tends to zero. Thus, one could expect to observe much larger distances under PPUA, with maximum distances equal to  $n - 1$  in the limit when  $\alpha = 1$ , as in DPUA. However, this is not the case. As we mentioned before, even when  $\alpha = 0$  or when  $\alpha = 1$ , the probability of choosing the server or a leaf decreases with  $n$ . Thus, it is extremely unlikely that a leaf node will always be chosen by a newcomer, which would lead to a line tree and large distances.

*Remark 6:* Under deterministic policies (DPUA, UNIFQ and K-C trees) the maximum distance equals 1 when  $\alpha = 0$  and increases as  $\alpha$  grows, reaching  $n - 1$  when  $\alpha = 1$  (not shown in Figure 10). Under probabilistic policies (PPUA and P2C), in contrast, the maximum distance is also an increasing function of  $\alpha$  but it varies in a much more restricted range of values as randomness in the choice of parents tends to balance distances preventing extreme topologies such as the star or line.

## 2) AVERAGE DISTANCE—DAVG

Figure 11 shows the average distance from the nodes to the root as a function of  $\alpha$  for all the trees. The average distance under URND and P2C trees are similar, and the rationale is similar to the one discussed above on the maximum distance: both URND and P2C leverage the uniform distribution for parent selection. Although the P2C tree chooses the best node between two randomized ones, its average distance will not be too much different from URND's.

For all the considered trees, except URND, the average distance is an increasing function of  $\alpha$ . Under PPUA, average distance ranges from around 6 to 14. Note that for  $\alpha \leq 0.4$  the average distances of PPUA trees are larger than those of P2C, K-C, UNIFQ and DPUA trees. For  $\alpha = 0.5$  the average distance values of PPUA begin to approach those of the other trees, and for  $\alpha \geq 0.85$  the average distance under PPUA is smaller than that of K-C and DPUA trees. Indeed, the average distances of P2C, K-C and DPUA trees are more sensitive to  $\alpha$  than PPUA, causing a more extreme change in topology when  $\alpha$  grows from 0 to 1 under the former when compared against the latter.

*Remark 7:* Under all considered policies, the average distance behavior is qualitatively similar to that of the maximum distance, slowly increasing as a function of  $\alpha$  from 1 up to 20 when  $0 \leq \alpha < 0.8$  ( $n = 60,000$ ).

## 3) CCDF OF NODE DISTANCES

Figures 12(a), 12(b) and 12(c) show the CCDF of the node distances for PPUA, DPUA, URND and P2C trees, for  $\alpha = 0$ ,  $\alpha = 0.5$  and  $\alpha = 1.0$ , respectively.<sup>4</sup>

<sup>4</sup>The distance distribution of K-C and UNIFQ trees can be easily obtained analytically, without resorting to simulations.

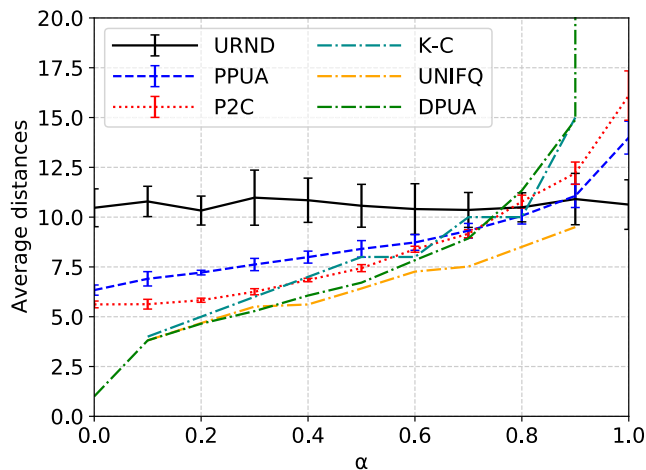


FIGURE 11. Average node distance as a function of  $\alpha$ .

For  $\alpha = 0$  (Figure 12(a)), K-C, UNIFQ and DPUA trees have a star topology, i.e., all nodes have distance equal 1 towards the root. Interestingly, the distances under P2C trees decay to zero faster than those under PPUA and URND trees. The PPUA tree, as expected, does not have an “extreme” topology, even when  $\alpha = 0$ , with most distances ranging from 1 to 8.

For  $\alpha = 0.5$  (Figure 12(b)), the CCDF of the DPUA tree corresponds to a topology with more than 90% of node distances almost uniformly distributed between 1 and 8. This implies that there are, approximately, 8,000 nodes at each level of the tree. For distances larger than 8, the distribution drops very sharply.

The distribution of distances for the P2C tree drops faster than PPUA, with almost 80% of distances ranging from 1 to 10. Note also that the CCDFs of distances for PPUA and P2C trees are not so different when  $\alpha$  grows from 0 to 0.5. This occurs due to their probabilistic nature.

For  $\alpha = 1$  (Figure 12(c)), the CCDF of the DPUA tree becomes a slowly decaying line. Indeed, under DPUA, node distances towards the root are uniformly distributed between 1 and  $n - 1$  (a line tree). As the URND tree is not dependent on  $\alpha$ , its CCDF values are smaller than those corresponding to the CCDFs of DPUA, PPUA and P2C trees, which account for the role of  $\alpha$ . This is because URND tends to produce trees that have smaller distances and correspondingly larger degrees than DPUA, PPUA and P2C when  $\alpha = 1$ . Further comparing PPUA and P2C, we observe that distances under PPUA trees tend to be smaller than those of P2C trees. Although the growth process of P2C trees chooses between two random nodes, the “deterministic choice at the second phase” leads to trees with greater distances, and correspondingly smaller degrees, than PPUA.

*Remark 8:* In trees where nodes connect to parents uniformly at random (URND), almost all nodes have distance to the root close to  $\ln n$ . When  $\alpha = 0$ , the optimal topology is a star, and under DPUA, P2C and PPUA the distances indeed decay to zero faster than URND, with P2C decay being faster

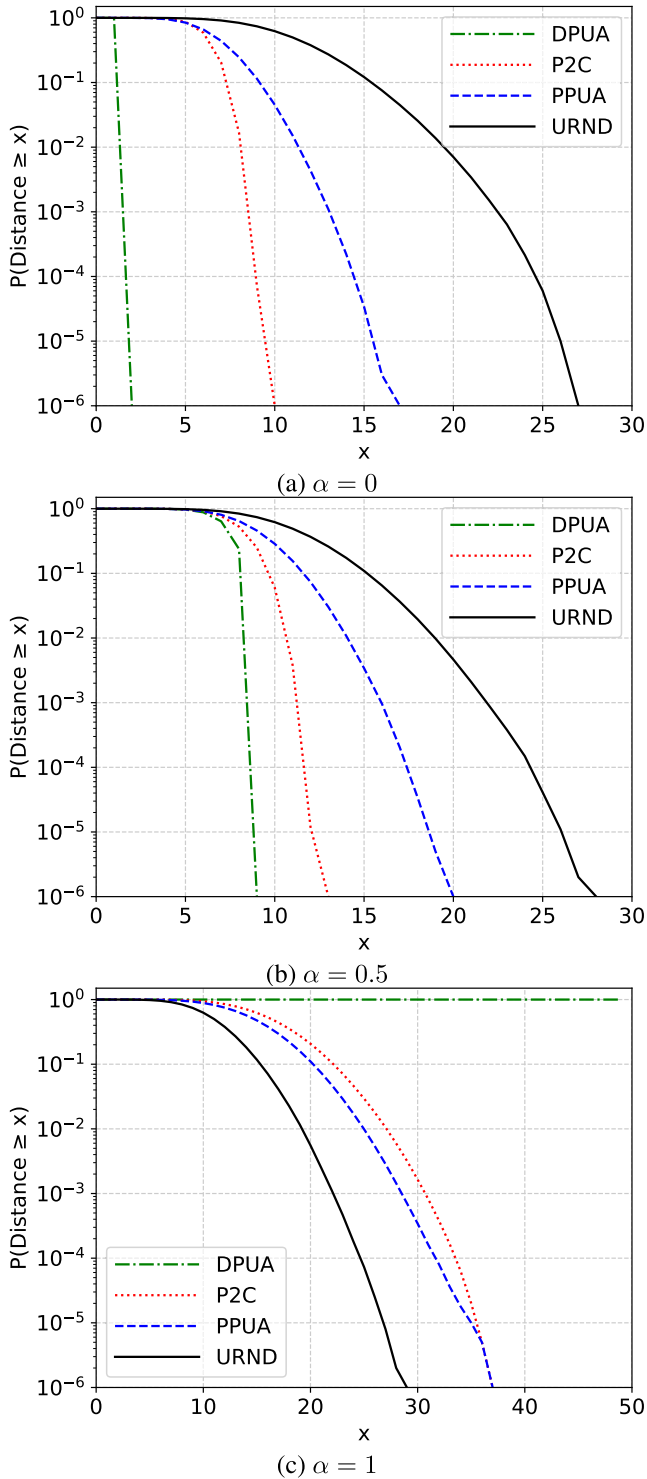


FIGURE 12. CCDF of node distance.

than PPUA. When  $\alpha = 0.5$ , the decay of the tail of node distances under DPUA, P2C and PPUA is still faster than URND. When  $\alpha = 1$ , in contrast, the optimal topology is a line, and distances are uniformly distributed under DPUA. In this case, the tail of distances decays slower under P2C and PPUA when compared against URND.

C. TREE QUALITY

Figure 13 shows the average quality for all considered trees, with  $\alpha$  varying between 0 and 1.

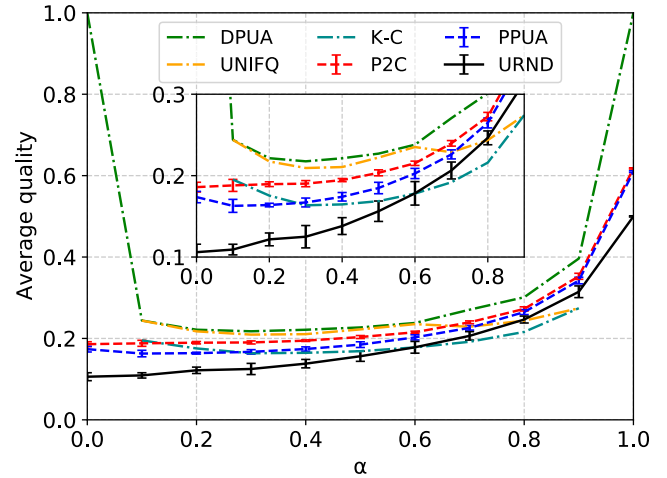


FIGURE 13. Average quality as a function of  $\alpha$  ( $n = 60,000$ ).

The quality of the URND tree increases with respect to  $\alpha$ . This occurs as the average node degree of URND trees is around 2 and the average node distance is around 10.5. Therefore, the larger the weight  $\alpha$  on the degree the higher its quality.

The DPUA tree provides the best quality for all  $\alpha$ . For some values of  $\alpha$  shown in the plots, the quality of DPUA is the same as that of the UNIFQ tree. At the extremes, when  $\alpha$  approaches zero or one, the quality of UNIFQ, DPUA and K-C trees is the same.

Next, we focus on a comparison between PPUA against its counterparts. The quality of DPUA is particularly larger than that of PPUA near the extremes  $\alpha = 0$  and  $\alpha = 1$ . Intuitively, this occurs because the growth process of the PPUA tree cannot generate extremely degenerate trees, like the line tree and the star tree, that yield the best tree quality when  $\alpha = 1$  and  $\alpha = 0$ , respectively.

Despite its simplicity, the quality of the PPUA tree is competitive against the considered alternatives. For  $\alpha \geq 0.35$  the quality of PPUA trees exceeds the quality of K-C trees. For  $\alpha \geq 0.7$ , its quality equals and even exceeds the quality of UNIFQ trees. Indeed, even for  $\alpha$  values where the quality of the PPUA tree is lower than K-C, UNIFQ and DPUA trees, the PPUA tree quality is at most 30% lower than its alternatives (except for  $\alpha \leq 0.1$ ).

The quality of P2C trees is also competitive against PPUA and their counterparts. Indeed, both PPUA and P2C count with randomness in the choice of nodes, but P2C leverages a “deterministic choice at the second phase.” We posit that the deterministic choice of the best of two parents, at the second phase, after they are chosen uniformly at random at the first phase, is one of the reasons for the slightly larger quality of P2C trees when compared against PPUA. Nonetheless, a detailed comparison between PPUA and P2C, including the

asymptotic analysis of the two when  $n \rightarrow \infty$ , is left as subject for future work.

*Remark 9:* The quality of PPUA trees is competitive against their counterparts when  $0.1 < \alpha < 0.9$ . When  $\alpha = 0$ , the asymptotic average quality of PPUA, for  $n \rightarrow \infty$ , remains to be determined (see Appendix B). When,  $\alpha = 1$ , the asymptotic average quality of PPUA equals 0.6 which is larger than 0.5 obtained through URND but lower than 1 obtained using DPUA (see Appendix C). Finally, the quality of PPUA and P2C is similar for  $n = 60,000$ , with P2C being slightly better than PPUA. The asymptotic comparison of PPUA and P2C, for  $n \rightarrow \infty$ , is left as subject for future work (see Appendix C).

#### D. DISCUSSION

Numerical evaluation of the PPUA model reveals some interesting observation about its behavior. Contrary to the classical “preferential attachment” model, where a few nodes tend to dominate the graph, attracting most newcomers, our model leads to a self-organization of the nodes in the tree, without the appearance of any degenerate structure. Even at extremes, when  $\alpha$  approaches zero or one, the PPUA model does not produce trees with degenerated structure. Of course, this may not be optimal at extremes, but otherwise it attests to the robustness of the model under various  $\alpha$  values.

To make this argument precise, consider the case where  $\alpha = 0$ . Assume that, after the arrival of  $n$  nodes to the system, we have a star topology, where the server is the center of the star and all the other nodes are directly connected to it. In this case, the probability that the next newcomer connects to the server, as determined by Equation (2), is  $p_s = 2/(2+n)$ , while the probability it connects to any other node is  $p_v = 1/(2+n)$ , for any leaf node  $v$ . Note that the probability of attaching to the server is twice the probability of attaching to any tagged node  $v$ . However, the probability of attaching to *any* leaf node is  $n/2$  larger than the probability of attaching to the server. Thus, it is very likely that the newcomer will break the star topology, connecting itself to a leaf node and not the server.

Note that throughout the work we considered average node quality as the key metric to be maximized. Alternatively, robustness is another aspect to be considered. Consider, for instance, the scenario  $\alpha = 1$ . Under DPUA, the algorithm converges to a line, which indeed maximizes average node quality. However, the produced topology is extremely fragile: a line will be disconnected whenever any of its nodes, except its single leaf, are removed. Under PPUA, in contrast, the algorithm converges to a topology whose average node quality is less than 1 (namely, 0.609, see Appendix C), but with the advantage of being more robust against node failures, given that there will be more than one leaf. We leave a thorough analysis of the tradeoff between performance and robustness as subject for future work.

Another interesting observation is that the quality of the trees generated by the model is good, comparable and sometimes superior to the quality of the other trees. This hints on the power of a self-organized mechanism to generate

trees. Note that constructing a complete  $k$ -ary (K-C), uniform quality (UNIFQ) or deterministic (DPUA) tree is much more entailed than using a simple probabilistic approach as proposed in the PPUA model (particularly in the DPUA tree, where we need to identify the node with the best quality among all the nodes already in the tree, for each newcomer). Nonetheless, the more rigid structure generated by the offline mechanisms do not necessarily lead to trees with much superior quality, at least when not considering the extreme cases of  $\alpha$  near 0 or 1.

Finally, we note that extreme cases for the utility function, when  $\alpha$  approaches zero or one, are likely not to be of interest in certain applications such as content distribution (extreme cases are discussed in Appendices B and C). As quality depends inherently on both node degree and node distance, these two aspects are likely to be present in the evaluation of system quality. Under this condition, the proposed PPUA model exhibit surprising properties, such as very good tree quality.

#### VI. CONCLUSION AND FUTURE WORK

Constructing efficient real-time dissemination trees is a real burden for tree-based systems. In this paper, we considered a tree growth mechanism based on the idea of “preferential attachment”, where a configurable utility function is used to determine the parent of an arriving node, and compare it against deterministic and other probabilistic tree growth strategies.

Through a numerical evaluation, we observe that the proposed model leads to effective self-organization of the nodes, generating trees that do not exhibit extreme topological properties (e.g., a *star* or *line* topology) but that can deliver very good quality under a wide range of quality measures (for various ranges for  $\alpha$ ). Moreover, the topological properties and quality of the PPUA tree were compared to several other trees generated by different growing processes, showing better or competitive performance. Also, the proposed model showed to have a comparable quality (and sometimes superior) to the deterministic tree and the carefully constructed offline trees (uniform quality and complete  $k$ -ary).

As future work, we are investigating the impact that selfish nodes (i.e., nodes that do not provide service to other nodes). In the current study, all nodes are altruistic and contribute to the system as requested by the mechanism and they never change their position in the tree. Heterogeneous nodes (nodes with different upload capacities) also will be investigated.

In this work we considered preferential recursive trees where the preference is given to nodes of low degrees and closer to the root. In particular, we used a convex linear combination of these two factors inversely. Other utility functions combining degrees and distances are subject for future work.

Throughout this work we considered a fine grained time scale under which arrivals can be assumed to be sequential, and according to which the system can update itself in between consecutive arrivals. We envision that as the system grows, the requirement that it updates after every single arrival can be removed, and that our results still hold if the



system updates after batch arrivals. A detailed analysis of the system accounting for such two time scales corresponding to user arrivals and system updates is left as subject for future work.

**APPENDIX A  
COMPLETE  $k$ -ARY TREES**

Next, we derive closed-form expressions for some of the quantities of interest related to complete K-C trees. The height  $h$  of a  $k$ -ary tree is such that

$$n_{k,h}^{(f)} - n = k^h - \text{mod} \left( n - n_{k,h-1}^{(f)}, k \right) - k \left\lfloor \frac{n - n_{k,h-1}^{(f)}}{k} \right\rfloor \quad (37)$$

where  $n_{k,h}^{(f)}$  is the number of nodes in a full  $k$ -ary tree with height  $h$ ,

$$n_{k,h}^{(f)} = \sum_{l=0}^h k^l = \begin{cases} \frac{1 - k^{h+1}}{1 - k}, & k > 1 \\ h + 1, & k = 1 \end{cases} \quad (38)$$

and

$$h = \begin{cases} \lceil \log_k(-(1-k)n + 1) - 1 \rceil, & k > 1 \\ n - 1, & k = 1. \end{cases} \quad (39)$$

The two last terms in Equation (37) correspond to the number of leaves at the last level whose parents have less than  $k$  and exactly  $k$  children, respectively. Therefore, the number of leaves is given by

$$\ell_{k,n}^{(K-C)} = k^{h-1} + \text{mod} \left( n - n_{k,h-1}^{(f)}, k \right) + \left\lfloor \frac{n - n_{k,h-1}^{(f)}}{k} \right\rfloor k - \left\lfloor \frac{n - n_{k,h-1}^{(f)}}{k} \right\rfloor \quad (40)$$

Let  $q_{k,n}^{(c)}$  denote the average node quality of this tree, as given by Equation (5). In what follows, we drop subscript  $n$  whenever it is clear from context. Thus, we have

$$q_k^{(c)} = \frac{\tilde{q}_k^{(c)}}{n - 1} \quad (41)$$

and

$$\tilde{q}_k^{(c)} = \sum_{i=1}^{h-1} \frac{k^i}{\alpha k + (1 - \alpha)i} + \frac{\lfloor (n - n_{k,h-1}^{(f)})/k \rfloor k}{\alpha k + (1 - \alpha)h} + \frac{\text{mod}(n - n_{k,h-1}^{(f)}, k)}{\alpha \text{mod}(n - n_{k,h-1}^{(f)}, k) + (1 - \alpha)h}. \quad (42)$$

The above expression is used to produce Figure 2.

**A. RANDOM RECURSIVE TREES**

In what follows, we consider two extremes with respect to node quality: networks where nodes are sensitive only to their parents degrees (the larger the degree, the worse), and where nodes are sensitive only to distances to root (the larger the distance, the worse). Let the expected number of nodes with

degree  $k$  (resp., height  $k$ ) in a tree be  $E(X_n^{(k)})$  (resp.,  $E(X_{n,k})$ ).<sup>5</sup> We compute the expected quality per node under three cases: deterministic preferential utility attachment (DPUA), probabilistic preferential utility attachment (PPUA), and uniformly at random node attachment (URND). The trees produced by the latter strategy are referred to as recursive trees in the literature [4].

Recall that  $S$  is the set of nodes in the tree, including the root, and

$$n = |S|.$$

**APPENDIX B  
ACCOUNTING FOR DEPTH**

Next, we consider the case where node qualities are sensitive to depth only ( $\alpha = 0$ ).

**A. TREE QUALITY**

Recall that the tree quality is given by  $\bar{q}$  (see (5)). As pointed above, let  $E(X_{n,k})$  be the expected number of nodes with height  $k$  in a network with  $n$  nodes. Then,

$$E(\bar{q}; n) = \frac{\sum_{k=1}^{n-1} E(X_{n,k})/k}{n - 1} \quad (43)$$

noting that  $E(X_{n,0}) = 1$ , as the root node is assumed to be the only node that has height 0.

Throughout this section, let

$$E(X_{n,k}) = g(k, n). \quad (44)$$

In the particular case where nodes deterministically select their parents (DPUA), all nodes connect to a node with minimum height. We obtain a star, with  $E(\bar{q}) = 1$ . In what follows, we consider uniform recursive trees where nodes connect to their parents uniformly at random (URND), and probabilistic preferential utility attachment (PPUA).

**B. UNIFORM AT RANDOM RECURSIVE TREE (URND)**

In a recursive tree, where nodes select their parents uniformly at random, we have

$$g(0, 1) = 1; \quad g(0, 2) = 1; \quad g(1, 2) = 1 \quad (45)$$

$$g(i, n) = 0, \quad n \leq i \quad (46)$$

$$g(i, n + 1) = g(i, n) + \frac{g(i - 1, n)}{n} \quad (47)$$

Therefore,

$$g(i, n + 1) = \sum_{l=1}^{n-i+1} \frac{1}{l + i - 1} g(i - 1, l + i - 1) \quad (48)$$

*Asymptotics:* According to Section 6.2 of [4] the expected number of nodes with height  $k$ ,  $E(X_{n,k})$ , is given in terms of the Stirling numbers. This result, in turn, is shown to imply that almost all nodes are concentrated around level  $\ln n$  (see Sections 6.2 and 6.3 in [4]),

$$E(X_{n,k}) \sim \frac{n}{\sqrt{2\pi \ln n}} e^{-\frac{(k - \ln n)^2}{2 \ln n}} \quad (49)$$

<sup>5</sup>We follow terminology from [4].

Thus,

$$E(\bar{q}; n) \sim 1/(\ln n).$$

As shown in Figure 14,  $1/\ln(n)$  accurately captures the behavior of expected quality.

*Theorem 1:* In a recursive tree where nodes select their parents uniformly at random (URND), when offered node quality depends only on node distance towards the root ( $\alpha = 0$ ) the expected node quality tends to zero as the tree size grows,

$$E(\bar{q}; n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

*Proof:* Follows immediately from Sections 6.2 and 6.3 and Theorem 6.17 of [4].  $\square$

Alternatively, the rationale behind the above theorem follows from a characterization of the degree distribution of recursive trees, which is known to be well approximated through an exponential distribution (see Remark 1). Therefore, degrees are “small”, which causes distances to grow logarithmically with respect to the number of nodes in the tree. Therefore, the expected tree quality tends to zero as the number of nodes grows to infinity.

### C. PROBABILISTIC PREFERENTIAL UTILITY ATTACHMENT (PPUA)

Nodes connect to each other according to inverse-height probabilistic preferential attachment, also referred to as probabilistic preferential utility attachment (PPUA): the probability of attaching to a node with height  $i$  is inversely proportional to  $1/(i + 1)$ . Then,

$$g(0, 1) = 1, \quad g(0, 2) = 1, \quad g(1, 2) = 1 \quad (50)$$

$$g(i, n) = 0, \quad n \leq i \quad (51)$$

$$g(i, n + 1) = g(i, n) + \frac{g(i - 1, n)/i}{\sum_{i=1}^n g(i - 1, n)/i} \quad (52)$$

$$E(\bar{q}; n) = \frac{\sum_{i=1}^{n-1} g(i, n)/i}{n - 1} \quad (53)$$

Figure 14 indicates that, for finite trees, probabilistic preferential utility attachment (PPUA) provides gains, in expected node quality, on top of the strategy wherein nodes select parents uniformly at random (URND). Does this gain still hold asymptotically? If the answer is affirmative, what is the asymptotic value of expected node quality? If the answer is negative, what is the rate at which expected node quality decays to zero?

Figure 15 shows the average distances towards the root (average node height) as a function of the number of vertices in the tree,  $n$ . DPUA, PPUA and URND correspond to average distances of 1, 6.3 and 10.5 when  $n = 60,000$ , in agreement with Figure 11. In addition, it indicates that average distances for URND grow roughly as  $\ln(n)$ .

Figure 16 shows the CCDF of node distance towards the root, as obtained using the above analytical expressions, for  $n = 60,000$ . The results are contrasted against those obtained

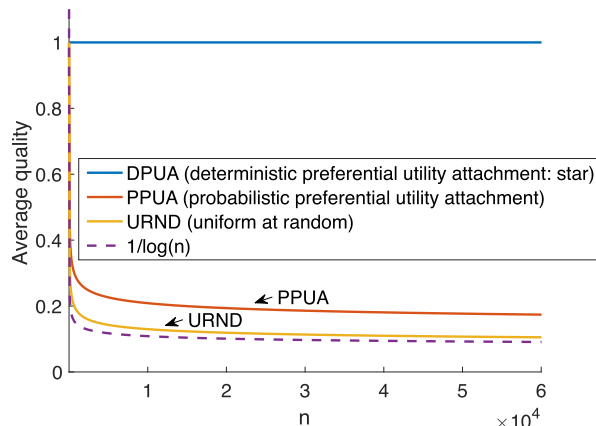


FIGURE 14. Tree quality; nodes sensitive to height, i.e., node quality equals inverse of distance to root ( $\alpha = 0$ ).

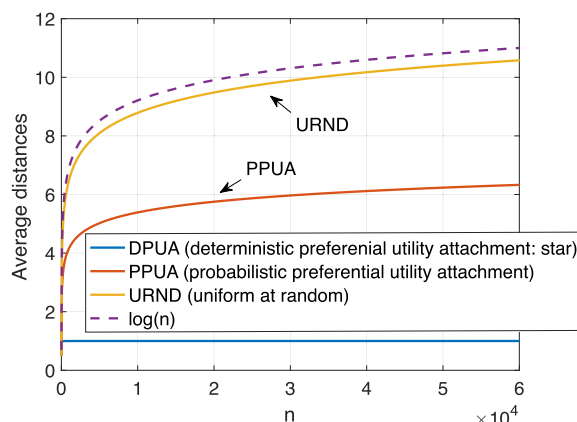


FIGURE 15. Average distances towards the root, i.e., average node height ( $\alpha = 0$ ).

through simulations, also for  $n = 60,000$  (Figure 12) indicating a close agreement between them.

Our preliminary numerical investigations suggest a negative answer on the benefits of PPUA when  $\alpha = 0$ , i.e., we envision that the expected quality per node, under PPUA, slowly decays to 0 as  $n$  grows; in particular, for  $n = 64,000$  and  $n = 1,000,000$  we have  $E(\bar{q}; n) = 0.1729$  and  $E(\bar{q}; n) = 0.1383$ , respectively (the case  $n = 60,000$  is illustrated in Figures 13 and 14), motivating the following conjecture.

*Conjecture 1 (Probabilistic preferential utility attachment is asymptotically neutral when  $\alpha = 0$ ):* Under probabilistic preferential utility attachment (PPUA), when offered node quality depends only on node distance towards the root ( $\alpha = 0$ ) the expected node quality tends to zero as the tree size grows,

$$E(\bar{q}; n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

### APPENDIX C ACCOUNTING FOR DEGREE

Next, we consider the case where node qualities are sensitive to degree only ( $\alpha = 1$ ).

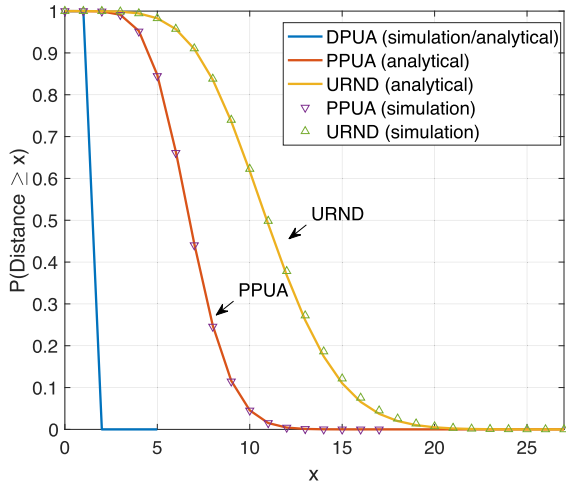


FIGURE 16. CCDF of node distance towards the root, i.e., node height ( $\alpha = 0$ ).

**A. TREE QUALITY**

Let the quality of a node be given by the inverse of the out-degree of its parent. As before, let the quality of a tree be the average of the qualities of its nodes (see (5)). We denote its mean by  $E(\bar{q})$ .

Interestingly,  $E(\bar{q})$  can be expressed as a function of the expected number of leaves, i.e., the expected number of nodes with out-degree zero,  $E(X_n^{(0)})$ . Indeed, let  $Q_i$  be the number of nodes whose parents have out-degree  $i$ . Then,

$$E(Q_i; n) = iE(X_n^{(i)})$$

Therefore,

$$\bar{q} = \frac{Q_1 + Q_2 \frac{1}{2} + \dots + Q_{n-1} \frac{1}{n-1}}{n-1} \tag{54}$$

$$E(\bar{q}; n) = \frac{E(X_n^{(1)}) + E(X_n^{(2)}) \frac{1}{2} + \dots + E(X_n^{(n-1)}) \frac{1}{n-1}}{n-1} \tag{55}$$

$$= \frac{\sum_{i=1}^{n-1} E(X_n^{(i)})}{n-1} = \frac{n - E(X_n^{(0)})}{n-1} \tag{56}$$

In particular, in the case where all nodes deterministically connect to a node with minimum out-degree, we obtain a line graph, and the average node quality is 1 for all  $n$ ,

$$E(\bar{q}) = 1.$$

Throughout this section, let

$$E(X_n^{(k)}) = f(k, n).$$

**B. UNIFORM AT RANDOM RECURSIVE TREE (URND)**

In this case,

$$E(\bar{q}; n) = \frac{n - E(X_n^{(0)})}{n-1} = \frac{n/2}{n-1} \tag{57}$$

where the above result follows from the fact that the expected number of leaves in a recursive tree with  $n$  nodes,  $n \geq 2$ ,

is  $n/2$  (see details below). In addition, we can derive the variance of the tree quality,  $V(\bar{q})$ , using results from [4],

$$V(\bar{q}) = V\left(\frac{X_n^{(0)}}{n-1}\right) = \frac{V(X_n^{(0)})}{(n-1)^2} = \frac{7n/12 + 1/3}{(n-1)^2} \tag{58}$$

**1) DEGREE DISTRIBUTION**

Next, we characterize the expected number of nodes with a given out-degree. To that aim, note that,

$$f(0, 1) = 1, f(0, 2) = 1, f(1, 2) = 1 \tag{59}$$

$$f(i, n) = 0, \quad n \leq i \tag{60}$$

$$f(0, n+1) = \frac{f(0, n)}{n} f(0, n) + \left(1 - \frac{f(0, n)}{n}\right) (f(0, n) + 1) \tag{61}$$

$$f(i, n+1) = \frac{f(i, n)}{n} (f(i, n) - 1) + \frac{f(i-1, n)}{n} (f(i, n) + 1) + \left(1 - \frac{f(i, n)}{n} - \frac{f(i-1, n)}{n}\right) f(i, n) \tag{62}$$

Therefore,

$$f(0, n+1) = f(0, n) + \left(1 - \frac{f(0, n)}{n}\right) \tag{63}$$

$$= \frac{n-1}{n} f(0, n) + 1 \tag{64}$$

$$f(i, n+1) = \frac{n-1}{n} f(i, n) + \frac{1}{n} f(i-1, n) \tag{65}$$

Next, we specialize the above results for  $i = 0, 1, 2$ .

**2) EXPECTED NUMBER OF LEAVES**

For  $i = 0$ ,

$$f(0, n+1) = \frac{f(0, 2)}{n} + \frac{1}{n} \left(\frac{n(n+1)}{2} - 1\right) = \frac{n+1}{2} \tag{66}$$

*Theorem 2:* In a recursive tree where nodes select their parents uniformly at random (URND), when offered node quality depends only on node degree ( $\alpha = 1$ ) the expected node quality tends to 0.5 as the tree size grows,

$$E(\bar{q}; n) \rightarrow 0.5 \text{ as } n \rightarrow \infty.$$

*Proof:* Follows immediately from (66) and (56).  $\square$

**3) GENERAL RECURSION**

Note that it follows from (65) and (66) that

$$f(i, n+1) = \frac{1}{n} \sum_{l=i}^n f(i-1, l). \tag{67}$$

In particular, the above recursion is in agreement with the average out-degree being equal to  $1 - 1/n$  in any tree with  $n$  vertices (see Equation (30) and Figure 5). Indeed,

$$\bar{d}_{n+1}^{(o)} = \frac{1}{n+1} \sum_{i=0}^n i f(i, n+1) \tag{68}$$

$$= \frac{1}{n(n+1)} \sum_{i=0}^n i \sum_{l=i}^n f(i-1, l) \quad (69)$$

$$= \frac{1}{n(n+1)} \sum_{l=1}^n (2l-1) = 1 - \frac{1}{n+1} \quad (70)$$

4) EXPECTED NUMBER OF NODES WITH OUT-DEGREE ONE

For  $i = 1$ ,

$$f(1, n+1) = \frac{1}{n} \sum_{l=1}^n f(0, l) = \frac{1}{n} \left( 1 + \sum_{l=2}^n \frac{l}{2} \right) \quad (71)$$

$$= \frac{1}{n} \left( \frac{2+n+n^2}{4} \right) = \frac{n}{4} + \frac{1}{n} \left( \frac{2+n}{4} \right) \quad (72)$$

5) EXPECTED NUMBER OF NODES WITH OUT-DEGREE TWO

For  $i = 2$ ,

$$f(2, n+1) = \frac{1}{n} \sum_{l=1}^n f(1, l) \quad (73)$$

$$= \frac{1}{n} \sum_{l=2}^n \frac{1}{l-1} \left( \frac{2+(l-1)+(l-1)^2}{4} \right) \quad (74)$$

$$= \left( \frac{1}{n} \sum_{l=2}^n \frac{1/2}{l-1} \right) + \frac{1}{2} \left( f(1, n+1) - \frac{1}{n} \right) \quad (75)$$

$$= \frac{n}{8} + \left( \frac{1}{2n} \sum_{l=2}^n \frac{1}{l-1} \right) + \frac{2+n}{8n} - \frac{1}{2n} \quad (76)$$

*Remark:* In [49] the authors provide an expression for the expected number of nodes with a given degree (sum of in-degrees and out-degrees). We envision that the above steps to derive an expression for  $f(i, n)$  for  $i = 0, 1, 2$  can be generalized for  $i \geq 3$  in order to find a general expression for  $f(i, n)$ , but leave it as subject for future work (see also Lemma 6.14 in [4]). For the purposes of this work,  $f(0, n)$  suffices in order to derive the expected tree quality.

**C. PROBABILISTIC PREFERENTIAL UTILITY ATTACHMENT (PPUA)**

Consider a growing tree where nodes attach to parents with probability inversely proportional to the parents out-degree plus one. As before, let  $E(X_n^{(k)}) = f(k, n)$  be the expected number of nodes with out-degree  $k$  in a network with  $n$  nodes. Then,

$$f(0, 1) = 1, \quad f(0, 2) = 1, \quad f(1, 2) = 1, \quad (77)$$

$$f(i, n) = 0, \quad n \leq i \quad (78)$$

In addition,

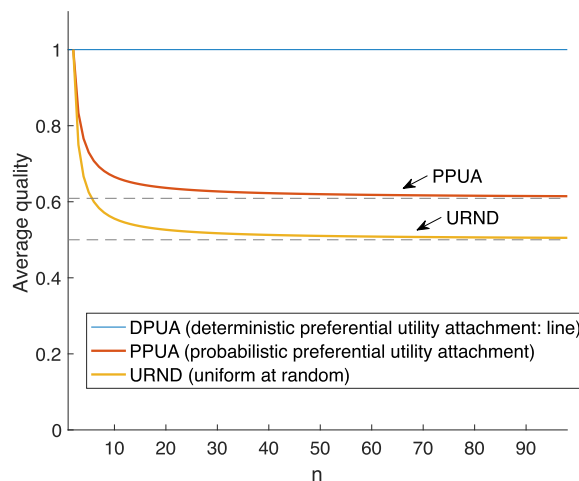
$$f(0, n+1) = \frac{f(0, n)}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} f(0, n) + \left( 1 - \frac{f(0, n)}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} \right) (f(0, n) + 1) \quad (79)$$

$$f(i, n+1) = \frac{f(i, n)/(i+1)}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} (f(i, n) - 1) + \frac{f(i-1, n)/i}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} (f(i, n) + 1) + \left( 1 - \frac{f(i, n)/(i+1) + f(i-1, n)/i}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} \right) f(i, n) \quad (80)$$

Then,

$$f(0, n+1) = f(0, n) + \left( 1 - \frac{f(0, n)}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} \right) \quad (81)$$

$$f(i, n+1) = f(i, n) - \frac{f(i, n)/(i+1)}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} + \frac{f(i-1, n)/i}{\sum_{i=0}^{n-1} f(i, n)/(i+1)} \quad (82)$$



**FIGURE 17. Tree quality: nodes sensitive to degrees, i.e., node quality equals inverse of parent out-degree ( $\alpha = 1$ ).**

Figure 17 indicates that for trees with  $n \leq 50$  probabilistic preferential utility attachment (PPUA) provides gains in terms of expected node quality on top of the uniform at random strategy (URND), where the expected node quality under PPUA is obtained from (56), (81) and (82). For  $n = 60, 000$ , average quality under PPUA and URND is depicted in Figure 13, again indicating the gains of PPUA on top of URND.

Figure 18 shows the CCDF of node degree, as obtained using the above analytical expressions, for  $n = 10, 000$ . The results are contrasted against those obtained through simulations for  $n = 60, 000$  (Figure 9(c)) indicating a close agreement between them.

Under the diffusion approximation considered in [6], the gain of PPUA on top of URND holds for finite  $n$  as well as asymptotically, noting that the asymptotic value of expected node quality under PPUA (resp., URND) equals 0.60905 (resp., 0.5). Indeed, under the diffusion approximation considered in [6] we prove the following result.

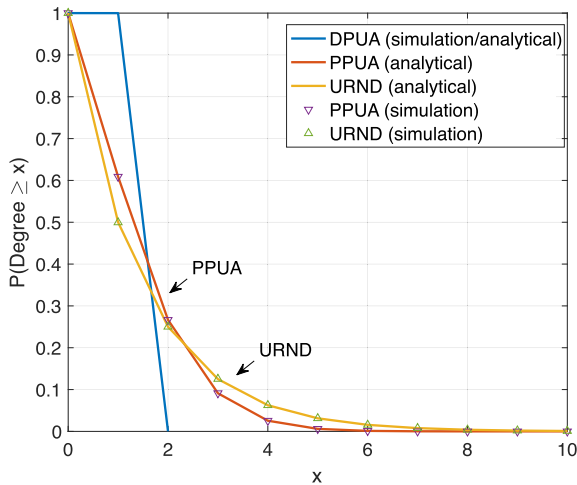


FIGURE 18. CCDF of out-degree ( $\alpha = 1$ ).

**Theorem 3 (Probabilistic Preferential Utility Attachment Is Asymptotically Beneficial When  $\alpha = 1$ ):** Under probabilistic preferential utility attachment (PPUA), when offered node quality depends only on node degree ( $\alpha = 1$ ) the expected node quality tends to a constant greater than 0.5 as the tree size grows,

$$E(\bar{q}; n) \rightarrow 0.60905 \text{ as } n \rightarrow \infty.$$

*Proof:* It follows from [6] that for large  $n$  we have

$$E(X_n^{(0)}) \rightarrow n \frac{1}{1 + \frac{1}{0.6419}} \quad (83)$$

Therefore, substituting (83) into (56),

$$E(\bar{q}; n) \rightarrow \frac{n(1 - 1/(1 + 1/0.6419))}{n - 1}$$

as  $n \rightarrow \infty$ .  $\square$

*Remark:* The proof of the above theorem relies on the diffusion approximation considered in [6]. Such an approximation is extremely accurate, specially for large networks. Nonetheless, a probabilistic proof of Theorem 3 or a rigorous argument showing that the diffusion approximation indeed captures the asymptotic behavior of the original stochastic system remains open.

Inspired by the above result and Conjecture 1, we pose a conjecture on the benefits of probabilistic preferential utility attachment when new nodes account for the distance of their parents towards the root, together with degrees, while determining how to connect to the network.

**Conjecture 2 (Asymptotic Benefits of Probabilistic Preferential Utility Attachment):** There is a threshold  $\alpha^*$  such that

- for  $\alpha \leq \alpha^*$  the expected node quality  $E(\bar{q}; n)$  under probabilistic preferential utility attachment (PPUA) is asymptotically equal to that obtained under the classical recursive tree model where nodes select their parents uniformly at random (URND). In this regime where

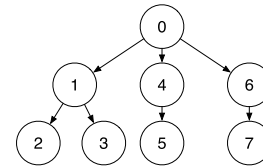


FIGURE 19. Optimal graph for  $\alpha = 0.2$  and  $n = 8$ .

$\alpha \leq \alpha^*$ , PPUA outperforms URND for finite  $n$ , but is neutral for  $n = \infty$

- for  $\alpha > \alpha^*$ , PPUA is asymptotically beneficial, i.e., the expected node quality  $E(\bar{q}; n)$  under PPUA is strictly greater than that obtained under URND for any finite value of  $n$  as well as in the limit when  $n = \infty$ .

Finally, we also pose a conjecture on the power of two choices:

**Conjecture 3 (Benefits of power of choice):** The expected node quality under power of two choices (P2C) is strictly greater than that obtained under URND or PPUA for any finite value of  $n$  as well as in the limit when  $n = \infty$ .

#### APPENDIX D OPTIMAL TREES

Next, we indicate through an example that optimal trees may be beyond the scope of the algorithms discussed in this work. Indeed, the optimal tree may not be obtained through any of the deterministic algorithms considered in this paper, and its characterization is left as subject for future work.

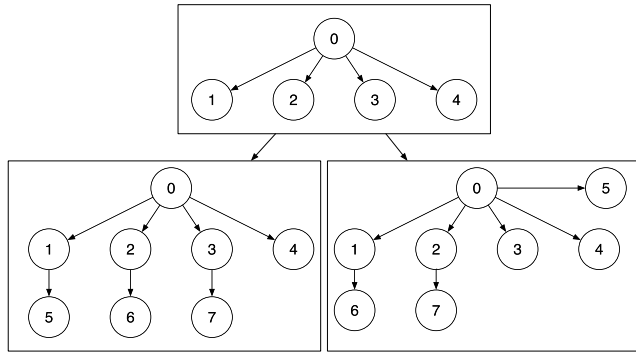
Let  $\alpha = 0.2$  and  $n = 8$ . The optimal tree has the root connected to 3 children, 1 of those children connected to 2 additional nodes, and the other 2 children connected to one additional node each (Figure 19). The average quality is

$$\begin{aligned} \bar{q} &= \frac{3/(0.6 + 0.8) + 2/(0.4 + 1.6) + 2/(0.2 + 1.6)}{7} \\ &= 0.6077 \end{aligned}$$

DPUA, in contrast, yields an average quality of 0.595238, with the root connected to 4 nodes, and 3 of those nodes connected to 1 other node (Figure 20).

$$\begin{aligned} \bar{q}^{(DPUA)} &= \frac{4/(0.8 + 0.8) + 3/(0.2 + 1.6)}{7} \\ &= 0.595238 \end{aligned}$$

Under DPUA, note that after inserting node 4 there is a tie with respect to where to insert node 5. If node 1 (resp., node 0) is chosen as the parent of node 5, we obtain the tree in the bottom left (resp., bottom right) of Figure 20. Indeed, in both cases node 5 will have a quality of  $1/(1.0 + 0.8) = 1/(0.2 + 1.6) = 1/1.8$ . However, if node 0 is set as the parent of node 5, node 5 produces a negative externality towards the other nodes connected to node 0, whereas if node 1 is selected as parent to node 5 the qualities of the nodes already in the tree do not change. In this particular example, choosing node 1 as parent of node 5 yields the best DPUA tree.



**FIGURE 20.** Graphs obtained through DP UA ( $\alpha = 0.2, n = 8$ ). Note that after inserting node 4 there is a tie with respect to where to insert node 5. If node 1 (resp., node 0) is chosen as the parent of node 5, we obtain the tree in the bottom left (resp., bottom right). The tree in the bottom left is the best DP UA tree, whereas the tree in the bottom right is a UNIFQ tree wherein all nodes have quality  $1/(0.2 \times 5 + 0.8) = 1/1.8$ .

**TABLE 5.** Average node quality for K-C trees with  $\alpha = 0.2$  and  $n = 8$ .

$k$	$\bar{q}$
1	0.406399659775493
2	0.578754578754579
3	0.580292723149866
4	0.551948051948052
5	0.539682539682540
6	0.507936507936508
7	0.454545454545455

The best K-C tree can also be verified to be suboptimal (see Table 5). Indeed,  $k = 3$  yields the best K-C tree, with average node quality equal to 0.58.

Finally, UNIFQ entails a root with degree 5 or larger (to satisfy the condition  $d_0 > 1/\alpha - 1 = 4$ ), which again is not optimal. In particular, in the tree where all 7 nodes except the root have the same quality, such quality equals  $1/1.8$  which is smaller than 0.6077. Interestingly, such tree wherein all nodes except the root have quality  $1/1.8$  is also obtained through DP UA, as discussed above (see Figure 20).

Clearly, the greedy DP UA strategy is suboptimal. It remains to be determined whether the construction of optimal trees is an NP-hard problem or whether efficient algorithms, possibly leveraging dynamic programming, can be designed to find such trees.

**REFERENCES**

[1] Y.-H. Chu, S. G. Rao, S. Seshan, and H. Zhang, "A case for end system multicast," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 8, pp. 1456–1471, Oct. 2002. [Online]. Available: <http://ieeexplore.ieee.org/document/1038577/>

[2] S. Banerjee, S. Lee, R. Braud, B. Bhattacharjee, and A. Srinivasan, "Scalable resilient media streaming," in *Proc. 14th Int. Workshop Netw. Operating Syst. Support Digit. Audio Video (NOSSDAV)*, New York, NY, USA, 2004, p. 4. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=1005847.1005851>

[3] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Rev. Modern Phys.*, vol. 74, no. 1, pp. 47–97, Jun. 2001.

[4] M. Drmota, *Random Trees: An Interplay Between Combinatorics and Probability*. New York, NY, USA: Springer, 2009.

[5] H. M. Mahmoud, "The power of choice in the construction of recursive trees," *Methodology Comput. Appl. Probab.*, vol. 12, no. 4, pp. 763–773, Dec. 2010.

[6] V. Sevim and P. A. Rikvold, "Effects of preference for attachment to low-degree nodes on the degree distributions of a growing directed network and a simple food-Web model," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 73, no. 5, May 2006, Art. no. 056115.

[7] S. Fujita, "Multi-tree-based peer-to-peer video streaming with a guaranteed latency," *IEICE Trans. Inf. Syst.*, vol. 102, no. 9, pp. 1707–1714, 2019.

[8] C. Yong Goh, H. Shyong Yeo, H. Lim, P. Kuan Hoong, J. W. Lim, and I. K. Tan, "A comparative study of tree-based and mesh-based overlay P2P media streaming," *Int. J. Multimedia Ubiquitous Eng.*, vol. 8, no. 4, pp. 97–106, 2013.

[9] S. Fujita, "On the cost of cloud-assistance in tree-structured P2P live streaming," in *Proc. IEEE Int. Parallel Distrib. Process. Symp. Workshops (IPDPSW)*, May 2018, pp. 820–828.

[10] O. C. Kwon and H. Song, "Adaptive tree-based P2P video streaming multicast system under high peer-churn rate," *J. Vis. Commun. Image Represent.*, vol. 24, no. 3, pp. 203–216, Apr. 2013. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1047320312001678>

[11] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross, "A measurement study of a large-scale P2P IPTV system," *IEEE Trans. Multimedia*, vol. 9, no. 8, pp. 1672–1687, Dec. 2007.

[12] N. Magharei, R. Rejaie, and Y. Guo, "Mesh or multiple-tree: A comparative study of live P2P streaming approaches," in *Proc. 26th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, May 2007.

[13] R. Lo Cigno, A. Russo, and D. Carra, "On some fundamental properties of P2P push/pull protocols," in *Proc. 2nd Int. Conf. Commun. Electron.*, Jun. 2008, pp. 67–73.

[14] T. Small, B. Li, and B. Liang, "On optimal peer-to-peer topology construction with maximum peer bandwidth contributions," in *Proc. 23rd Biennial Symp. Commun.*, 2006, pp. 157–160.

[15] S. Liu, R. Zhang-Shen, W. Jiang, J. Rexford, and M. Chiang, "Performance bounds for peer-assisted live streaming," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 36, no. 1, pp. 313–324, Jun. 2008.

[16] M. Sayit, S. Demirci, Y. Kaymak, and E. T. Tunali, "Adaptive, incentive and scalable dynamic tree overlay for P2P live video streaming," *Peer-to-Peer Netw. Appl.*, vol. 9, no. 6, pp. 1074–1088, Nov. 2016.

[17] M. R. Korulkar and S. Sulochana, "An efficient tree-based self-organizing protocol for Internet of Things," *IEEE Access*, vol. 4, pp. 3535–3546, 2016.

[18] L. Maccari, N. Facchi, L. Baldesi, and R. Lo Cigno, "Optimized P2P streaming for wireless distributed networks," *Pervas. Mobile Comput.*, vol. 42, pp. 335–350, Dec. 2017.

[19] H. Terelius and K. H. Johansson, "Peer-to-peer gradient topologies in networks with churn," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 4, pp. 2085–2095, Dec. 2018.

[20] R. Hasimoto-Beltran, F. D. A. Lopez-Fuentes, and M. Vera-Lopez, "Hierarchical P2P architecture for efficient content distribution," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 4, pp. 724–739, Jul. 2019.

[21] S. Budhkar and V. Tamarapalli, "An overlay management strategy to improve QoS in CDN-P2P live streaming systems," *Peer-to-Peer Netw. Appl.*, vol. 13, no. 1, pp. 190–206, Jan. 2020.

[22] M. Zhao, P. Aditya, A. Chen, Y. Lin, A. Haeberlen, P. Druschel, B. Maggs, B. Wishon, and M. Ponec, "Peer-assisted content distribution in akamai netsession," in *Proc. Conf. Internet Meas. Conf.*, Oct. 2013, pp. 31–42.

[23] N. Anjum, D. Karamshuk, M. Shikh-Bahaei, and N. Sastry, "Survey on peer-assisted content delivery networks," *Comput. Netw.*, vol. 116, pp. 79–95, Apr. 2017.

[24] X. Zhang and H. Hassanein, "A survey of peer-to-peer live video streaming schemes—An algorithmic perspective," *Comput. Netw.*, vol. 56, no. 15, pp. 3548–3579, Oct. 2012. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1389128612002393>

[25] A. WaysiAITuhafi, "A review on peer-to-peer live video streaming topology," *Int. J. Comput. Appl.*, vol. 68, no. 5, pp. 5–14, 2013.

[26] J.-F. Huang, S.-S. Kao, S.-Y. Hsieh, and R. Klasing, "Top-down construction of independent spanning trees in alternating group networks," *IEEE Access*, vol. 8, pp. 112333–112347, 2020.

[27] A. Hameed, R. Dai, and B. Balas, "A decision-tree-based perceptual video quality prediction model and its application in FEC for wireless multimedia communications," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 764–774, Apr. 2016.

[28] K. Li, C.-C. Hua, X. You, and X.-P. Guan, "Distributed consensus control for nonlinear multiagent systems under directed graphs of dynamic frequency switches," *IEEE Trans. Autom. Control*, vol. 66, no. 2, pp. 841–848, Feb. 2021.

- [29] W. Dong, S. Liu, Y. Ding, X. Sheng, and X. Zhu, "An artificially weighted spanning tree coverage algorithm for decentralized flying robots," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 4, pp. 1689–1698, Oct. 2020.
- [30] L. Ma, S. Cheng, and Y. Shi, "Enhancing learning efficiency of brain storm optimization via orthogonal learning design," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 11, pp. 6723–6742, Nov. 2021.
- [31] L. Ma, M. Huang, S. Yang, R. Wang, and X. Wang, "An adaptive localized decision variable analysis approach to large-scale multiobjective and many-objective optimization," *IEEE Trans. Cybern.*, early access, Jan. 21, 2021, doi: 10.1109/TCYB.2020.3041212.
- [32] S. Pandey, S. Park, and M. J. Choi, "Community oriented in-network caching and edge caching for over-the-top services in adaptive network conditions to improve performance," *Int. J. Netw. Manage.*, vol. 30, no. 4, p. e2104, Jul. 2020.
- [33] D. Carra, R. Lo Cigno, and E. W. Biersack, "Graph based analysis of mesh overlay streaming systems," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 9, pp. 1667–1677, Dec. 2007.
- [34] T. Small, B. Li, and B. Liang, "Outreach: Peer-to-peer topology construction towards minimized server bandwidth costs," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 1, pp. 35–45, Jan. 2007.
- [35] R. Kumar, Y. Liu, and K. Ross, "Stochastic fluid theory for P2P streaming systems," in *Proc. 26th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2007, pp. 919–927.
- [36] D. Carra, R. L. Cigno, and E. W. Biersack, "Stochastic graph processes for performance evaluation of content delivery applications in overlay networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 2, pp. 247–261, Feb. 2008.
- [37] T. Bonald, L. Massoulié, F. Mathieu, D. Perino, and A. Twigg, "Epidemic live streaming: Optimal performance trade-offs," in *Proc. ACM SIGMETRICS Int. Conf. Meas. Model. Comput. Syst.*, 2008, pp. 325–326.
- [38] R. Li, W. Wang, and Z. Di, "Effects of human dynamics on epidemic spreading in Côte d'Ivoire," *Phys. A, Stat. Mech. Appl.*, vol. 467, pp. 30–40, Feb. 2017.
- [39] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics and endemic states in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 63, no. 6, p. 066117, 2001.
- [40] H. M. Mahmoud, R. T. Smythe, and J. Szymański, "On the structure of random plane-oriented recursive trees and their branches," *Random Struct. Algorithms*, vol. 4, no. 2, pp. 151–176, 1993.
- [41] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Sep. 1999.
- [42] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, "Scale-free networks from varying vertex intrinsic fitness," *Phys. Rev. Lett.*, vol. 89, no. 25, Dec. 2002, Art. no. 258702.
- [43] M. Hofri and H. Mahmoud, *Algorithmics of Nonuniformity: Tools and Paradigms*. Boca Raton, FL, USA: CRC Press, 2018.
- [44] M. R. Lyon and H. M. Mahmoud, "Trees grown under young-age preferential attachment," *J. Appl. Probab.*, vol. 57, no. 3, pp. 911–927, Sep. 2020.
- [45] M. Karsai, M. Kiveliä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, "Small but slow world: How network topology and burstiness slow down spreading," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 83, no. 2, Feb. 2011, Art. no. 025102.
- [46] Á. Backhausz and T. F. Móri, "Degree distribution in the lower levels of the uniform recursive tree," 2011, *arXiv:1112.1250*.
- [47] V. Sevim and P. A. Rikvold, "Network growth with preferential attachment for high indegree and low outdegree," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 11, pp. 2631–2636, Apr. 2008.
- [48] H. M. Mahmoud, "The degree profile in some classes of random graphs that generalize recursive trees," *Methodology Comput. Appl. Probab.*, vol. 16, no. 3, pp. 527–538, Sep. 2014.
- [49] H. S. Na and A. Rapoport, "Distribution of nodes of a tree by degree," *Math. Biosci.*, vol. 6, pp. 313–329, Jan. 1970.



**DANIEL RATTON FIGUEIREDO** received the Ph.D. degree in computer science from the University of Massachusetts, Amherst, in 2005. He worked as a Postdoctoral Researcher with the Swiss Federal Institute of Technology, Lausanne (EPFL). In 2007, he joined the Department of Systems Engineering and Computer Science (PESC), Federal University of Rio de Janeiro (UFRJ), Brazil, where he is currently an Associate Professor. His research interests include network science and probabilistic modeling of network dynamics.



**DANIEL FRAÇÃO LUIZ** received the B.S. degree in computer science and the M.S. degree in informatics from the Federal University of Amazonas (UFAM), in 2010 and 2014, respectively, where he is currently pursuing the Ph.D. degree in electrical engineering.

In 2014, he joined the Sidia Institute of Science and Technology as a Software Developer. His research interests include the IoT, embedded systems, and wireless/mobile networks.



**CELSO BARBOSA CARVALHO** received the B.S. degree in electrical engineering from the Federal University of Amazonas (UFAM), in 1999, and the M.S. and Ph.D. degrees in electrical engineering from the Federal University of Rio de Janeiro (COPPE/UFRJ), Brazil, in 2005 and 2012, respectively.

From 2000 to 2006, he integrated the research and development (R&D) teams of the Foundation of Analysis Research and Technological Innovation-Amazonas/Brazil (FUCAPI), Siemens Mobile (BenQ-Amazonas/Brazil), and the Center for Research in Electronics and Information-Amazonas/Brazil (CETELI), where he worked as a Research and Development Engineer and a Project Manager in projects in the areas of industrial automation, telecommunications, improvement of software development processes, and industrial processes improvement. Since 2007, he has been with the Federal University of Amazonas as a full-time Professor in both undergraduate and graduate courses of electrical and computer engineering. His research interests include the IoT, Industry 4.0, wireless embedded systems, and wireless/mobile networks.



**DANIEL SADO MENASCHÉ** (Member, IEEE) received the Ph.D. degree in computer science from the University of Massachusetts, Amherst, in 2011. Currently, he is an Associate Professor with the Computer Science Department, Federal University of Rio de Janeiro, Brazil. His research interests include modeling, analysis, security, and performance evaluation of computer systems. He is an Alumni Affiliated Member of the Brazilian Academy of Sciences. He was a recipient of best paper awards at GLOBECOM 2007, CoNEXT 2009, INFOCOM 2013, and ICGSE 2015.

• • •



**MARCIO NUNES DE MIRANDA** received the Ph.D. degree in electrical engineering from the Federal University of Rio de Janeiro (COPPE/UFRJ), Rio de Janeiro, Brazil, in 2002. Currently, he is a Professor with the Computer Science Department, Federal Rural University of Rio de Janeiro, Brazil. His research interests include modeling, analysis, and performance evaluation of computer networks, focusing on mobile networks and wireless embedded systems.