Characterizing several properties of high-dimensional random Apollonian networks

Panpan Zhang

January 23, 2019

Abstract. In this article, we investigate several properties of high-dimensional random Apollonian networks (HDRANs), including two types of degree profiles, the small-world effect (clustering property), sparsity, and several distance-based properties. The methods that we use to characterize the degree profiles are two-dimensional mathematical induction, analytic combinatorics, and Pólya urns, etc. The small-world property and sparsity are respectively measured by the local clustering coefficient and a proposed Gini index. Finally, we look into three distance-based properties, which are total depth, diameter and the Wiener index.

AMS subject classifications.
Primary: 90B15; 05C82
Secondary: 05C07; 05C12

Key words. Degree profile; distance; random Apollonian networks; small world; sparsity; topological index

1 Introduction

Due to the surge of interest in networks, such as the Internet (e.g., The Internet Mapping Project by Hal Burch and Bill Cheswick), resistor networks [40], the World Wide Web (WWW) [6], and social networks (e.g., friendship network [37]), a plethora of network models have been proposed and studied in the last several decades. In this paper, we study a network model that recently caught researchers’ attention—Apollonian networks (ANs). ANs arise from the problem of space-filling packing of spheres, proposed by the ancient Greek mathematician Apollonius of Perga. ANs are popular since this class of networks possesses various typical network characteristics, which are summarized in the title of [2]: scale free, small world, Euclidean, space filling and matching graphs. Each of these phrases is a significant area of modern network research itself.

The counterpart of an AN in the field of random network analysis is a Random Apollonian Network (RAN). The study of RAN first appeared in [55], where

1Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.
the power-law and the clustering coefficient were investigated. From then on, many more properties of RANs were uncovered by applied mathematicians and probabilists: The degree distribution was characterized by \cite{24}; the diameter was calculated by \cite{21,24}; the length of the longest path in RANs was determined by \cite{12,14,21}. All these resources, however, only focused on planar RANs, the evolution of which is based on continuing triangulation. Triangulated RANs are a special class of (general) RANs with network index 3, and they are maximal planar graphs according to the Kuratowski criterion \cite{34}.

Due to the increasing complexity of real-world networks, there is a high demand of research in high-dimensional networks. High-dimensional random Apollonian Networks (HDRANs) refer to RANs with a general network index $k \geq 3$. HDRANs were first introduced by \cite{54}, in which an iterative algorithm was designed to characterize several properties, including degree distribution, clustering coefficient and diameter. The exact degree distribution of a vertex with a fixed label and the total weight, a macro metric, were determined by \cite{52}. Other than these two resources, to the best of our knowledge, few work has been done for HDRANs.

The goal of this paper is to give a complete study of HDRANs. We survey and extend several known results of HDRANs, and uncover a couple of new properties of common interest. The rest of the paper is organized as follows. In Section 2, we briefly review the evolutionary process of HDRANs as well as some basic graph invariants thereof. In the next five sections, we investigate several properties of HDRANs. To make it clearer, we summarize these properties and associated methods in Table 1. In Section 8 we give some concluding remarks and propose some future work. For better readability, we present main results and short derivations in the main body of the article, but long proofs in the appendix.

2 Evolution of random Apollonian networks

In this section, we review the evolution of a RAN of index $k \geq 3$. At time $n = 0$, we start with a complete graph on $k$ vertices all of which are labeled with 0. At each subsequent time point $n \geq 1$, a $k$-clique is chosen uniformly at random among all active cliques in the network. A new vertex labeled with $n$ is linked by $k$ edges to all the vertices of the chosen clique. Then, the recruiting clique is deactivated. An explanatory example of a RAN with index $k = 5$ is given in Figure 1.

\footnote{In graph theory, a complete graph is a graph such that each pair of vertices therein is connected by an edge. A complete graph on $k$ vertices is also called $k$-clique or $k$-simplex. We shall interchangeably use these terms through the manuscript.}
According to the evolutionary process described above, we obtain some basic and deterministic graph invariants of a RAN with index $k$ at time $n$: the number of vertices $V_n^{(k)} = k + n$, the number of edges $E_n^{(k)} = k + nk$, and the number of active cliques $C_n^{(k)} = 1 + (k - 1)n$. We note that RANs of indices 1 and 2 are not considered in this paper, as their structure lacks research interest. A RAN of index 1 at time $n$ is a single vertex labeled with $n$, while a RAN of index 2 at time $n$ is a path of length $n$.

![Figure 1: An example of the evolution of a HDRAN of index 5 in two steps; active cliques are those containing at least one solid edge.](image)
3 Degree profile I

In this section, we investigate the degree profile of a RAN of index \( k \geq 3 \). The random variable of prime interest is \( X_{n,j}^{(k)} \), the number of vertices of degree \( j \) in a RAN of index \( k \) at time \( n \), for \( j \geq k \), where the boundary condition arises from the natural lower bound of the degree of vertices in RANs. It is also worthy of noting that the natural upper bound for \( j \) at time \( n \) is \( k + n - 1 \).

The degree random variable that we consider in this section is different from that investigated in [52], and the methods developed in [52] are not amenable to this study, which will be explained in detail in the sequel. To distinguish the two kinds of degree profiles, we call the one discussed in this section degree profile I. Specifically, we present two results of \( X_{n,j}^{(k)} \), which are respectively shown in Theorems 1 and 2. In Theorem 1, we prove that the difference between the expectation of \( X_{n,j}^{(k)} \) and a linear function of \( n \) is uniformly bounded, where the bound is determined. In Theorem 2, we show that \( X_{n,j}^{(k)} \) concentrates on its expectation with high probability, i.e., a focusing property.

Theorem 1 Let \( X_{n,j}^{(k)} \) be the number of vertices of degree \( j \) in a RAN of index \( k \) at time \( n \), for \( j \geq k \). For each \( n \in \mathbb{N} \) and any \( k \geq 3 \), there exists a constant \( b_{j,k} \) such that

\[
\left| \mathbb{E} \left[ X_{n,j}^{(k)} \right] - b_{j,k} n \right| \leq \frac{2k^2}{2k - 1}.
\]

In particular, we have \( b_{j,k} = \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k-1)} \).

The proof of Theorem 1 is based on an elementary mathematical tool—induction. As suggested in [24], we split the cases of \( j = k \) and \( j > k \) in the proof. For the case of \( j = k \), we apply the traditional mathematical induction directly, whereas we develop a two-dimensional induction based on an infinite triangular array for the case of \( j > k \). For the better readability of the paper, we present the major steps of the proof in Appendix A.

In the proof of Theorem 1, we show that the mean of \( X_{n,j}^{(k)} \) scaled by \( n \) converges to \( b_{j,k} \) when \( n \) is large. Let \( j \) go to infinity as well, we discover that \( b_{j,k} \sim j^{-k} \) according to the Stirling’s approximation. This implies that the degree distribution in HDRANs follows a power-law property, where the exponent is the network index \( k \). Consequently, HDRANs are scale-free networks. The power-law property for planar RANs (i.e., \( k = 3 \)) has been recovered in [55] numerically and in [24] analytically.

In addition, we are interested in the deviation of the random variable \( X_{n,j}^{(k)} \) from its expectation. In Theorem 2, we develop a Chebyshev-type inequality.

3 Upon joining into the network, every newcomer is connected with \( k \) existing vertices, leading to minimal possible degree \( k \).
**Theorem 2** Let \( X_{n,j}^{(k)} \) be the number of vertices of degree \( j \) in a RAN of index \( k \) at time \( n \), for \( j \geq k \). For any \( \lambda > 0 \), we have

\[
P \left( \left| X_{n,j}^{(k)} - \mathbb{E} \left[ X_{n,j}^{(k)} \right] \right| \geq \lambda \right) \leq e^{-\lambda^2/(8kn)}.
\]

The proof of Theorem 2 is presented in Appendix B. The main idea is to employ the *Azuma-Hoeffding inequality* [3] based on a martingale sequence. We remark that the exact same concentration result is found for *random k-trees* [26]. The author of [26] tackled the problem by using the methods from tree realization theory. The intrinsic reason of the identicality is due to the similarity in the evolutionary processes of HDRANs with index \( k \) and random \( k \)-trees.

Before ending this section, we would like to point out that the methods in the proofs of Theorems 1 and 2 are extended from the ideas in [24]. The results for planar RANs (a special case for \( k = 3 \)) can be found in [24, Theorem 1.1].

4 Degree profile II

Another type of degree profile that we look into is node-specified. Let \( D_{n,j}^{(k)} \) denote the degree of the node labeled with \( j \) in a HDRAN of index \( k \) at time \( n \). This property was investigated in [52], where the growth of HDRANs was represented by a two-color Pólya urn scheme [36]. Pólya urn appears to be an appropriate model since it successfully captures the evolutionary characteristics of highly dependent structures.

Noticing that the degree of a vertex is equal to the number of cliques incident with it, the authors of [52] introduced a color code such that the active cliques incident with the node labeled with \( j \) were colored white, while all the rest were colored blue. The associated urn scheme is governed by the *replacement matrix*

\[
\begin{pmatrix}
  k - 2 & 1 \\
  0 & k - 1
\end{pmatrix}.
\]

This replacement matrix is triangular, so the associated Pólya urn is called *triangular urn*. This class of urns has been extensively studied in [23, 32, 51]. The next proposition specifies the exact distribution of \( D_{j,n}^{(k)} \) as well as its moments.

**Proposition 1** Let \( D_{n,j}^{(k)} \) be the degree of the node labeled with \( j \) in a RAN of index \( k \) at time \( n \).
at time $n$, for $n \geq j$. The distribution of $D_{n,j}^{(k)}$ is given by

$$
P \left( D_{n,j}^{(k)} = k + \delta \right) = \frac{\Gamma(n - j + 1) \Gamma \left( j + \frac{1}{k-1} \right) \left( \delta + 2 \right)}{\Gamma \left( n + \frac{1}{k-1} \right)} \left( \delta + \frac{2}{k-2} \right)
\times \sum_{r=1}^{\delta} (-1)^{\delta} \binom{\delta}{i} \binom{n - 2 - \frac{k-2}{k-1} r}{n - j},
$$

for $\delta = 1, 2, \ldots, n - j$. The $s$-th moment of $D_{n,j}^{(k)}$ is

$$
\mathbb{E} \left[ \left( D_{n,j}^{(k)} \right)^s \right] = \frac{1}{(k-2)^s} \left[ (k(k-3))^s + \sum_{r=1}^{s} \binom{s}{r} (k(k-3))^{s-r}(k-2)^r \right.
\times \sum_{i=1}^{r} (-1)^{r-i} \binom{r}{i} \left( \frac{k}{k-2} \right)_i \left( \frac{j + \frac{1}{k-1} + \frac{k-2}{k-1} i}{n-j} \right),
$$

where $\langle \cdot \rangle$ represents the Pochhammer symbol of rising factorial, and $\{ \cdot \}$ represents Stirling numbers of the second kind.

The probability distribution function of $D_{n,j}^{(k)}$ is obtained by exploiting the results in [23, Proposition 14], and the moments are recovered from [52, Proposition 1]. The asymptotic moments of $D_{n,j}^{(k)}$ are obtained directly by applying the Stirling’s approximation to Equation (2); namely,

$$
\mathbb{E} \left[ \left( D_{n,j}^{(k)} \right)^s \right] = \frac{\Gamma \left( j + \frac{1}{k-1} \right) \Gamma \left( s + \frac{k}{k-2} \right)}{\Gamma \left( j + \frac{k-2}{k-1} s + \frac{1}{k-1} \right) \Gamma \left( \frac{k}{k-2} \right)}.
$$

In particular, the asymptotic mean of $D_{n,j}^{(k)}$ is given by

$$
\mathbb{E} \left[ D_{n,j}^{(k)} \right] \sim \frac{k}{k-2} \frac{\Gamma \left( j + \frac{1}{k-1} \right)}{\Gamma(j + 1)} n^{(k-2)/(k-1)},
$$

implying a phase transition in $j = j(n)$:

$$
\mathbb{E} \left[ D_{n,j}^{(k)} \right] \sim \begin{cases} 
\frac{k}{k-2} \left( \frac{n}{j(n)} \right)^{(k-2)/(k-1)}, & j = o(n), \\
\frac{k}{k-2} \left( k - 3 + \alpha^{-(k-2)/(k-1)} \right), & j \sim \alpha n,
\end{cases}
$$

for some $\alpha > 0$. 
5 Small world

In the previous sections, we show that HDRANs are scale-free. In this section, we look into the small-world property of HDRANs. The idea of small-world was introduced by [49]. In the paper, the authors suggested to use the average of local clustering coefficients to assess the small-world effect; that is,

\[ \hat{C}(n) = \frac{1}{V} \sum_v C_v(n), \]

where \( V = V_n^{(k)} \) is the number of vertices and \( C_v(n) \) is the local clustering coefficient of vertex \( v \) at time \( n \). The local clustering coefficient of vertex \( v \) is defined as the proportion of the number of edges in the open neighborhood of \( v \), i.e.,

\[ C_v(n) = \frac{|\{e_{uw}, u, w \in N_v(n)\}|}{|N_v(n)|(|N_v(n) - 1|)/2}, \]

where \( N_v(n) \) is the open neighborhood of \( v \) at time \( n \), \( e_{ij} \) denotes an edge between vertices \( i \) and \( j \), and \( |·| \) represents the cardinality of a set.

For each newcomer, say \( v^* \), to an HDRAN of index \( k \), the open neighborhood of \( v^* \) is comprised by \( k \) vertices of the simplex chosen for recruiting \( v^* \). Thus, the order of the open neighborhood of \( v^* \) is \( k \), and the number of edges in the neighborhood of \( v^* \) is \( \binom{k}{2} \). Upon the first appearance of \( v^* \) in the network, the degree of \( v^* \), denoted \( d_{v^*}(n) \), is \( k \). As an active simplex containing \( v^* \) is selected for recruiting a newcomer in any subsequent time point, \( d_{v^*}(n) \) increases by 1, and the number of edges of the neighborhood of \( v^* \) increases by \( k - 1 \). In general, for a vertex \( v \) of \( \text{deg}_v(n) = j \) at time \( n \), the clustering coefficient is given by

\[ C_v(n) = \frac{(k - 1)(j - k) + \binom{k}{2}}{\frac{j}{2}} = \frac{(k - 1)(2j - k)}{j(j - 1)}. \]

Accordingly, the clustering coefficient of the entire network at time \( n \) is

\[ \hat{C}(n) = \frac{1}{n + k} \sum_v C_v(n) = \frac{1}{n + k} \sum_{j=k}^{k+n-1} \frac{(k - 1)(2j - k)}{j(j - 1)} \times X_{n,j}^{(k)}. \]

where \( X_{n,j}^{(k)} \) denotes the number of vertices of degree \( j \) in the network at time \( n \). When the network is large (i.e., \( n \to \infty \)), the asymptotic clustering coefficient is
given by

$$\hat{C}(\infty) \approx \sum_{j=k}^{\infty} \frac{(k-1)(2j-k)}{j(j-1)} \lim_{n \to \infty} \frac{\mathbb{E}[X_{n,j}^{(k)}]}{n+k}$$

$$= \sum_{j=k}^{\infty} \frac{(k-1)(2j-k)}{j(j-1)} \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k+1)},$$

where the second equality in the last display holds according to Theorem 1. We simplify the expression of $\hat{C}(\infty)$ by applying several algebraic results of gamma function, and get

$$\hat{C}(\infty) \approx \frac{(k-1)\Gamma(2k-1)}{\Gamma(k-1)} \sum_{j=k}^{\infty} \frac{2(j-k)\Gamma(j-1)}{j\Gamma(j+k)}.$$  

We evaluate the two terms in the summand one after another. The first sum is

$$\sum_{j=k}^{\infty} \frac{2\Gamma(j-1)}{\Gamma(j+k)} = \frac{2(2k-1)\Gamma(k-1)}{k\Gamma(2k)};$$

The second sum is

$$\sum_{j=k}^{\infty} \frac{k\Gamma(j-1)}{j\Gamma(j+k)} = \frac{\Gamma(k-1)_{3}F_{2}(1, k-1, k; 2k, k+1; 1)}{\Gamma(2k)}.$$

where $_{3}F_{2}(\cdot, \cdot, \cdot; \cdot, \cdot; \cdot)$ is a generalized hypergeometric function. Putting them together, we thus have

$$\hat{C}(\infty) \approx \frac{k-1}{2k-1} \left( \frac{2(2k-1)}{k} - _{3}F_{2}(1, k-1, k; 2k, k+1; 1) \right).$$

Although hypergeometric functions cannot be written in closed forms in general, we derive the analytical results of $\hat{C}(\infty)$ for several small values of $k$, and present them in Table 2. In particular, the estimated clustering coefficient for triangulated RANs (i.e., $k = 3$) based on our calculation is $12\pi^2 - 353/3 \approx 0.7686$, which is more accurate than $46/3 - 36 \log(3/2) \approx 0.7366$, the result from [55, Equation (6)], according to a simulation experiment (0.7683 based on the average of 50 independent samples, each of which is run over 10000 iterations).
Network index \((k)\) & \(\hat{C}(\infty)\) \\
3 & \(12\pi^2 - \frac{353}{3}\) \\
4 & \(120\pi^2 - \frac{2367}{2}\) \\
5 & \(\frac{2800}{3}\pi^2 - \frac{138161}{15}\) \\
6 & \(6300\pi^2 - \frac{746131}{12}\) \\
7 & \(38808\pi^2 - \frac{134056533}{350}\) \\
8 & \(224224\pi^2 - \frac{663906367}{300}\) \\
9 & \(1235520\pi^2 - \frac{26887974331}{2205}\) \\
10 & \(6563700\pi^2 - \frac{253941996039}{3920}\) 

Table 2: Asymptotic clustering coefficients of HDRANs with small indices \(k\)

6 Sparsity

\textit{Sparsity} is a property of common interest in network modeling [42, 44, 45], as well as in data analytics [1, 11]. As opposed to “dense,” this topology plays a key role when one defines sparse networks. Sparse networks have fewer links than the maximum possible number of links in the (complete) network of same order. In computer science, sparse networks are considered to be somewhere dense or nowhere dense. The investigation of sparsity of HDRANs is inspired by an article recently published on American Physics Society [18]. It was analytically and numerically proven in the article that the probability that a scale-free network is dense is 0, given that the power-law coefficient falls between 0 and 2.

One of the most commonly-used network topology to measure the sparsity of a network \(G(V, E)\) is the \textit{link density} (also known as edge density in the literature):

\[
\text{density}(G) = \frac{|E|}{\binom{|V|}{2}}.
\]

For a HDRAN of index \(k\), denoted \(A^{(k)}_n\), its link density at time \(n\) is a decreasing function of \(n\), viz.,

\[
\text{density } (A^{(k)}_n) = \frac{E^{(k)}_n}{\binom{V^{(k)}_n}{2}} = \frac{k + nk}{\binom{k+n}{2}} = \frac{2(k + nk)}{(k+n)(k+n-1)}.
\]

Observing that the link density of an HDRAN in any form is deterministic given \(k\) and \(n\), we assert that this topology indeed fails to expose the randomness or
to capture the structure of HDRANs. Other topologies that have been proposed to measure the sparsity of both nonrandom and random networks include degeneracy, arboricity, maximum average degree, etc. We refer the interested readers to [39] for textbook style expositions of these topologies and their properties.

In this section, we measure the sparsity of HDRANs via a classical metric—the Gini index [27]. The Gini index which appears more often in economics is commonly used to measure the inequality of income or wealth [16, 27]. The utilization of the Gini index as a sparsity measurement originates in electrical engineering [10]. More often, the Gini index was used to evaluate regularity of graphs [5, 20]. The Gini index debuted as a sparsity measurement of networks in [28].

A graphical interpretation of the Gini index is the Lorenz curve. As portrayed in Figure 6, the Lorenz curve (thick) splits the lower triangle of a unit square into $A$ and $B$. A well-established relationship between the Gini index and the Lorenz curve is that the Gini index of the associated Lorenz curve is equal to the ratio of Area($A$) and Area($A + B$), equivalent to $1 - 2 \times \text{Area}(B)$.

![Figure 2: An example of typical Lorenz curve](image)

We construct the Gini index of HDRANs based on vertex degrees. At time $n$, there is a total of $k + n$ vertices in $A_n^{(k)}$, and the admissible degree set is $\mathcal{J} = \{k, k+1, \ldots, k+n\}$. According to Theorem 1, the mean of the proportion of the number of vertices having degree $j \in \mathcal{J}$ can be approximated by $(\Gamma(j)\Gamma(2k-1))/(\Gamma(j+k)\Gamma(k-1))$, when $n$ is large. For simplicity, let us denote this mean proportion for each pair of $j$ and $k$ by $\gamma(j,k)$. These $\gamma(j,k)$’s altogether naturally form the Lorenz curve after being rearranged in an ascending order. Note that

$$\frac{\partial}{\partial j} \gamma(j,k) = \frac{((\Psi(j) - \Psi(j + k))2^{2k-2}(k-1)\Gamma\left(k - \frac{1}{2}\right)\Gamma(j)}{\Gamma\left(\frac{1}{2}\right)\Gamma(j+k)} < 0,$$

where $\Psi(\cdot)$ is the digamma function, known to be increasing on the positive real line. Hence, the function $\gamma(j,k)$ is decreasing with respect to $j$. 

10
Specifically, we build the Lorenz curve as follows. The bottom of the unit square is equispaced into \((n + 1)\) segments. The bottom left vertex is marked 0 along with vertical value 0. The cumulative proportion value \(\sum_{j=k}^{k+n} \gamma(j,k)\) is assigned to the \(i\)th segmentation from the left. There is a total of \(n\) segmentations between the bottom left and bottom right vertices. Lastly, the vertical value for the bottom right vertex is \(\sum_{j=k}^{k+n} \gamma(j,k)\). The Lorenz curve is comprised by smoothly connecting these assigned values in order, from left to right.

In the next lemma, we show that the Lorenz curve that we established in the last paragraph is well defined, i.e., the two ends of the Lorenz curve respectively coincide with the bottom left and the top right corners of the unit square.

**Lemma 1** We claim that
\[
\lim_{n \to \infty} \sum_{j=k}^{k+n} \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k-1)} = 1 \quad \text{and} \quad \lim_{n \to \infty} \sum_{j=k}^{k+n} \frac{\Gamma(j)\Gamma(2k-1)}{i/n} = 0.
\]
The proof of Lemma 1 is presented in Appendix C. Next, we calculate \(\text{Area}(B)\), equivalent to integrating the Lorenz curve from 0 to 1. For large value of \(n\), the integration can be approximated by applying the trapezoid rule; that is,
\[
\text{Area}(B) \approx \frac{1}{2(n+1)} \left[ \sum_{j=k}^{k+n} \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k-1)} \right. \\
+ \left. \left( \sum_{j=k}^{k+n} \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k-1)} + \sum_{j=k+n-1}^{k+n} \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k-1)} \right) \right. \\
+ \cdots + \left. \left( \sum_{j=k+1}^{k+n} \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k-1)} + \sum_{j=k}^{k+n} \frac{\Gamma(j)\Gamma(2k-1)}{\Gamma(j+k)\Gamma(k-1)} \right) \right] \\
= \frac{1}{2(n+1)} \left( \frac{3k-2}{k-2} - \frac{2^{2k-1}((k-1)n+2)\Gamma\left(k - \frac{1}{2}\right)\Gamma(k+n+1)}{(k-2)\Gamma\left(\frac{1}{2}\right)\Gamma(2k+n)} \right)
\sim n^{-1} - n^{1-k},
\]
In what follows, the Gini index of an HDRAN of index \(k\) at time \(n\) is given by
\[
\text{Gini} \left( A_{n}^{(k)} \right) = 1 - 2 \times \text{Area}(B)
= 1 - \frac{1}{n+1} \left( \frac{3k-2}{k-2} - \frac{2^{2k-1}((k-1)n+2)\Gamma\left(k - \frac{1}{2}\right)\Gamma(k+n+1)}{(k-2)\Gamma\left(\frac{1}{2}\right)\Gamma(2k+n)} \right),
\]
the asymptotic equivalent of which is equal to 1. A large value of Gini index (ranging from 0 to 1) indicates an extremely nonuniform distribution of vertex
degrees, implying that all vertex degrees are dominated by only a few classes, whereas a small value of Gini index suggests vertex degrees are evenly distributed in different degree classes. Thus, we conclude (asymptotically) high sparseness of HDRANs.

We further verify our conclusion by conducting some simulation experiments. In general, each network $G(V, E)$ is associated with a unique $|V| \times |V|$ adjacency matrix, denoted $A = (A_{ij})$, in which $A_{ij} = 1$ only when there is an edge linking vertices $i$ and $j$, for $i, j \in V$; 0, otherwise. If $G$ is undirected, $A$ is symmetric. The degree of vertex $i$ thus can be represented by the sum of $i$th row or the $i$th column in $A$, allowing us to compute the Gini index of each simulated network $G$ through $A$ accordingly. For each $k = 3, 10, 30$, we generate 100 independent HDRANs at time $n = 5000$. The comparison of Lorenz curves (based on the average of cumulative degree proportion sequences) is given in Figure 3.

![Comparison of Lorenz Curves](image)

**Figure 3:** Comparison of Lorenz Curves for simulated HDRANs of $k = 3, 10, 30$ at time $n = 5000$

Besides, we calculate the Gini index of each of the 100 simulated HDRANs (of $k = 3, 10, 30$) at time 50000, and take the average; The estimated Gini indices are 0.9970330 (for $k = 3$), 0.9990327 (for $k = 10$), and 0.9997262 (for $k = 30$). We do not show the corresponding Lorenz curves as they are not visually distinguishable.

7 **Depth, diameter and distance**

In this section, we investigate several distance-based properties of HDRANs. The first measure that we look at is clique-based—*depth*, which is defined (for
HDRANs) recursively as follows. At time 1, the original $k$-clique is divided into $k$ simplexes, and then is deactivated. The depth of each of the active $k$-cliques equals 1. At time $n > 1$, an existing active clique $C^*$ is chosen uniformly at random, and subdivided into $k$ new cliques $C_1, C_2, \ldots, C_k$. Then, we have

$$\text{depth}(C_i) = \text{depth}(C^*) + 1,$$

for all $i = 1, 2, \ldots, k$. An explanatory example of a RAN of index $m = 5$ is shown in Figure 4, where (active) cliques $(1, 0^{(1)}, 0^{(2)}, 0^{(3)}, 0^{(5)})$, $(1, 0^{(1)}, 0^{(2)}, 0^{(4)}, 0^{(5)})$, $(1, 0^{(1)}, 0^{(3)}, 0^{(4)}, 0^{(5)})$, and $(1, 0^{(2)}, 0^{(3)}, 0^{(4)}, 0^{(5)})$ have depth 1; all the rest have depth 2.

![Figure 4: An example of a HDRAN of index 5 at step 2.](image)

In contrast, distance, also known as geodesic distance, is a property based on pairwise vertices. In a given network $G(V, E)$, the distance between a pair of arbitrary vertices $i, j \in V$, denoted $d(i, j)$, is the number of edges in the shortest path (or one of the shortest paths) connecting $i$ and $j$. A related property, diameter of network $G$, denoted $\text{diameter}(G)$, is defined in a max-min manner: the greatest length of the shortest paths between every two vertices in $G$, i.e., $\max_{i,j \in V} \{d(i, j)\}$. See [9, page 82] for fundamental properties of the diameter of a graph. For instance, the diameter of the HDRAN given in Figure 4 is 2, referring to the distance between the vertices respectively labeled with 2 and $0^{(5)}$.

It was introduced in [17] that there exists an one-to-one relation between the evolution of HDRANs (of index $k$) and that of $k$-ary trees\(^4\). An illustrative example is presented in Figure 5. Active and inactive cliques in HDRANs (of index $k$) respectively correspond to external and internal nodes in $k$-ary trees. Thus, the

\(^4\)See [43, page 224] for the definition of $k$-ary tree
Figure 5: The evolution of the 5-ary tree corresponding to that of the HDRAN of index 5 given in Figure 4. Elliptic (internal) nodes refer to inactive cliques, whereas rectangular (active) nodes refer to active ones.

total depth of active cliques in $A_n^{(k)}$ is equivalent to the total depth\footnote{In tree structure, the depth of a node is the number of links between the node and the root (of the tree).} of external nodes in the corresponding $k$-ary tree at time $n$, denoted $T_n^{(k)}$. In the literature, the total depth of external nodes in $T_n^{(k)}$ is also known as the total external path, denoted by $E_n^{(k)}$ in our manuscript. For uniformity, we use $E_n^{(k)}$ as the notation for the total depth of active cliques in $A_n^{(k)}$ as well.

**Proposition 2** Let $E_n^{(k)}$ be the total depth of active cliques in a HDRAN of index $k$ at time $n$. The first two moments of $E_n^{(k)}$ are given by

$$
\mathbb{E} \left[ E_n^{(k)} \right] = (kn - n + 1) \sum_{i=0}^{n-1} \frac{k}{k + (k-1)i}.
$$

$$
\mathbb{E} \left[ (E_n^{(k)})^2 \right] = ((k-1)n + m)((k-1)n + 1)kE(k, n) + O(n^2 \log n),
$$

where $E(k, n)$ is a function of $k$ and $n$, given in Appendix D.
The proof of Proposition 2 also can be found in Appendix D. As we know that 
\[ \sum_{i=0}^{n-1} \frac{k}{k + (k - 1)i} \sim \frac{k}{k - 1} \log n, \]
for large \( n \), we hence conclude that the leading order of the asymptotic expectation of \( E_n^{(k)} \) is \( kn \log n \).

The diameter of HDRANs is also considered. In [24], the authors established an upper bound for the diameter of planar RANs by utilizing a known result of the height of weighted \( k \)-ary trees [10, Theorem 5], i.e.,
\[ \text{diameter} (A_n^{(3)}) \leq \rho \log n, \]
where \( \rho = 1/\eta \), and \( \eta \) is the unique solution greater than 1 for \( \eta - 1 - \log \eta = \log 3 \). This upper bound can be extended to \( A_n^{(k)} \) effortlessly; that is,
\[ \text{diameter} (A_n^{(k)}) \leq \frac{2}{\rho^* (k - 1)} \log n, \]
where \( \rho^* = 1/\eta^* \) is the unique solution greater than 1 for \( \eta^* - 1 - \log \eta^* = \log k \). In addition, the authors of [21] proved \( \text{diameter} (A_n^{(3)}) \sim c \log n \) by estimating the height of a class of specifically-designed random trees. The value of \( c \) is approximately 1.668. The asymptotic expression of the diameter of more general \( A_n^{(k)} \) was developed by [13] and by [33]. The approach in [13] is to utilize known results of continuous-time branching processes coupled with recurrence method, and the authors of [33] coped with difficulties by characterizing vertex generations. We only state (without repeating the proof) the weak law of the diameter of \( A_n^{(k)} \) from [13] (with a minor tweak) in the next theorem.

**Theorem 3 ([13, Theorem 2])** For \( k \geq 3 \), with high probability, we have
\[ \text{diameter} (A_n^{(k)}) \sim c \log n, \]
where \( c \) is the solution of
\[ \frac{1}{c} = \sum_{\ell=0}^{k-1} \frac{k - 1}{\ell + a(k - 1)}, \]
in which the value of \( a \) is given by
\[ \frac{\Gamma(k+1)\Gamma(k\alpha)}{\Gamma((k-1)a+k)} \exp \left\{ \sum_{\ell=0}^{k-1} \frac{(k-1)(a+1) - 1}{\ell + (k-1)a} \right\} = 1. \]
Especially, as \( k \to \infty \),
\[ c \sim \frac{1}{k \log 2}. \]
A topological measure related to distance is the Wiener index, which was proposed by the chemist Harry Weiner \[50\] to study molecular branching of chemical compounds. For a network \(G(V, E)\), the Wiener index is defined as the sum of distances of all paired vertices, i.e., \(W(G) = \sum_{i,j \in V} d(i, j)\). The Weiner index has been extensively studied for random trees \[19, 38\]. For other random structures, we refer the readers to \[7, 25, 31, 46, 47, 48\].

The methodologies for the Wiener index of random trees, however, are not adaptable to the study of RANs, as the bijection between RANs and \(k\)-ary trees is based on a clique-to-node mapping. The high dependency of active cliques (sharing vertices and edges) substantially increases the challenge of formulating mathematical relation between distance (vertex-based) and depth (clique-based).

There is only a few articles studying distance or related properties in RANs. In \[33\], the authors proved that the distance of two arbitrary vertices in a HDRAN has both mean and variance of order \(\log n\), and that this distance follows a Gaussian law asymptotically. However, it seems difficult to extend this result to the Weiner index, as the covariance structure of the distances (of all paired vertices) is unspecified. Planar RANs (\(A_n^{(3)}\)) were considered in \[8\]. In this article, the dominant term of the total distance of all pairs of vertices was shown to be \(\sqrt{3\pi n^{5/2}}/22\). The main idea was to consider an enumerative generating function of the total distance, and then decompose the total distance into interdistance and extradistance. This approach can be extended to HDRANs of small network index \(k\), but seemingly not applicable to HDRANs with general index \(k\). Therefore, the Wiener index of HDRANs remains an open problem up to date.

We numerically look into the Wiener index of HDRANs via a series of simulations. For \(k = 3, 5, 8, 10\), we generate 500 independent HDRANs at time 2000, calculate the Wiener index for each simulated HDRAN, and use the kernel method to estimate the density. The plots of the estimated densities are presented in Figure 6 where we find that they are approximately bell-shaped, but not symmetric (positively skewed). By observing these patterns, we conjecture that the limiting distribution of the Wiener index of HDRANs does not follow a Gaussian law. In addition, for each \(k\), we apply the Shapiro-Wilk test to the simulated data comprising 500 Wiener indices, and receive the following \(p\)-values: 0.0003 for \(k = 3\); 0.0024 for \(k = 5\); 9.56 \times 10^{-8} for \(k = 8\); and 0 for \(k = 10\). These \(p\)-values are all statistically significant, in support of our conjecture.

8 Concluding remarks

In this section, we give some concluding remarks and propose some future work. We investigate several properties of high-dimensional random Apollonian networks in this paper. Two types of degree profiles are considered. For the first
Figure 6: Density estimation of the Wiener indices of HDRANs for $k = 3, 5, 8, 10$

type, we show that the number of vertices of a given degree concentrates on its
expectation with high probability. In the proof of Theorem 1, we derive the
$L_1$ limit of $X^{(k)}_{n,j}$, i.e.,

$$
\lim_{n \to \infty} \mathbb{E} \left[ X^{(k)}_{n,j} \right] = \frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j + k)\Gamma(k - 1)} n,
$$

which suggests that the asymptotic expectation of $X^{(k)}_{n,j}$ experiences a phase transition. There are two regimes. According to the Stirling’s Approximation, we have

$$
\mathbb{E} \left[ X^{(k)}_{n,j} \right] \sim \begin{cases} 
\frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j + k)\Gamma(k - 1)} n, & \text{for fixed } j; \\
\frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j + k)\Gamma(k - 1)} \frac{n}{j^2}, & \text{for } j \to \infty,
\end{cases}
$$

as $n \to \infty$.

For the second type of degree profile, the degree of a vertex of a given label, we develop the probability mass function and the exact moments by applying the analytic combinatorics methods and the results in triangular Pólya urns.

The next two properties that we investigate are the small world property measured by the local clustering coefficient and the sparsity measured by a proposed Gini index. We conclude that HDRANs are highly clustered and sparse.

The last several properties that we look into are distance-based. According to an one-to-one relation between HDRANs and $k$-ary trees, we compute the first two moments of the total depth of active cliques in HDRANs. We also numerically
study the Wiener index, and conjecture that its limiting distribution is not normal based on simulation results. The diameter of HDRANs is retrieved from [14].

Finally, we propose some future work. Our conjecture of non-normality of the Wiener index is based on numerical experiments. A more rigorous proof is needed. There remain many open problems for HDRANs, such as the length of the longest path and the highest vertex degree. One suggests studies of stochastic processes that take place on HDRANs, especially the processes with applications to mathematical physics, such as percolation and diffusion. We will investigate these open problems and report the results elsewhere.

References

[1] Arnold, G., Höltl, J., Köksal, A., Bodík, R. and Sagiv, M. (2010) Specifying and verifying sparse network codes. *ACM SIGPLAN Not.* 45 249–260.

[2] Andrade, J.S. Jr., Herrmann, H.J., Andrade, R.F.S. and da Silva, L.R. (2005) Apollonian networks: Simultaneously scale-free, small world, Euclidean, space filling, and with matching graphs. *Phys. Rev. Lett.* 94 018702.

[3] Azuma, K. (1967) Weighted sums of certain dependent random variables. *Tohoku Math. J.* 19 357–367.

[4] Bagchi, A. and Pal. A.K. (1985) Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods* 6 394–405.

[5] Balaji, H. and Mahmoud, H.M. (2017) The Gini index of random trees with an application to caterpillars. *J. Appl. Probab.* 54 701–709.

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

[7] Bereg, S. and Wang, H. (2007) Wiener indices of balanced binary trees. *Discrete Appl. Math.* 155 457–467.

[8] Bodini, O., Darrasse, A. and Soria, M. (2008) Distance in Apollonian network structures. In *Proceedings of Annual International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC’08)* 307–318. Valparaíso, Chile.

[9] Bondy, J.A. and Murty, U.S.R. (2008) *Graph theory.* Springer. New York, NY.
[10] Broutin, N. and Devroye, L. (2006) Large deviations for the weighted height of an extended class of trees. *Algorithmica* **46** 271–297.

[11] Buluç, A., Gilbert, J. and Shah, V.B. (2011) Implementing Sparse Matrices for Graph Algorithms. In J. Kepner and J. Gilbert (Eds.) *Graph Algorithms in the Language of Linear Algebra*, 287–313. S.I.A.M. Philadelphia, PA.

[12] Collevecchio, A., Mehrabian, A. and Wormald, N. (2014) Longest paths in random Apollonian networks and largest $r$-ary subtrees of random $d$-ary recursive trees. *J. Appl. Probab.* **53** 846–856.

[13] Cooper, C., Frieze, A. and Uehara, R. (2014) The height of random $k$-trees and related branching processes. *Random Structures Algorithms* **45** 675–702.

[14] Cooper, C. and Frieze, A. (2015) Long paths in random Apollonian networks. *Internet Math.* **11** 308–318.

[15] Corless, R.M., Gonnet G.H., Hare D.E.G., Jeffrey D.J. and Knuth D.E. (1996) On the LambertW function. *Adv. Comput. Math.* **5** 329–359.

[16] Dalton, H. (1920) The measurement of the inequality of incomes. *Econom. J.* **30** 348–361.

[17] Darrasse, A. and Soria, M. (2007) Degree distribution of random Apollonian network structures and Boltzmann sampling. In *Proceedings of Conference on Analysis of Algorithms (AofA’07)* 313–324. Juan-les-Pins, France.

[18] Del Genio, C.I., Gross, T. and Bassler, K.E. (2011). All scale-free networks are sparse. *Phys. Rev. Lett.* **107** 178701.

[19] Dobrynin, A.A., Entringer, R. and Gutman, I. (2001) Wiener index of trees: Theory and applications. *Acta Appl. Math.* **66** 211–249.

[20] Domicolo, C. and Mahmoud, H. (2019) Degree-based Gini index for graphs. To appear in *Probab. Eng. Informational Sci*.

[21] Ebrahimzadeh, E., Farczade, L., Gao, P., Mehrabian, A., Sato, C.M., Wormald, N. and Zung, J. (2014) On longest paths and diameter in random Apollonian networks. *Random Structures Algorithms* **45** 703–725.

[22] Fan, C. and Lu, L. (2006) *Complex graphs and networks*. American Mathematical Society. Providence, RI.
[23] Flajolet, P., Duma, P. and Puyhaubert, V. (2006) Some exactly solvable models of urn process theory. In Proceedings of Fourth Colloquium on Mathematics and Computer Science Algorithms, Trees, Combinatorics and Probabilities 59–118. Nancy, France.

[24] Frieze, A. and Tsourakakis, C.E. (2014) Some properties of random Apollonian networks. Internet Math. 10 162–187.

[25] Fuchs, M. and Lee, C.-K. (2015) The Wiener index of random digital trees. SIAM J. Discrete Math. 29 586–641.

[26] Gao, Y. (2009) The degree distribution of random $k$-trees. Theoret. Comput. Sci. 410 688–695.

[27] Gini, C. (1921) Measurement of inequality of incomes. Econom. J. 31 124–126.

[28] Goswami, S., Mruthy, C.A. and Das, A.K. (2018) Sparsity measure of a network graph: Gini index. Inf. Sci. 462 19–36.

[29] Gradshteyn, I.S. and Ryzhik, I.M. (2007) Table of Integrals, Series and Products. Academic Press. Cambridge, MA.

[30] Herley, N. and Rickard, S. (2009). Comparing measures of sparsity. IEEE Trans. Inf. Theory 55 4723–4741.

[31] Janson, S. (2003) The Wiener index of simply generated random trees. Random Structures Algorithms 22 337–358.

[32] Janson, S. (2006) Limit theorems for triangular urn schemes. Probab. Theory Related Fields 134 417–452.

[33] Kolossváry, I., Komjáthy, J. and Vágó, L. (2016) Degrees and distances in random and evolving Apollonian networks. Adv. Appl. Prob. 48 865–902.

[34] Kuratowski, C. and Menger, K. (1930) Remarques sur la théorie axiomatique de la dimension. Monatsh. Math. Phys. 37 169–174.

[35] Lorenz, M.O. (1905) Methods for measuring the concentration of wealth. J. Amer. Statist. Assoc. 9 209–219.

[36] Mahmoud, H.M. (2009) Pólya urn models. CRC Press. Boca Raton, FL.

[37] Moody, J. (2001) Race, School Integration, and Friendship Segregation in America Am. J. Sociol. 107 679–716.
[38] Neininger, R. (2002) The Wiener index of random trees. Combin. Probab. Comput. 11 587–597.

[39] Nešetřil, J. and Ossona de Mendez, P. (2012) Sparsity—Graphs, Structures and Algorithms. Springer. New York, NY.

[40] Finding and evaluating community structure in networks Newman, M.E.J. and Girvan, M. (2004) Finding and evaluating community structure in networks. Phys. Rev. E 69 026113.

[41] Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis. Chapman & Hall, London, UK.

[42] Singh, A. and Humphries, M.D. (2015) Finding communities in sparse networks. Sci. Rep. 5 8828.

[43] Storer, J.A. (2002) An introduction to data structures and algorithms. Birkhäuser. Boston, MA.

[44] Verzelen, N. and Arias-Castro, E. (2015) Community detection in sparse random networks. Ann. Appl. Probab. 25 3645–3510.

[45] Vinciotti, V. and Hashem, H. (2013) Robust methods for inferring sparse network structures. Comput. Statist. Data Anal. 67 84–94.

[46] Wagner, S. (2006) A class of trees and its Wiener index. Acta Appl. Math. 91 119–132.

[47] Wagner, S. (2007) On the Wiener index of random trees. Discrete Math. 312 1502–1511.

[48] Wagner, S. (2007) On the average Wiener index of degree-restricted trees. Australas. J. Comb. 37 187–203.

[49] Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of “small-world” networks. Nature 393 440–442.

[50] Wiener, H. (1947) Structural determination of paraffin boiling points. J. Am. Chem. Soc. 69 17–20.

[51] Zhang, P., Chen, C. and Mahmoud, H.M. (2015) Explicit characterization of moments of balanced triangular Pólya urns by an elementary approach. Statist. Probab. Lett. 96 149–153.

[52] Zhang, P. and Mahmoud, H.M. (2016) The degree profile and weight in Apollonian networks and $k$-trees. Adv. in Appl. Probab. 48 163–175.
[53] Zhang, P. (2016) On properties of several random networks. Ph.D. thesis, Department of Statistics, The George Washington University, Washington, D.C., U.S.A.

[54] Zhang, Z., Comellas, F., Fertin, G. and Rong, L. (2006) High-dimensional Apollonian networks. J. Phys. A. 39 1811-1818.

[55] Zhou, T., Yan, G. and Wang, B.-H. (2005) Maximal planar networks with large clustering coefficient and power-law degree distribution. Phys. Rev. E 71 046141.
A Proof of Theorem 1

To establish a sharp bound for the difference between the expectation of $X^{(k)}_{n,j}$ and its $L_1$ limit (after being properly scaled) for $j \geq k$, we distinguish the case of $j = k$ and the case of $j > k$.

A.1 The case of $j = k$

The vertices of degree $k$ form a special class in RANs of index $k$—terminal vertices. Terminal vertices never recruit newcomers since their first appearance in the network. A stronger almost sure limit of $X^{(k)}_{n,k}$ was developed via a two-color Pólya urn model\textsuperscript{6} in [52]; that is, $X^{(k)}_{n,k} \xrightarrow{a.s.} \frac{k - 1}{2k - 1} n$.

According to the result in Theorem 1, the $L_1$ bound for $X^{(k)}_{n,k}$ is given by

$$
\left| \mathbb{E}[X^{(k)}_{n,k}] - \frac{k - 1}{2k - 1} n \right| \leq \frac{2k^2}{2k - 1}.
$$

(3)

We prove this result via an induction on $n \in \mathbb{N}$. Obviously, Equation (3) is valid for $n = 1$, as we have $\mathbb{E}[X^{(k)}_{1,k}] = X^{(k)}_{1,k} = 1$. Assume that Equation (3) holds up to some integer $m$. Consider a recursive relation between $X^{(k)}_{n+1,k}$ and $X^{(k)}_{n,k}$ for all $n \geq 1$. Denote $\text{deg}_v(n)$ the degree of vertex $v$ at time $n$, and let $\mathbb{1}\{\text{deg}_v(n) = j\}$ be an indicator function which equals 1, if $\text{deg}_v(n) = j$; 0, otherwise. In general, the expected value of $X^{(k)}_{n,j}$ can be written in terms of the following:

$$
\mathbb{E}[X^{(k)}_{n,j}] = \sum_v \mathbb{E}[\mathbb{1}\{\text{deg}_v(n) = j\}] .
$$

(4)

Besides, we have the following almost-sure relation of the degree of $v$ between time $n + 1$ and $n$:

$$
\mathbb{1}\{\text{deg}_v(n + 1) = k\} = \mathbb{1}\{\text{deg}_v(n) = k\} \mathbb{P}(v \text{ is not chosen at time } n + 1).
$$

In what follows, we obtain a recurrence between the first moments of $X^{(k)}_{n+1,k}$ and $X^{(k)}_{n,k}$; that is,

$$
\mathbb{E}[X^{(k)}_{n+1,k}] = \mathbb{E}[X^{(k)}_{n,k}] \left(1 - \frac{k}{(k - 1)n + 1}\right) + 1.
$$

We refer the interested readers to [36] for text style exposition of Pólya urn models.
For the inductive step $n = m + 1$, we have

\[
\begin{align*}
|\mathbb{E}[X_{m+1,k}^{(k)}] - \frac{k-1}{2k-1}(m+1)| \\
= \left| \mathbb{E}[X_{m,k}^{(k)}] \left(1 - \frac{k}{(k-1)m+1}\right) + 1 - \frac{k-1}{2k-1}(m+1) \right| \\
\leq \left| \left( \mathbb{E}[X_{m,k}^{(k)}] - \frac{k-1}{2k-1}m \right) \left(1 - \frac{k}{(k-1)m+1}\right) \right| \\
+ \frac{k}{2k-1} \frac{k(k-1)m}{(2k-1)((k-1)m+1)} \\
\leq \frac{2k^2}{2k-1} \left(1 - \frac{k}{(k-1)m+1}\right) + \frac{k}{(2k-1)((k-1)m+1)},
\end{align*}
\]

which completes the proof.

A.2 The case of $j > k$

Due to high dependency of HDRAN structure in the network growth, it is difficult to use classical probabilistic methods, such as Pólya urns or recurrence methods, to determine the $L_1$ limit or establish an $L_1$ bound for $X_{n,j}^{(k)}$ for general $j > k$. The reason is that the recurrence for the expectation of $X_{n,j}^{(k)}$ does not have an analytic solution.

For this case, we prove the theorem by a two-dimensional induction on $n = \{1, 2, 3, \ldots\}$ and $j = \{k, k+1, \ldots, k+n-1\}$. Consider an infinite lower triangle table in which the rows are indexed by $n$ and the columns are indexed by $j$. A illustrative diagram of the inductive progression can be found in [53, page 69]. The leftmost column and the diagonal of the triangle jointly form the bases of the induction. Notice that the leftmost column refers to the case of $j = k$, which has already been verified for all $n$. The basis on the diagonal can be proved in an analogous manner. We omit the details here.

Assume that the result stated in the theorem holds up to $j = \ell > k$. Before proving the inductive step, we establish a two-dimensional recursive relation for $\mathbb{E}[X_{n,j}^{(k)}]$ for $n$ and $j$. Since the degree for each vertex in the network increases at most by one at each evolutionary step, we observe an almost-sure relation for $\deg_v(n)$ as follows:

\[
1 \{\deg_v(n+1) = j\} = 1 \{\deg_v(n) = j\} \mathbb{P}(v \text{ is not chosen at time } n+1) \\
+ 1 \{\deg_v(n) = j-1\} \mathbb{P}(v \text{ is chosen at time } n+1) \\
= 1 \{\deg_v(n) = j\} \left(1 - \frac{j}{(k-1)n+1}\right)
\]
\[ + 1 \{ \text{deg}_v(n) = j - 1 \} \frac{j - 1}{(k - 1)n + 1}. \]

According to Equation (4), we then obtain a recurrence for \( E[X_{n,j}] \); namely,

\[
E[X_{n+1,j}] = E[X_{n,j}] \left( 1 - \frac{j}{(k - 1)n + 1} \right) + E[X_{n,j-1}] \frac{j - 1}{(k - 1)n + 1}.
\]

To the best of our knowledge, the recurrence above does not have an analytic solution. We exploit a well-known result in [22], to compute the asymptotic expectation of \( E[X_{n,j}] \), and to determine the value of \( b_{j,k} \) subsequently. It was shown in [22] that a sequence \( \{ \alpha_n \} \) which satisfies the recurrence

\[
\alpha_{n+1} = \left( 1 - \frac{\beta_n}{n + \xi} \right) \alpha_n + \gamma_n
\]

for \( n \geq n_0 \) such that \( \lim_{n \to \infty} \beta_n = \beta > 0 \) and \( \lim_{n \to \infty} \gamma_n = \gamma \) has the following limiting result:

\[
\lim_{n \to \infty} \frac{\alpha_n}{n} = \frac{\gamma}{1 + \beta}.
\]

Consider the following settings: \( \alpha_n = E[X_{n,j}] \), \( \beta_n = j/(k - 1) \), \( \xi = 1/(k - 1) \), and \( \gamma_n = E[X_{n,j-1}] \frac{j - 1}{(k - 1)n + 1} \). We then have

\[
\lim_{n \to \infty} \frac{E[X_{n,j}]}{n} = b_{j-1,k} \frac{j - 1}{j + k - 1},
\]

which in fact establishes a heirarchical recurrence for \( b_{j,k} \) for \( j \geq k \); that is,

\[
b_{j,k} = b_{j-1,k} \frac{j - 1}{j + k - 1},
\]

with the initial value \( b_{k,k} = (k - 1)/(2k - 1) \). We solve the recurrence to get

\[
b_{j,k} = \frac{(k - 1)\Gamma(j)\Gamma(2k)}{(2k - 1)\Gamma(j + k)\Gamma(k)} = \frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j + k)\Gamma(k - 1)}.
\]

We are now at the position to prove the inductive step. For \( j = \ell + 1 \), we have

\[
E[X_{n+1,\ell+1}] - b_{\ell+1,k}(n + 1)
\]

\[
= E[X_{n,\ell}] \left( 1 - \frac{\ell}{(k - 1)n + 1} \right) + E[X_{n,\ell-1}] \frac{\ell - 1}{(k - 1)n + 1} - b_{\ell+1,k}(n + 1)
\]

25
Let $(M_n)_{n=0}^N$ be a martingale sequence such that $|M_{n+1} - M_n| \leq c$ for all $0 \leq n \leq N - 1$. For an arbitrary $\lambda > 0$, we have

$$P(|M_n - M_0| \geq \lambda) \leq e^{-\frac{\lambda^2}{2c^2N}}.$$ 

Let $(\Omega, \mathcal{F}, P)$ be the probability space induced by a HDRAN after $n$ insertions. Fix $j \geq k$, let $(M_i)_{i=0}^n$ be the martingale sequence defined as $M_i = \mathbb{E} \left[ X_{n,j}^{(k)} \mid \mathcal{F}_i \right]$, where $\mathcal{F}_i$ is the $\sigma$-field generated by the HDRAN in the first $i$ steps. Recall that $M_0 = \mathbb{E} \left[ X_{n,j}^{(k)} \mid \mathcal{F}_0 \right] = \mathbb{E} \left[ X_{n,j}^{(k)} \right]$ and $M_n = \mathbb{E} \left[ X_{n,j}^{(k)} \mid \mathcal{F}_n \right] = X_{n,j}^{(k)}$. Consider two stochastic sequences of choosing cliques, $S := C_1, C_2, \ldots, C_{s-1}, C_s, \ldots$ and $S' := C_1, C_2, \ldots, C_{s-1}, C'_s, \ldots$ in which the first different choices of cliques appear at time $s$. The change of the number of vertices of degree $j$ at time $s$ is at most $2k$, referring to the $2k$ vertices involved in the cliques $C_s$ and $C'_s$. We manipulate the choices of cliques in the two sequences after time $s$ as follows:

- If an active clique inside of $C_s$ in $S$ is chosen, we select the same clique in $S'$;

- If an active clique outside of $C_s$ in $S$ is chosen, we select the a clique inside $C'_s$ according to a preserved isomorphic mapping in $S'$.

Noticing that $C_s$ and $C'_s$ are arbitrary, we conclude that the difference between the number of vertices with degree $j$ in $S$ and $S'$ is at most $2k$ in average. Thus, we have $|M_{i+1} - M_i| \leq 2k$, for all $0 \leq i \leq n - 1$. The result in Theorem 2 is then immediately obtained by applying the Azuma-Hoeffding inequality.
C Proof of Lemma 1

C.1 The first part

The identity in the left-hand side of the equation is summable:

\[ \sum_{j=k}^{k+n} \frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j+k)\Gamma(k-1)} = \frac{\Gamma(2k - 1)}{\Gamma(k-1)} \sum_{j=k}^{k+n} \frac{\Gamma(j)}{\Gamma(j+k)} \]

\[ = \frac{\Gamma(2k - 1)}{\Gamma(k-1)} \left( -\frac{\Gamma(k+n+1)}{(k-1)\Gamma(2k+n)} + \frac{\Gamma \left( \frac{1}{2} \right) 2^{-2k+2}(2k-1)}{2(k-1)\Gamma \left( k + \frac{1}{2} \right)} \right) \]

\[ = -\frac{\Gamma(2k-1)\Gamma(k+n+1)}{\Gamma(k)\Gamma(2k+n)} + 1. \]

As \( n \to \infty \), we apply the Stirling’s approximation to get

\[ \lim_{n \to \infty} \left( -\frac{\Gamma(2k-1)\Gamma(k+n+1)}{\Gamma(k)\Gamma(2k+n)} + 1 \right) \sim \lim_{n \to \infty} (n^{1-k}) + 1 = 1, \]

for \( k \geq 3 \).

C.2 The second part

We change the lower bound of the summation to \( k + n - i + 1 \), and get

\[ \sum_{j=k+n-i+1}^{k+n} \frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j+k)\Gamma(k-1)} = \sum_{j=k}^{k+n} \frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j+k)\Gamma(k-1)} - \sum_{j=k}^{k+n-i} \frac{\Gamma(j)\Gamma(2k - 1)}{\Gamma(j+k)\Gamma(k-1)} \]

\[ = \frac{\Gamma(2k-1)\Gamma(k+n-i+1)}{\Gamma(k)\Gamma(2k+n-i)} - \frac{\Gamma(2k-1)\Gamma(k+n+1)}{\Gamma(k)\Gamma(2k+n)} \]

\[ \sim n^{1-k}, \]

by the Stirling’s approximation. For \( k \geq 3 \), we have \( n^{1-k} = o(n^{-1}) \), which completes the proof.

D Proof of Proposition 2

The proof is based on \( T^{(k)}_n \). A related property of \( E^{(k)}_n \) is the number of external nodes in \( T^{(k)}_n \), denoted by \( L^{(k)}_n \). According to the evolution of \( k \)-ary trees, we have

\[ L^{(k)}_n = 1 + n(k-1). \]
At time \( n \), we enumerate all external nodes in \( T_n^{(k)} \) with respect to a preserved isomorphic mapping, e.g., from top to bottom, and from left to right for the external nodes at the same level. Let \( D_{i,n}^{(k)} \) be the depth of the external node labeled with \( i \) in \( T_n^{(k)} \). We thus have

\[
\mathcal{E}_n^{(k)} = \sum_{i=1}^{L_n^{(k)}} D_{i,n}^{(k)}.
\]

Note that the quantity of \( \mathcal{E}_n^{(m)} \) increases monotonically with respect to \( n \), and the amount of increase depends on the depth of the node sampled at each time point. Suppose that the external node labeled with \( i \) is selected upon time \( n \). The increment from \( \mathcal{E}_n^{(k)} - \mathcal{E}_{n-1}^{(k)} \) to \( \mathcal{E}_n^{(k)} \) is

\[
k \left( D_{i,n-1}^{(k)} + 1 \right) - D_{i,n-1}^{(k)} = (k - 1)D_{i,n-1}^{(k)} + k,
\]

leading to the following almost-sure relation conditioning on \( T_n^{(m)} \) and label \( i \), i.e.,

\[
\mathcal{E}_n^{(k)} = \mathcal{E}_{n-1}^{(k)} + (k - 1)D_{i,n-1}^{(k)} + k. \tag{5}
\]

We obtain a recurrence for \( \mathbb{E} \left[ \mathcal{E}_n^{(k)} \right] \) by averaging \( i \) out in Equation (5); that is,

\[
\mathbb{E} \left[ \mathcal{E}_n^{(k)} \middle| T_n^{(k)} \right] = \mathcal{E}_{n-1}^{(k)} + \frac{k - 1}{(k - 1)(n - 1) + 1} \mathcal{E}_{n-1}^{(k)} + k
\]

\[
= \frac{(k - 1)n + 1}{(k - 1)(n - 1) + 1} \mathcal{E}_{n-1}^{(k)} + k,
\]

equivalent to (after taking another expectation both sides)

\[
\mathbb{E} \left[ \mathcal{E}_n^{(k)} \right] = \frac{(k - 1)n + 1}{(k - 1)(n - 1) + 1} \mathbb{E} \left[ \mathcal{E}_{n-1}^{(k)} \right] + k. \tag{6}
\]

Noting the initial condition \( \mathcal{E}_0^{(k)} = 0 \), we solve Equation (6) recursively for \( \mathbb{E} \left[ \mathcal{E}_n^{(k)} \right] \) to get

\[
\mathbb{E} \left[ \mathcal{E}_n^{(k)} \right] = \left( \frac{k n - n + 1}{k - 1} \right)^m \left[ \Psi \left( n + \frac{k}{k - 1} \right) - \Psi \left( \frac{k}{k - 1} \right) \right], \tag{7}
\]

where \( \Psi(\cdot) \) represents the \textit{digamma function}. We finally apply a known formula for difference equation of digamma functions \([29, \text{Eq. 3.231.5}]\) to obtain the result stated in the theorem.
For the second moment of \( E_n^{(k)} \), we consider the following convolution variable:

\[
D_n^{(k)} = \sum_{i=1}^{k_n} \left( D_{i,n}^{(k)} \right)^2,
\]

the sum of squared depths of external nodes in \( T_n^{(k)} \). We implement a strategy analogous for the calculation of \( E_n^{(k)} \) to compute the expectation of \( D_n^{(k)} \), which, later on, is used in the computation of the second moment of \( E_n^{(k)} \). For better readability of the article, we omit the details of the derivation of \( E \left[ D_n^{(k)} \right] \), but just state the result:

\[
E \left[ D_n^{(k)} \right] = k \left[ (k-1)n + 1 \right] \left( \sum_{j=1}^{n-1} \left( \frac{2k \Psi \left( \frac{(k-1)j + 2k-1}{k-1} \right) - 2k \Psi \left( \frac{k}{k-1} \right)}{(k-1)j + k} + \frac{(k-1) ((k-1)j - m)}{(k-1)j + k} \right) \right)^2.
\] (8)

Conditional on \( T_n^{(k)} \) and label \( i \), we square the almost-sure relation in Equation (5) to get

\[
(E_n^{(k)})^2 = (E_{n-1}^{(k)})^2 + (k-1)^2 \left( D_{i,n-1}^{(k)} \right)^2 + k^2 + 2E_n^{(k)}D_{i,n-1}^{(k)} + 2kE_{n-1}^{(k)} + 2kD_{i,n-1}^{(k)}.
\]

We average out \( i \) to obtain

\[
E \left[ (E_n^{(k)})^2 \mid T_n^{(k)} \right] = \left( 1 + \frac{2(k-1)}{(k-1)(n-1) + 1} \right) \left( E_{n-1}^{(k)} \right)^2 + (k-1)^2E_{n-1}^{(k)} + 2kE_n^{(k)} + 2kD_{i,n-1}^{(k)}.
\]

Taking the expectation with respect to \( T_n^{(k)} \) and plugging in the results of \( E \left[ D_n^{(k)} \right] \) (c.f. Equation (8)) and \( E \left[ E_n^{(k)} \right] \) (c.f. Equation (7)), we obtain a recurrence for

\[
E \left[ (E_n^{(k)})^2 \right] = \left( 1 + \frac{2(k-1)}{(k-1)(n-1) + 1} \right) E \left[ (E_{n-1}^{(k)})^2 \right] + C(k, n),
\] (9)

where \( C(k, n) \) is a known function of \( m \) and \( n \). Solving the recurrence relation for \( E \left[ (E_n^{(k)})^2 \right] \) with the initial condition \( E \left[ (E_1^{(k)})^2 \right] = m^2 \), we obtain the stated result in the theorem.

29
At last, we present the exact expressions of the two functions: $E(k, n)$ and $C(k, n)$.

$$E(k, n) = \frac{1}{((k - 1)i + 2k - 1)((k - 1)i + k)} \sum_{i=1}^{n-1} \left(2k^2 - 2k\right) \sum_{j=1}^{n-1} \frac{\Psi\left(\frac{(k-1)j+2k-1}{k-1}\right)}{(k-1)j + k}$$

$$- 2\Psi\left(\frac{k}{k-1}\right) - k + 1 \Psi\left(\frac{(k-1)i + 2k - 1}{k-1}\right)$$

$$+ 2k \left[i\Psi\left(\frac{(k-1)i + k}{k-1}\right) + \Psi\left(1, \frac{(k-1)i + 2k - 1}{k-1}\right)\right]$$

where $\Psi(1, \cdot)$ is the first order derivative of the digamma function;

$$C(k, n) = \frac{2k}{k-1} \left\{ k(k - 1)^2 \sum_{i=0}^{n-1} \frac{\Psi\left(\frac{(k-1)i+2k-1}{k-1}\right)}{(k - 1)i + k} - \left[(k^2 - k)n + \frac{k^2 + 1}{2}\right] \Psi\left(\frac{k}{k-1}\right)\right\}$$

$$+ (k^2 - k) \left[\Psi\left(1, \frac{(k-1)n + 1}{k-1}\right) - \Psi\left(1, \frac{k}{k-1}\right) + \Psi\left(1, \frac{k}{k-1}\right)\right]$$

$$+ \Psi\left(\frac{(k-1)n + 1}{k-1}\right) \left[(k^2 - k)n + \frac{k^2 + 1}{2} - (k^2 - k)\Psi\left(\frac{k}{k-1}\right)\right] + \frac{k^2 - k}{2}.$$