Ordered increasing k-trees: Introduction and analysis of a preferential attachment network model

Alois Panholzer and Georg Seitz †

Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8-10/104
A-1040 Wien, Austria.
Alois.Panholzer@tuwien.ac.at, Georg.Seitz@tuwien.ac.at

We introduce a random graph model based on k-trees, which can be generated by applying a probabilistic preferential attachment rule, but which also has a simple combinatorial description. We carry out a precise distributional analysis of important parameters for the network model such as the degree, the local clustering coefficient and the number of descendants of the nodes and root-to-node distances. We do not only obtain results for random nodes, but in particular we also get a precise description of the behaviour of parameters for the j-th inserted node in a random k-tree of size n, where $j = j(n)$ might grow with n. The approach presented is not restricted to this specific k-tree model, but can also be applied to other evolving k-tree models.

Keywords: network model, increasing k-trees, degree distribution, local clustering coefficient, root-to-node distances, limiting distributions

1 Introduction

Since the pioneering work of [WS98] on real-world networks (as social networks, biological networks or computer networks), various random network models have been introduced that capture at least part of the typical properties observed frequently. Such properties (see, e.g., [WS98, BA99]) are, e.g., a small average node-to-node distance, a high clustering coefficient, and a power-law degree distribution.

One of the most famous of such random graph models has been introduced by [BA99]. It uses the idea of "preferential attachment" (or "success breeds success"), where, starting with a set of nodes, successively nodes are added and linked to a set of nodes by using a specific stochastic growth rule, namely that the probability that a new node is attached to an already existing node is proportional to the degree of that node. A mathematically rigorous definition of this model together with an analysis of important

† This work has been supported by the Austrian Science Foundation FWF, grant S9608-N23.

parameters has been given in [BR03]. It has been pointed out in that work that plane-oriented recursive trees, an important and heavily analyzed tree model (see, e.g., [MS95] for a definition and early results), are a special instance of the Bárabasi-Albert graph model.

In this work we introduce a random graph model, which is based on so-called k-trees[(i)] (see, e.g., [BP69, Moo69]), but where we apply a preferential attachment rule in order to generate them. Starting with a k-clique (a complete connected graph with k vertices) of nodes (the so-called root-clique) labelled by $0_1, 0_2, \ldots, 0_k$, successively the nodes with labels $1, 2, \ldots, n$ are inserted, where in each step the inserted node will be attached to all of the nodes of an already existing k-clique. But instead of choosing a clique at random we use the probabilistic growth rule that the probability that a new node is attached to an already existing k-clique is proportional to one plus the number of nodes that have been previously attached to this k-clique (the so-called children of the k-clique). In order to also obtain a combinatorial description of these graph families we will consider increasingly labelled ordered k-trees and speak about the model of "ordered increasing k-trees"; a precise definition will be given in Section 2.

From the construction of k-trees it is apparent that for $k = 1$ one obtains the model of plane-oriented recursive trees; thus the here studied ordered increasing k-trees can be considered as graph families that are generalizations of plane-oriented recursive trees. Quite recently k-trees have been introduced as network models in [Gao09, ADS10] and an analysis of important parameters has been given. In contrast to the model we are introducing the considered k-trees are there generated by a uniform attachment rule, i.e., in each step a new node is attached to a randomly chosen already existing k-clique. Combinatorially one might speak then about the model of "unordered increasing k-trees", which leads for the special instance $k = 1$ to the model of (uniform) recursive trees.

We will give a precise distributional analysis of important parameters in ordered increasing k-trees such as the degree, the local clustering coefficient and the number of descendants of the nodes and root-to-node distances. We are here not only interested in a study of quantities for random nodes, but a main emphasis is given on describing the behaviour of parameters for the j-th inserted node in a random k-tree of size n, depending on the growth of $j = j(n)$: we can give a complete characterization of the limit laws appearing; partially we even obtain exact results. Thus the local behaviour of the nodes during the graph evolution process is described quite well. Furthermore, using this precise information on the behaviour of the parameters for the j-th inserted node we will easily deduce also the limiting behaviour for randomly selected nodes in the k-tree (partially we obtain again even exact results). In particular we can show that the distribution of the node-degrees follows asymptotically a so-called power law, i.e., the probability that a randomly selected node has degree d behaves asymptotically as $\sim c d^{-2-\frac{1}{k}}$, and that the expected local clustering coefficient is rather high (e.g., for $k = 2$ it is asymptotically, for $n \to \infty$, given by $23 - \frac{9}{4}\pi^2 \approx 0.793390\ldots$). Moreover the root-to-node distance of node n (but also of a random node), is asymptotically Gaussian with expectation $\sim \frac{1}{(k+1)H_k} \log n$ and variance $\sim \frac{H_k^{(2)}}{(k+1)H_k^3} \log n$, where $H_k = \sum_{\ell=1}^{n} \frac{1}{\ell}$ and $H_k^{(2)} = \sum_{\ell=1}^{n} \frac{1}{\ell^2}$ denote the first and second order harmonic numbers.

To show our results we use both descriptions of ordered increasing k-trees, namely (i) the description via the graph evolution process which often gives rise to a "bottom-up approach" when considering the parameter before and after inserting node n, and (ii) the combinatorial description as ordered increasing k-trees which often allows a "top-down approach" when using a decomposition of the k-tree with respect to the root-clique, see Section 2. The latter approach has been applied with success in [DS09, ADS10] to

[(i)] Here $k \geq 1$ is always an integer. The term k-trees, also called k-dimensional trees, is somewhat misleading, since, for $k \geq 2$, these graphs are no more trees. In particular they should not be confused with k-ary trees, which are indeed trees.

other k-tree models as in particular to randomly labelled k-trees. Both descriptions turn out to be quite useful when analyzing parameters in ordered increasing k-trees; to show our results for the node-degree, the local clustering coefficient and the number of descendants we use the bottom-up approach, whereas for obtaining results on the root-to-node distance we use the top-down approach.

2 Ordered increasing k-trees

k-trees are families of simple graphs, which have been introduced by [BP69]. k-trees might be defined recursively in a way analogous to trees: a k-tree T is either a k-clique (i.e., a complete connected graph with k vertices) or there exists a node u (one might call u endnode), which is incident to exactly k edges that connect this node to all of the vertices of a k-clique, such that, when removing u and the k incident edges from T, the remaining graph is itself a k-tree. In this paper we will always consider rooted k-trees, which means that in each k-tree one k-clique is distinguished as the root-clique (the nodes contained in the root-clique are called root nodes, whereas the remaining nodes are non-root nodes; for the k-tree model studied in this work we will also call the non-root nodes "inserted nodes"). Then, apart from the edges connecting the root nodes with each other, this induces a natural orientation on the edges. Thus, for each non-root node, we can distinguish between ingoing edges (coming from the direction of the root-clique) and outgoing edges, which also defines the in-degree $d^-(u)$ and the out-degree $d^+(u)$ of a node u; for a root node we will only define the out-degree. It is immediate from the definition that each non-root node u has exactly k ingoing edges, and these edges connect u with a k-clique $K = \{w_1, \dots, w_k\}$. We might then say that u is a child of the k-clique K or that u is attached to K and that $w_1, \dots, w_k$ are the parents of u. For the degree $d(u)$ of a node u it holds that $d(u) = d^+(u) + k$ for a non-root node and $d(u) = d^+(u) + k - 1$ for a root node. We also define the out-degree $d^+(K)$ of a k-clique K as the number of children of K.

Unlike in previously considered k-tree models it is for our purpose important to introduce ordered k-trees, i.e., we assume that the children of each k-clique are linearly ordered (thus one might speak about the first, second, etc. child of a k-clique). Furthermore, we introduce specific labellings of the nodes of ordered k-trees, which might be called increasing labellings (in analogy to the corresponding term for trees, see, e.g., [FBS92]). Given an ordered k-tree with n non-root nodes we label the set of root nodes by $\{0_1, 0_2, \dots, 0_k\}$, whereas the non-root nodes are labelled by $\{1, 2, \dots, n\}$ in such a way that the label of a node is always larger than the labels of all its parent nodes (of course, in this context the value of 0_ℓ, $1 \le \ell \le k$, is defined as 0). The graph family "ordered increasing k-trees" can then be described combinatorially as the family of all increasingly labelled ordered k-trees. It is apparent from the definition that for $k = 1$ one gets the tree family of plane-oriented recursive trees. In what follows we will often use as an abbreviation the term k-tree without further specification, but the meaning should always be "ordered increasing k-tree". Furthermore, we will often identify a node with its label, so node j always has the meaning of "the node labelled by j".

Throughout this paper we use the convention that the size $|T|$ of a k-tree T is given by the number of non-root nodes; thus the k-tree consisting only of the root-clique $K_0 = \{0_1, \dots, 0_k\}$ has size 0. Let T_n denote the number of ordered increasing k-trees of size n (we do not explicitly express the dependence on k, which is of course given). Obviously it holds $T_0 = T_1 = 1$. To get an enumeration formula for T_n we observe that when inserting a node into a k-tree this always increases the number of possible ways of attaching a further node by $k + 1$ (k due to the newly generated k-cliques and a further one due to a new available position at the parent k-clique). Thus there are always $1 + (k + 1)(n - 1)$ possible ways of

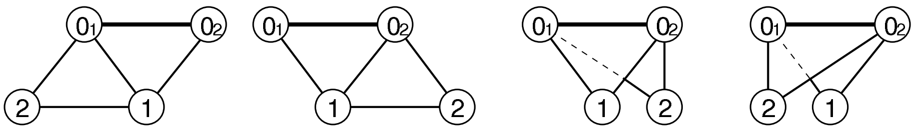

Fig. 1: All 4 different 2-trees of size 2. In the third and the fourth 2-tree in the picture the linear order on the children 1 and 2 of the root-clique is expressed by drawing 1 in front of 2 or vice versa.

inserting node n into a k-tree of size $n-1$. Since each k-tree of size n is uniquely obtained from a k-tree of size $n-1$ and inserting node n in a possible way it holds that $T_n = (1+(k+1)(n-1))T_{n-1}$, which shows that the number of different ordered increasing k-trees of size n is given by

$$T_n = \prod_{\ell=0}^{n-1} \big(1+(k+1)\ell\big) = n!(k+1)^n \binom{n-\frac{k}{k+1}}{n}, \quad \text{for } n \geq 0. \tag{1}$$

In Figure 1 we give all 4 different 2-trees of size 2.

When studying parameters in ordered increasing k-trees we always assume the "random ordered increasing k-tree model", which means that we assume that each of the T_n ordered increasing k-trees of size n appears with the same probability. It is easily seen that for this model the k-trees can be obtained indeed by the probabilistic preferential-attachment growth rule figured out in the introduction. One just has to take in mind that when a k-clique K in a k-tree has ℓ children, i.e., $d^+(K) = \ell$, then there are always exactly $\ell+1$ possible ways of attaching a new node to K, namely as the first child, second child, $\dots$, $(\ell+1)$-th child. Thus the following evolution process generates ordered increasing k-trees uniformly at random:

- Step 0: start with the root clique labelled by $0_1, 0_2, \dots, 0_k$.
- Step n: the node with label n is attached to any k-clique K in the already grown k-tree of size $n-1$ with a probability $p(K)$ given by

$$p(K) = \frac{d^+(K)+1}{1+(k+1)(n-1)}.$$

We will also use the combinatorial decomposition of ordered increasing k-trees with respect to the root-clique. To describe this decomposition it is advantagous to introduce two families $\mathcal{T}$ and $\mathcal{S}$ of combinatorial objects (they depend on k, but we do not explicitly express this). $\mathcal{T}$ is just the family of ordered increasing k-trees, whereas $\mathcal{S}$ consists of all ordered increasing k-trees, where the root-clique has exactly one child. Of course, an object of $\mathcal{T}$, where the root-clique has exactly ℓ children, can be obtained, after identification of the root nodes and an order-preserving relabelling, by a sequence of ℓ objects of $\mathcal{S}$. Furthermore, when considering objects in $\mathcal{S}$ the child-node attached to the root-clique has to be labelled by 1, and together with all choices of $k-1$ nodes from the root-clique it is forming exactly k different k-cliques, which, after relabelling, can themselves be considered as root-cliques of objects of $\mathcal{T}$. Thus we obtain the following formal description of the families $\mathcal{T}$ and $\mathcal{S}$ (see, e.g., [FS09] for an explanation of such formal specifications):

$$\begin{aligned} \mathcal{T} &= \mathcal{S}^0 \,\dot{\cup}\, \mathcal{S}^1 \,\dot{\cup}\, \mathcal{S}^2 \,\dot{\cup}\, \mathcal{S}^3 \,\dot{\cup}\, \cdots \;=\; \textsc{Seq}(\mathcal{S}), \\ \mathcal{S} &= \{1\} \times \mathcal{T}^k = \mathcal{Z}^{\Box} * \mathcal{T}^k. \end{aligned} \tag{2}$$

When denoting by T_n and S_n the number of objects in the families $\mathcal{T}$ and $\mathcal{S}$, respecively, of size n and by $T(z) := \sum_{n\geq 0} T_n \frac{z^n}{n!}$ and $S(z) := \sum_{n\geq 0} S_n \frac{z^n}{n!}$ their exponential generating functions, we obtain by using the symbolic method (see again, e.g., [FS09]) immediately the following system of equations:

$$T(z) = \frac{1}{1 - S(z)}, \quad S'(z) = T(z)^k, \quad S(0) = 0,$$

which has the solution

$$T(z) = \big(1 - (k+1)z\big)^{-\frac{1}{k+1}} \quad \text{and} \quad S(z) = 1 - \big(1 - (k+1)z\big)^{\frac{1}{k+1}}. \tag{3}$$

Extracting coefficients from $T(z)$ and $S(z)$ shows again that the number T_n of ordered increasing k-trees of size n is given by (1), whereas $S_n = (n-1)!(k+1)^{n-1}\binom{n-1-\frac{1}{k+1}}{n-1}$, $n \geq 1$.

3 Parameters studied and results

3.1 Parameters studied

Next we give a definition of the quantities studied in random ordered increasing k-trees. For better readability we do not explicitly express the dependence of the quantities on k, which is of course always given, in the notations.

The r.v. $Y_{n,j}$ counts the out-degree (see Section 2) of node j in a random k-tree of size n, whereas the r.v. $\bar{Y}_n$ counts the out-degree of a random inserted node in a random k-tree of size n, i.e., the out-degree of a node picked at random from the non-root nodes $\{1, 2, \dots, n\}$. The r.v. $Y_{n,0}$ counts the out-degree of the root-node 0_1 in a random k-tree of size n; of course, due to symmetry, the corresponding r.v. are identically distributed for each of the root-nodes $0_1, \dots, 0_k$ and do not have to be introduced separately.

The r.v. $C_{n,j}$ counts the local clustering coefficient of node j in a random k-tree of size n. The local clustering coefficient has been introduced by [WS98] and is considered as an important parameter in the study of real-world networks. The local clustering coefficient $C_G(u)$ of a node u in a graph $G(V, E)$ is defined as the proportion of edges between neighbours of u divided by the number of edges between the neighbours that could possibly exist; formally $C_G(u)$ is given by

$$C_G(u) = \begin{cases} \frac{|\{e\in E\,:\, e=(x,y) \text{ with } x,y\in N(u)\}|}{\binom{d(u)}{2}}, & \text{if } d(u) \geq 2, \\ 0, \quad \text{if } d(u) = 0 \text{ or } d(u) = 1, \end{cases} \tag{4}$$

where $N(u)$ denotes the set of neighbours (i.e., adjacent nodes) of u. The r.v. $\bar{C}_n$ counts the local clustering coefficient of a randomly selected node (amongst the root nodes and the inserted nodes) in a random k-tree of size n.

The r.v. $X_{n,j}$ counts the number of descendants of node j in a random k-tree of size n. Whether a node w is a descendant of u might be defined recursively: w is a descendant of u if either $u = w$ or if w has a parent node, which is a descendant of u. One might then also say that u is an ancestor of w. The r.v. $\bar{X}_n$ counts the number of descendants of a random inserted node in a random k-tree of size n.

The r.v. D_n counts the distance between the root node 0_1 and node n in a random k-tree of size n. As usual the distance between nodes in a graph is measured by the minimal number of edges contained in a path amongst all paths connecting these nodes. Again the corresponding r.v. are identically distributed

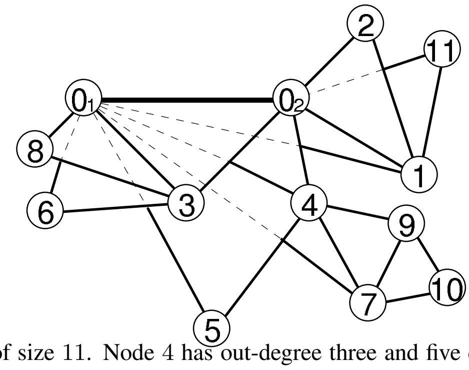

Fig. 2: An example of a 2-tree of size 11. Node 4 has out-degree three and five descendants (counting the node as a descendant of itself). The local clustering coefficient of node 4 is 0.4, since there are four edges between the five neighbours of this node (see the definition of the local clustering coefficient). The distance of node 4 to the root node 0_1 is one.

for each of the root-nodes $0_1, \dots, 0_k$. Furthermore, as a direct consequence of the evolution process of random k-trees one obtains that the distance $D_{n,j}$ between the root node 0_1 and node j in a random k-tree of size n is distributed as D_j and thus does not have to be studied separately. The r.v. $\bar{D}_n$ counts the distance between the root node 0_1 and a random inserted node in a random k-tree of size n.

In Figure 2 we give an example of a 2-tree together with the parameters studied.

3.2 Results

Degree of the nodes

Theorem 1 *The r.v. $Y_{n,j}$, which counts the out-degree of node j in a random k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{Y_{n,j} = m\} = \frac{\binom{j-\frac{k}{k+1}}{j}}{\binom{n-\frac{k}{k+1}}{n}\binom{n}{j}} \sum_{\ell=0}^{m} \binom{m}{\ell} (-1)^{\ell} \binom{n-\frac{k(2+\ell)}{k+1}}{n-j}, \quad \text{for } n \geq j \geq 1 \text{ and } m \geq 0.$$

The limiting distribution behaviour of $Y_{n,j}$ is, for $n \to \infty$ and depending on the growth of j, characterized as follows.

- *The region for j fixed. The normalized random variable $n^{-\frac{k}{k+1}} Y_{n,j}$ converges in distribution to a r.v. Y_j, i.e., $n^{-\frac{k}{k+1}} Y_{n,j} \xrightarrow{(d)} Y_j$, which is fully characterized by its moments. The s-th moments of Y_j are, for $s \geq 0$, given by*
$$\mathbb{E}(Y_j^s) = \frac{s!\Gamma(j + \frac{1}{k+1})}{\Gamma(j + 1 + \frac{k(s-1)}{k+1})}.$$
- *The region for j small: $j \to \infty$ such that $j = o(n)$. The normalized random variable $\left(\frac{j}{n}\right)^{\frac{k}{k+1}} Y_{n,j}$ is asymptotically exponentially distributed with parameter 1:*
$\left(\frac{j}{n}\right)^{\frac{k}{k+1}} Y_{n,j} \xrightarrow{(d)} \mathrm{Exp}(1)$, i.e., $\left(\frac{j}{n}\right)^{\frac{k}{k+1}} Y_{n,j} \xrightarrow{(d)} Y$, where the s-th moments of Y are, for $s \geq 0$, given by
$$\mathbb{E}(Y^s) = s!.$$

- *The central region for j: $j \to \infty$ such that $j \sim \rho n$, with $0 < \rho < 1$. The random variable $Y_{n,j}$ is asymptotically geometrically distributed with success probability $\rho^{\frac{k}{k+1}}$:*
  $Y_{n,j} \xrightarrow{(d)} \mathrm{Geom}(\rho^{\frac{k}{k+1}})$*, i.e.,* $Y_{n,j} \xrightarrow{(d)} Y_\rho$*, where the probability mass function of Y_ρ is given by*

$$\mathbb{P}\{Y_\rho = m\} = \rho^{\frac{k}{k+1}}\big(1 - \rho^{\frac{k}{k+1}}\big)^m, \quad \textit{for } m \geq 0.$$

- *The region for j large: $j \to \infty$ such that $\tilde{j} := n - j = o(n)$. It holds that $\mathbb{P}\{Y_{n,j} = 0\} \to 1$.*

Theorem 2 *The r.v. $Y_{n,0}$, which counts the out-degree of the root node 0_1 in a random k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{\bar{Y}_{n,0} = m\} = \frac{\binom{m-\frac{k-1}{k}}{m}}{\binom{n-\frac{k}{k+1}}{n}} \sum_{\ell=0}^{m} \binom{m}{\ell}(-1)^\ell \binom{n-1-\frac{k\ell}{k+1}}{n}, \quad \textit{for } m \geq 0.$$

For $n \to \infty$, the normalized random variable $n^{-\frac{k}{k+1}} Y_{n,0}$ converges in distribution to a r.v. Y_0, i.e., $n^{-\frac{k}{k+1}} Y_{n,0} \xrightarrow{(d)} Y_0$, which is fully characterized by its moments. The s-th moments of Y_0 are, for $s \geq 0$, given by

$$\mathbb{E}(Y_0^s) = \frac{\Gamma(\frac{1}{k+1})\Gamma(s+\frac{1}{k})}{\Gamma(\frac{1}{k})\Gamma(\frac{k}{k+1}s + \frac{1}{k+1})}.$$

Theorem 3 *The r.v. $\bar{Y}_n$, which counts the out-degree of a random inserted node in a random k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{\bar{Y}_n = m\} = \frac{1}{n\binom{n-\frac{k}{k+1}}{n}} \sum_{\ell=0}^{m} \frac{\binom{m}{\ell}(-1)^\ell}{k(\ell+2)+1} \left(\binom{n+\frac{1}{k+1}}{n} - \binom{n-\frac{k(\ell+2)}{k+1}}{n} \right), \quad \textit{for } m \geq 0.$$

For $n \to \infty$, $\bar{Y}_n$ converges in distribution to a discrete r.v. $\bar{Y}$, i.e., $\bar{Y}_n \xrightarrow{(d)} \bar{Y}$, with

$$\mathbb{P}\{\bar{Y} = m\} = p_m := \frac{k+1}{k(m+1)\binom{m+2+\frac{1}{k}}{m+1}}, \quad \textit{for } m \geq 0.$$

Since $p_m \sim \frac{k+1}{k}\Gamma\big(2+\frac{1}{k}\big)m^{-2-\frac{1}{k}}$, for $m \to \infty$, it follows that $\bar{Y}_n$ follows asymptotically a power-law distribution with exponent $2 + \frac{1}{k}$.

Local clustering coefficient

Lemma 1 *For any k-tree T the local clustering coefficient $C_T(u)$ of a node u only depends on the degree $d(u)$ of u. It holds then for $d(u) \geq k \geq 2$:*

$$C_T(u) = \frac{2(k-1)}{d(u)} - \frac{(k-1)(k-2)}{d(u)(d(u)-1)}.$$

k	$c_k = \lim_{n\to\infty} \mathbb{E}(\bar{C}_n)$	numerically	k	$c_k = \lim_{n\to\infty} \mathbb{E}(\bar{C}_n)$	numerically
2	$23 - \frac{9}{4}\pi^2$	$0.793390\ldots$	6	$\frac{148003}{57024} - \frac{2695}{62208}\Psi'(-\frac{11}{6})$	$0.903449\ldots$
3	$-5 + \frac{16}{3}\Psi'(\frac{4}{3})$	$0.843184\ldots$	10		$0.933975\ldots$
4	$\frac{1051}{96} - \frac{75}{128}\Psi'(\frac{1}{4})$	$0.871356\ldots$	50		$0.982804\ldots$
5	$\frac{512}{125} - \frac{72}{625}\Psi'(-\frac{4}{5})$	$0.889998\ldots$	100		$0.990885\ldots$

Tab. 1: The limit c_k of the expected local clustering coefficient $\mathbb{E}(\bar{C}_n)$ for small values of k.

Theorem 4 *Let the r.v. $\bar{C}_n$ count the local clustering coefficient of a random node in a random k-tree of size n. Then the expected local clustering coefficient $\mathbb{E}(\bar{C}_n)$ behaves, for $n \to \infty$, as follows (here $\Psi(x) = (\ln\Gamma(x))'$ denotes the Psi-function and $\Psi'(x)$ its derivative):*

$$\begin{aligned}
\mathbb{E}(\bar{C}_n) \to c_k &:= \sum_{m\ge 0} \frac{(k+1)(k-1)}{k(m+1)(m+k)\binom{m+2+\frac{1}{k}}{m+1}}\Big(2 - \frac{k-2}{m+k-1}\Big) \\
&= (k+1)\Bigg(\frac{1}{k-1} + \frac{3}{k}\binom{k-4-\frac{1}{k}}{k-2}\sum_{\ell=1}^{k-2}\frac{1}{\ell^2\binom{\ell+3+\frac{1}{k}-k}{\ell}} + \frac{3}{k}\sum_{\ell=0}^{k-3}\frac{(-1)^\ell}{(\ell+1)(\ell-1-\frac{1}{k})\binom{k-3}{\ell}} \\
&\qquad - \frac{3}{k}\binom{k-4-\frac{1}{k}}{k-2}\Psi'\big(4+\frac{1}{k}-k\big)\Bigg).
\end{aligned}$$

It further holds that $c_k \to 1$, for $k \to \infty$.

Number of descendants

Theorem 5 *The r.v. $X_{n,j}$, which counts the number of descendants of node j in a random k-tree of size n, has the following exact distribution:*

$$\mathbb{P}\{X_{n,j} = m\} = \frac{\binom{m-1-\frac{1}{k+1}}{m-1}\binom{n-m-1+\frac{2}{k+1}}{n-m-j+1}}{\binom{n-\frac{k}{k+1}}{n-j}}, \quad \text{for } n \ge j \ge 1 \text{ and } m \ge 1.$$

The limiting distribution behaviour of $X_{n,j}$ is, for $n \to \infty$ and depending on the growth of j, characterized as follows.

- *The region for j fixed. The normalized random variable $\frac{X_{n,j}}{n}$ is asymptotically Beta-distributed, $\frac{X_{n,j}}{n} \xrightarrow{(d)} \beta(\frac{k}{k+1}, j-1+\frac{2}{k+1})$, i.e., $\frac{X_{n,j}}{n} \xrightarrow{(d)} X_j$, where the s-th moments of X_j are, for $s \ge 0$, given by*

$$\mathbb{E}(X_j^s) = \frac{\binom{s-\frac{1}{k+1}}{s}}{\binom{s+j-\frac{k}{k+1}}{s}}.$$

- *The region for j small: $j \to \infty$ such that $j = o(n)$. The normalized random variable $\frac{j}{n}X_{n,j}$ is asymptotically Gamma-distributed, $\frac{j}{n}X_{n,j} \xrightarrow{(d)} \gamma(\frac{k}{k+1}, 1)$, i.e., $\frac{j}{n}X_{n,j} \xrightarrow{(d)} X$, where the s-th moments of X are, for $s \geq 0$, given by*
$$\mathbb{E}(X^s) = s!\binom{s-\frac{1}{k+1}}{s}.$$
- *The central region for j: $j \to \infty$ such that $j \sim \rho n$, with $0 < \rho < 1$. The shifted random variable $X_{n,j} - 1$ is asymptotically negative binomial-distributed, $X_{n,j} - 1 \xrightarrow{(d)} \mathrm{NegBin}(\frac{k}{k+1}, \rho)$, i.e., $X_{n,j} - 1 \xrightarrow{(d)} X_\rho$, where the probability mass function of X_ρ is given by*
$$\mathbb{P}\{X_\rho = m\} = \binom{m-\frac{1}{k+1}}{m}\rho^{\frac{k}{k+1}}(1-\rho)^m, \quad \textit{for } m \geq 0.$$
- *The region for j large: $j \to \infty$ such that $\tilde{j} := n - j = o(n)$. It holds that $\mathbb{P}\{X_{n,j} = 1\} \to 1$.*

Theorem 6 *The r.v. $\bar{X}_n$, which counts the number of descendants of a random inserted node in a random k-tree of size n, has the following exact distribution (with $m \geq 1$):*
$$\mathbb{P}\{\bar{X}_n = m\} = \frac{\binom{m-1-\frac{1}{k+1}}{m-1}}{n\binom{n-\frac{k}{k+1}}{n}}\sum_{\ell=0}^{m-1}\frac{\binom{m-1}{\ell}(-1)^\ell}{(k+1)(\ell+1)+k}\left(\binom{n+\frac{1}{k+1}}{n} - \binom{n-\ell-2+\frac{2}{k+1}}{n}\right).$$
For $n \to \infty$, $\bar{X}_n$ converges in distribution to a discrete r.v. $\bar{X}$, i.e., $\bar{X}_n \xrightarrow{(d)} \bar{X}$, with
$$\mathbb{P}\{\bar{X} = m\} = \frac{k}{(k+1)(m+\frac{k}{k+1})(m-\frac{1}{k+1})}, \quad \textit{for } m \geq 1.$$

Root-to-node distance

Theorem 7 *The r.v. D_n, which measures the distance between the root node 0_1 and node n in a random k-tree of size n, is, for $n \to \infty$, asymptotically Gaussian distributed, where the rate of convergence is of order $\mathcal{O}\big(\frac{1}{\sqrt{\log n}}\big)$:*
$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{D_n - \mathbb{E}(D_n)}{\sqrt{\mathbb{V}(D_n)}} \leq x\right\} - \Phi(x)\right| = \mathcal{O}\Big(\frac{1}{\sqrt{\log n}}\Big),$$
and the expectation $\mathbb{E}(D_n)$ and the variance $\mathbb{V}(D_n)$ satisfy
$$\mathbb{E}(D_n) = \frac{1}{(k+1)H_k}\log n + \mathcal{O}(1), \quad \mathbb{V}(D_n) = \frac{H_k^{(2)}}{(k+1)H_k^3}\log n + \mathcal{O}(1).$$

Corollary 1 *The r.v. $\bar{D}_n$, which measures the distance between the root node 0_1 and a random inserted node in a random k-tree of size n, is, for $n \to \infty$, asymptotically Gaussian distributed: $\mathbb{P}\left\{\frac{\bar{D}_n - \mathbb{E}(\bar{D}_n)}{\sqrt{\mathbb{V}(\bar{D}_n)}} \leq x\right\} \to \Phi(x)$, for all $x \in \mathbb{R}$, with expectation and variance satisfying $\mathbb{E}(\bar{D}_n) = \frac{1}{(k+1)H_k}\log n + \mathcal{O}(1)$ and $\mathbb{V}(\bar{D}_n) = \frac{H_k^{(2)}}{(k+1)H_k^3}\log n + \mathcal{O}(1)$.*

Here $\Phi(x)$ denotes the distribution function of the standard normal distribution $\mathcal{N}(0,1)$.

4 Brief outline of the proof of the results

4.1 Degree of the nodes

In order to get a suitable description of the r.v. $Y_{n,j}$ we consider the graph evolution process of k-trees. The following observation is crucial to our approach: each node x attached to node j increases the number of possibilities of attaching a new node to j by exactly k (1 possibility more at the k-clique where node x is attached and $k-1$ possibilities more due to the $k-1$ new k-cliques containing x and j). Thus if node $j \geq 1$ has out-degree m there are exactly $(m+1)k$ possibilities of attaching a new node that increases the out-degree of node j, whereas the remaining possibilities will keep the out-degree unchanged.

Thus if we count by $T_{n,j,m} := T_n \mathbb{P}\{Y_{n,j} = m\}$ the number of k-trees of size n such that node j has out-degree m, we immediately get the following recurrence:

$$T_{n,j,m} = \big((k+1)n - km - 2k\big)T_{n-1,j,m} + kmT_{n-1,j,m-1}, \quad \text{for } n > j \geq 1 \text{ and } m \geq 0,$$

with $T_{j,j,0} = T_j$, for $j \geq 1$, and $T_{j,j,m} = 0$, for $m > 0$. Introducing the generating functions $T^{[j]}(z,v) := \sum_{n \geq j} \sum_{m \geq 0} T_{n,j,m} \frac{z^{n-j}}{(n-j)!} v^m$ leads to the following linear first order partial differential equation:

$$\big(1-(k+1)z\big)T_z^{[j]}(z,v) + kv(1-v)T_v^{[j]}(z,v) - \big(k(j-1+v)+j+1\big)T^{[j]}(z,v) = 0, \quad T^{[j]}(0,v) = T_j,$$

which can be solved by applying the method of characteristics. The solution is given by the following expression:

$$T^{[j]}(z,v) = \frac{T_j}{\Big(1 - v\big(1 - (1-(k+1)z)^{\frac{k}{k+1}}\big)\Big)\big(1-(k+1)z\big)^{\frac{kj-k+j+1}{k+1}}},$$

and extracting coefficients immediately shows the exact formula for the probabilities $\mathbb{P}\{Y_{n,j} = m\}$ given in Theorem 1. To show the limiting distribution results given in Theorem 1 we use, depending on the growth behaviour of $j = j(n)$, different approaches (see [KP07] for similar considerations on the node-degree of increasing trees). For the two cases j fixed and $j \to \infty$, such that $j = o(n)$, we use the method of moments, where we study the explicit expression for the s-th factorial moments obtained after extracting coefficients from the s-th derivative of $T^{[j]}(z,v)$ w.r.t. v evaluated at $v = 1$. For the remaining two cases $j \sim \rho n$, with $0 < \rho < 1$, and $n - j = o(n)$ we directly study the exact expression for the probabilities.

To show Theorem 2 concerning the out-degree of the root node 0_1 one can use the same approach as for a non-root j, but one has to start with a slightly different recurrence. The asymptotic considerations are similar to the case j fixed.

For obtaining Theorem 3 one simply uses the relation $\mathbb{P}\{\bar{Y}_n = m\} = \frac{1}{n}\sum_{j=1}^{n} \mathbb{P}\{Y_{n,j} = m\}$ and Theorem 1. In order to get the explicit expression for the probabilities given in the theorem we use a hypergeometric identity for simplifying it.

4.2 Local clustering coefficient

The crucial observation for analyzing the local clustering coefficient in k-trees is that the local clustering coefficient $C_T(u)$ of a node u in a k-tree T only depends on the degree $d(u)$ of the corresponding node; the exact relation is expressed in Lemma 1. To show this we will, according to the definition (4), count the number $M(u)$ of edges between neighbours of u. Consider a node u in a k-tree; then it always holds that

$d(u) \geq k-1$. If $d(u) = k-1$ then the k-tree can consist only of a single root-clique and u is one of the root nodes; thus all $k-1$ neighbours of u are connected with each other, which implies $M(u) = \binom{k-1}{2}$. In order to determine $M(u)$ when $d(u) \geq k$ we observe that in any k-tree holds that when increasing the degree of a node u by 1 then the number of edges between neighbours of u increases exactly by $k-1$; this holds since a new node w adjacent to u generates a k-clique, such that w is also adjacent to $k-1$ neighbours of u. Thus $M(u) = \binom{k-1}{2} + (k-1)(d(u)-k+1)$, for $d(u) \geq k-1$, which implies Lemma 1.

Of course, due to Lemma 1, one can immediately obtain distributional relations between r.v. measuring the degree (or out-degree) and the local clustering coefficient of nodes in k-trees. In particular we are interested in the r.v. $\bar{C}_n$ measuring the local clustering coefficient of a random node in a random k-tree of size n (of course, similar considerations for the local clustering coefficient of node j can be made also, but we skip them here). One gets then

$$\bar{C}_n \stackrel{(d)}{=} \frac{2(k-1)}{\tilde{Y}_n} - \frac{(k-1)(k-2)}{\tilde{Y}_n(\tilde{Y}_n-1)},$$

where $\tilde{Y}_n$ measures the degree of a randomly selected node (amongst the root nodes and inserted nodes) in a k-tree of size n. Of course, the distribution of $\tilde{Y}_n$, and thus also the distribution of $\bar{C}_n$ is fully determined by the previously studied r.v. $\bar{Y}_n$ and $Y_{n,0}$. In particular it easily follows that $\tilde{Y}_n \xrightarrow{(d)} \bar{Y}+k$, where the distribution of the discrete random variable $\bar{Y}$ is characterized in Theorem 3. The main quantity of interest in this context is the expected local clustering coefficient of a random node. Since $\tilde{Y}_n \xrightarrow{(d)} \bar{Y}+k$, with $\bar{Y}$ a discrete r.v., and since the function $f(m) = \frac{2(k-1)}{m} - \frac{(k-1)(k-2)}{m(m-1)}$ is uniformly bounded for $m \geq k$, it immediately follows that

$$\mathbb{E}(\bar{C}_n) \to c_k := \sum_{m \geq k} \mathbb{P}\{\bar{Y}+k = m\}\Big(\frac{2(k-1)}{m} - \frac{(k-1)(k-2)}{m(m-1)}\Big),$$

which leads to the first expression for c_k given in Theorem 4. The second one, which is advantageous when computing c_k for small k, can be obtained by rather lengthy manipulations with beta integrals and their derivatives.

4.3 Number of descendants

For a recursive description of the r.v. $X_{n,j}$ we consider the graph evolution process of k-trees. Here the following observation is crucial: each node x attached to a descendant of node j increases the number of possibilities of attaching a new node to a descendant of j by exactly $k+1$ (1 possibility more at the k-clique where node x is attached and k possibilities more due to the k new k-cliques containing x). Thus if node $j \geq 1$ has m descendants there are exactly $(k+1)m-1$ possibilities of attaching a new node that increases the number of descendants of node j, whereas the remaining possibilities will keep the number of descendants unchanged.

This description allows a recursive approach analogeous to the one sketched in Subsection 4.1. Also the asymptotic considerations are very similar to the ones discussed there, but somewhat simpler, due to the closed formulæ for the exact results.

4.4 Root-to-node distance

In order to study the distance between node n and the root node 0_1 in a k-tree it is natural to study the distance between node n and all root nodes $0_1, \ldots, 0_k$ simultaneously. To do this we first introduce the notion dist(u, K), which gives the distance between a node u and a k-clique $K = \{w_1, \dots, w_k\}$ via dist$(u, K) := \min_{1 \le \ell \le k}\{\text{dist}(u, w_\ell)\}$ (where dist(x, y) denotes the distance between the nodes x and y). We introduce then the r.v. $\tilde{D}_n$, which counts the distance between node n and the root-clique $K_0 = \{0_1, \dots, 0_k\}$ in a random k-tree of size n. Since the distance between node n and an arbitrary root node in a k-tree is always either the same as the distance between n and the root-clique K_0 or one more, there are always ℓ root nodes, with $1 \le \ell \le k$, which are at the same distance to node n like the root-clique is, and $k - \ell$ root nodes with a distance one larger. Due to symmetry it suffices to introduce the following k different events $\mathcal{E}_\ell$, $1 \le \ell \le k$, which describe the different situations that can occur:

$$\mathcal{E}_\ell := \big\{\text{dist}(n, 0_1) = \cdots = \text{dist}(n, 0_\ell) < \text{dist}(n, 0_{\ell+1}) = \cdots = \text{dist}(n, 0_k)\big\}.$$

Then the distribution of the r.v. D_n we are interested in can be described as follows (amongst the $\binom{k}{\ell}$ possible situations symmetric to event $\mathcal{E}_\ell$ one has to distinguish whether node 0_1 is at the same distance to n or is one larger than the distance between n and the root-clique):

$$\mathbb{P}\{D_n = m\} = \sum_{\ell=1}^{k} \mathbb{P}\{\tilde{D}_n = m \wedge \mathcal{E}_\ell \text{ occurs}\}\binom{k-1}{\ell-1} + \sum_{\ell=1}^{k} \mathbb{P}\{\tilde{D}_n = m-1 \wedge \mathcal{E}_\ell \text{ occurs}\}\binom{k-1}{\ell}. \quad (5)$$

When introducing the generating functions

$$T_\ell(z, v) := \sum_{n\ge1}\sum_{m\ge0} T_n \mathbb{P}\{\tilde{D}_n = m \ \wedge \ \mathcal{E}_\ell \text{ occurs}\}\frac{z^{n-1}}{(n-1)!}v^m, \quad 1 \le \ell \le k,$$
$$S_\ell(z, v) := \sum_{n\ge1}\sum_{m\ge0} S_n \mathbb{P}\{\tilde{D}_n^{[S]} = m \ \wedge \ \mathcal{E}_\ell \text{ occurs}\}\frac{z^{n-1}}{(n-1)!}v^m, \quad 1 \le \ell \le k,$$

where $\tilde{D}_n^{[S]}$ denotes the corresponding r.v. for objects in the family $\mathcal{S}$, one obtains by using the combinatorial decomposition of k-trees w.r.t. the root-clique (and after a study of the possibilities for the distance between node 1 and node n in objects of $\mathcal{S}$ leading to event $\mathcal{E}_\ell$) given by (2) the following system of equations, with $T(z)$ and $S(z)$ given in (3):

$$T_\ell(z, v) = \frac{S_\ell(z, v)}{(1 - S(z))^2}, \quad \frac{\partial}{\partial z}S_\ell(z, v) = (k - \ell)T(z)^{k-1}\big(T_\ell(z, v) + T_{\ell+1}(z, v)\big), \ \ 1 \le \ell \le k - 1,$$
$$T_k(z, v) = \frac{S_k(z, v)}{(1 - S(z))^2}, \quad \frac{\partial}{\partial z}S_k(z, v) = kvT(z)^{k-1}T_1(z, v).$$

This leads to the following system of linear differential equations for the functions $S_\ell(z, v)$:

$$\frac{\partial}{\partial z}S_\ell(z, v) = \frac{(k - \ell)\big(S_\ell(z, v) + S_{\ell+1}(z, v)\big)}{1 - (k + 1)z}, \quad 1 \le \ell \le k - 1, \quad \frac{\partial}{\partial z}S_k(z, v) = \frac{kvS_1(z, v)}{1 - (k + 1)z}.$$

Since it is possible to get from this system of differential equations a single differential equation for $S_k(z, v)$, which is of Euler type, it can be solved explicitly; thus all functions $T_\ell(z, v)$, $1 \le \ell \le k$, can also be given explicitly.

Since the generating function $T(z,v) := \sum_{n\ge 1}\sum_{m\ge 0} T_n \mathbb{P}\{D_n = m\}\frac{z^{n-1}}{(n-1)!}v^m$ is due to equation (5) completely determined by the functions $T_\ell(z,v)$, $1 \le \ell \le k$, via

$$T(z,v) = \sum_{\ell=1}^{k} \left(\binom{k-1}{\ell-1} + \binom{k-1}{\ell} v \right) T_\ell(z,v),$$

it is possible to also get an explicit solution for $T(z,v)$. One eventually obtains that

$$T(z,v) = \sum_{j=1}^{k} \frac{B_j(v)}{(1-(k+1)z)^{\alpha_j(v)+\frac{2}{k+1}}}, \tag{6}$$

with $\alpha_1(v), \dots, \alpha_k(v)$ the roots of the equation

$$\frac{k!v}{(k+1)^k} = \prod_{r=0}^{k-1} \left(\alpha - \frac{r}{k+1}\right), \tag{7}$$

and where the functions $B_j(v)$, $1 \le j \le k$, can be given explicitly (see [Pan04] for such considerations on a related problem); for our purpose it is sufficient to state that all functions $B_j(v)$ are analytic in a neighbourhood of $v = 1$. By considerations as in [Mah92] one can show that for $v = 1$ all roots $\alpha_j(1)$, $1 \le j \le k$, of (7) are simple. It is easily observed that $\frac{k}{k+1}$ is a root of (7) when $v = 1$; moreover, it is the root with largest real part. Let $\alpha_1(v)$ denote the root of (7), which satisfies $\alpha_1(1) = \frac{k}{k+1}$. Then from (6) we obtain the following asymptotic expansion of the moment generating function of D_n:

$$\mathbb{E}\big(e^{D_n s}\big) = e^{(\alpha_1(e^s)-\frac{k}{k+1})\log n + \log\left(\frac{\Gamma(1+\frac{1}{k+1})B_1(e^s)}{\Gamma(\alpha_1(e^s)+\frac{2}{k+1})}\right)} \cdot \big(1+\mathcal{O}(n^{-\eta})\big), \quad \text{with an } \eta > 0.$$

An application of the quasi-power theorem of Hwang, see [HKH98], immediately shows Theorem 7. Corollary 1 can be deduced from it easily.

5 Conclusion

We introduced a network model which is based on k-trees and which can either be described by a probabilistic growth rule using preferential attachment or combinatorially by considering increasing labellings of the nodes and a linear ordering of the children of k-cliques. We gave a precise analysis of various parameters and could show that the distribution of the node-degrees follows asymptotically a power law, that the expected local clustering coefficient is high, and that the root-to-node distance of node n is asymptotically Gaussian distributed with expectation and variance of order $\log n$. The approach we used is not restricted to the introduced model, but can also be easily applied to further evolving k-tree models; in particular the previously introduced uniform attachment model for k-trees can be treated in the same way. But also further k-tree models such as, e.g., ones with a "saturation rule", where at most up to d children can be attached to a k-clique and where the probability that a new node is attached is proportional to the number of "free places", can be introduced and analyzed (the special instance $d = 1$ gives the so-called Apollonian networks). One can even go a step further and introduce weighted ordered k-trees (with or without increasing labellings), where each k-clique in the k-tree gets a weight depending on the number of attached children. This, in analogy to simply generated tree families, see [FS09], leads then to "simple families of k-trees" (if unlabelled or arbitrarily labelled) or "simple families of increasing k-trees" (if increasingly labelled). By choosing specific weights for increasingly labelled ordered k-trees all the before-mentioned evolution models for k-trees can be obtained; a complete characterization of possible k-tree evolution models as has been given in [PP07] for simple families of increasing trees is possible.

References

[ADS10] O. Bodini A. Darrasse, H.-K. Hwang and M. Soria. The connectivity-profile of random increasing k-trees. In *Proceedings of ANALCO'10: Workshop on Analytic Algorithmics and Combinatorics*, 2010. to appear.

[BA99] A.-L. Bárabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[BP69] L. W. Beineke and R. E. Pippert. The number of labeled k-dimensional trees. *Journal of Combinatorial Theory*, 6:200–205, 1969.

[BR03] B. Bollobás and O. M. Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34, Weinheim, 2003. Wiley-VCH.

[DS09] A. Darrasse and M. Soria. Limiting distribution for distances in k-trees. In *Proceedings of the 20th International Workshop on Combinatorial Algorithms (IWOCA 2009)*, Lecture Notes in Computer Science. Springer-Verlag, 2009. to appear.

[FBS92] P. Flajolet F. Bergeron and B. Salvy. Varieties of increasing trees. *Lecture Notes in Computer Science*, 581:24–48, 1992.

[FS09] P. Flajolet and R. Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[Gao09] Y. Gao. The degree distribution of random k-trees. *Theoretical Computer Science*, 410:688–695, 2009.

[HKH98] H.-K-Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19:329–343, 1998.

[KP07] M. Kuba and A. Panholzer. On the degree distribution of the nodes in increasing trees. *Journal of Combinatorial Theory, Series A*, 114:597–618, 2007.

[Mah92] H. M. Mahmoud. *Evolution of random search trees*. John Wiley & Sons, New York, 1992.

[Moo69] J. W. Moon. The number of labeled k-trees. *Journal of Combinatorial Theory*, 6:196–199, 1969.

[MS95] H. M. Mahmoud and R. T. Smythe. A survey of recursive trees. *Theoretical Probability and Mathematical Statistics*, 51:1–37, 1995.

[Pan04] A. Panholzer. Distribution of the Steiner distance in generalized m-ary search trees. *Combinatorics, Probability & Computing*, 13:717–733, 2004.

[PP07] A. Panholzer and H. Prodinger. Level of nodes in increasing trees revisited. *Random Structures and Algorithms*, 31:203–226, 2007.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.