The degree distribution of random *k*-trees

Yong Gao *

Department of Computer Science, Irving K. Barber School of Arts and Sciences, University of British Columbia Okanagan, Kelowna, Canada V1V 1V7

ARTICLE INFO

ABSTRACT

A power law degree distribution is established for a graph evolution model based on the graph class of *k*-trees. This *k*-tree-based graph process can be viewed as an idealized model that captures some characteristics of the preferential attachment and copying mechanisms that existing evolving graph processes fail to model due to technical obstacles. The result also serves as a further cautionary note reinforcing the point of view that a power law degree distribution should not be regarded as the only important characteristic of a complex network, as has been previously argued [D. Achlioptas, A. Clauset, D. Kempe, C. Moore, On the bias of traceroute sampling, or power-law degree distribution in regular graphs, in: Proceedings of the 37th ACM Symposium on Theory of Computing, STOC'05, 2005, pp. 694–703; L. Li, D. Alderson, J. Doyle, W. Willinger, Towards a theory of scale-free graphs: Definition, properties, and implications, Internet Mathematics 2 (4) (2005) 431–523; M. Mitzenmacher, The future of power law research, Internet Mathematics, 2 (4) (2005) 525–534].

1. Introduction

Since the discovery of the power-law degree distribution of the web graphs and other complex large-scale networks, many random models for such networks have been proposed [3,6,9,10]. By studying a variety of graph models with a power law degree distribution, it is hoped that one can gain insight into the characteristics of real-world complex networks that are algorithmically exploitable, and can use these models as a tool for empirical studies [7]. It is therefore desirable to have random models that not only exhibit power law degree distributions, but also have other structural features specified in a controlled manner.

Most of the existing models for complex networks define a graph evolution process in which vertices are added to the current graph one at a time. In each time step, the newly-added vertex is connected to a number of existing vertices selected according to some probability distribution. Two popular ways to specify the probability distribution for vertex selection are *preferential attachment* and *copying* (also known as *duplication*). In the preferential attachment model, an existing vertex is selected with probability in proportion to its vertex degree. In the copying model, neighbors of an existing vertex (selected uniformly at random) are sampled to determine the vertices to connect to.

Bollobas et al. [6] proved the first rigorous result on the power law degree distribution of such graph evolution models, showing that with high probability the degree distribution of the Barabasi–Albert model [3] obeys a power law d^{-3}. Since then, many variants of the preferential attachment model have been proposed by introducing additional parameters that manipulate the probability with which an existing vertex is to be selected. The motivation is to construct models that obey a power law degree distribution with the exponent depending on some adjustable parameters so that a variety of power law distributions observed in the real-world setting can be modelled. Jordan [13] analyzed a slightly generalized model investigated by Dorogovtsev et al. [11] and showed that for large constant $d > 0$, the proportion of vertices of degree d

* Tel.: +1 250 807 9503.
  E-mail address: yong.gao@ubc.ca.

follows a power law $d^{-\gamma}$ with the exponent $\gamma \in (2, \infty)$ determined by two adjustable parameters. Aiello, Chung, and Lu [2] and Cooper and Frieze [9] studied even more general preferential-attachment models with a set of parameters. These parameters specify the number of existing vertices to be selected in each step and control in a probabilistic way how these vertices are selected. A vertex can be selected by sampling uniformly at random from existing vertices or by the preferential attachment mechanism. Among the other results, Cooper and Frieze showed that in their general model the proportion of vertices of degree $d > 0$ follows a power law with the exponent $\gamma \in (2, \infty)$ determined by the model parameters. In all of the preferential attachment models, it is an essential assumption that the vertices to be connected to the new vertex are selected independently of each other.

The first model with copying mechanism for the web graphs is proposed in [15]. A similar model, called the duplication model, arises in the context of biological networks [8]. With the copying mechanism, a new vertex v_{n+1} is connected to a set of existing vertices using the following scheme:

(1) An existing vertex v_i is selected uniformly at random to copy from.
(2) Let $N(v_i)$ be the set of neighbors of v_i in $\{v_1, v_2, \ldots, v_{i-1}\}$. The vertex v_{n+1} is then connected to a subset of $N(v_i)$ selected in a probabilistic fashion. The number of neighbors that v_{n+1} is connected to is called the out-degree of v_{n+1}.

Without any extra work, the above copying mechanism generates a star-like graph centered on the initial graph G_0. To overcome this limitation, Kumar et al. [15] require that the out-degree (i.e., the number of out-edges) is a constant and implement this by connecting the new vertex to either its neighbors or other vertices selected uniformly at random which is crucial for the construction to work. For the case that the out-degree is 1, it was proved in [15] that the in-degree sequence has a power law distribution with high probability.

In the duplication models studied in [8,5], $N(v_i)$ is extended to contain all the neighbors of v_i and each vertex in this extended $N(v_i)$ is connected to v_{n+1} independently with a certain probability. As noted in [5], a correction step has to be employed to avoid the generation of degenerate graph processes. Power law distributions for the expected fraction of vertices of a given degree are proved in [8,5]. Cooper and Frieze [9] use a copying scheme in which the neighbors of v_{n+1} is selected one at a time by repeating the process a number of times independently. This makes the (highly-complicated) analysis more approachable, but spoils to a large extent the idea of the copying mechanism that is intended to capture the phenomenon that neighboring vertices are likely to be connected together to a new vertex.

In this paper, we study a random model for the well-known graph class of k-trees, which may serve as an alternative (and idealized) model in the study of complex networks. The notion of k-trees is a generalization of trees and is closely related to the concept of treewidth in graph theory [14]. We show that the degree distribution of a graph evolution process obtained by a straightforward randomization of the recursive definition of k-trees obeys the power law

$$d^{-(1+\frac{k}{k-1})}$$

with high probability for large d, where k is the parameter that characterizes the degree to which a graph is tree-like. In addition to introducing an alternative model with preferential attachment and copying mechanisms, we hope that the fact that a power law degree distribution exists in such a graph class with quite unique structural characteristics serves as a further cautionary note, reinforcing the viewpoint that a power law degree distribution should not be regarded as the only important characteristic of a complex network, as has been previously argued in [16,17]. We note that in [1], the inherent bias of existing approaches in the empirical study of the Internet graph was identified — it was shown that the widely-used traceroute sampling method "can make power laws appear where none existed in the underlying graphs!"

In the next section, we introduce the construction of the random k-trees and discuss its relation to existing models of complex networks. In Section 3, we prove the power law degree distribution of random k-trees. We conclude in Section 4 with a discussion on the construction of random partial k-trees.

2. Random k-trees: The construction

Throughout this paper, the degree of a vertex v in a graph G is denoted by $\deg_G(v)$. A k-clique of a graph is understood as a complete subgraph on a set of k vertices. All the graphs considered in this paper are undirected.

The construction of a random k-tree is based on the following simple randomization of the recursive definition of k-trees [14]. Starting with an initial clique $G^k(k + 1)$ of size $k + 1$, a sequence of graphs $\{G^k(n), n \geq k + 1\}$ is constructed by adding vertices to the graph one at a time. To construct $G^k(n + 1)$, we add a new vertex v_{n+1} and then connect it to the k vertices of a k-clique selected uniformly at random from all the k-cliques in $G^k(n)$. We call the graph process $\{G^k(n), n \geq k + 1\}$ a k-tree process.

2.1. Relations to existing models

In this subsection, we discuss some basic properties of the k-tree process, including the number of k-cliques in $G^k(n)$ and the probability that an existing vertex of a given degree is connected to a new vertex. These properties are needed in the proof of our main result. They also enable us to illustrate further the relations between the k-tree process and existing graph evolution models.

Let $\mathcal{C}_n$ be the set of cliques of size k in the graph $G^k(n)$. It is easy to see that when a new vertex is added, exactly $\binom{k}{k-1}$ new k-cliques are created and none of the existing k-cliques is destroyed. So, taking into consideration the initial clique of size $k + 1$, we see that the total number of k-cliques in $G^k(n)$ is

$$|\mathcal{C}_n| = (n - k - 1)k + (k + 1).$$

Consider a vertex v in $G^k(n)$. Since every time a new vertex is added and connected to the vertex v, exactly $\binom{k-1}{k-2}$ new k-cliques are created that contain v as one of its vertices, the total number of k-cliques in $G^k(n)$ containing v is

$$n^* = \binom{k}{k-1} + \binom{k-1}{k-2}(\deg_{G^k(n)}(v) - k), \tag{2.1}$$

where the first term is the number of the k-cliques containing v that are created when v is added to the graph and the second term is the total number of k-cliques containing v that are created later on when v is connected to new vertices.

Therefore, given $G^k(n)$ (i.e., conditional on $G^k(n)$), the conditional probability for v to be connected to the new vertex v_{n+1} is

$$\mathbb{P}\left[v \text{ is connected to } v_{n+1} \mid G^k(n)\right] = \frac{n^*}{|\mathcal{C}_n|} = \frac{a_k \deg_{G^k(n)}(v) - b_k}{c_k n} \tag{2.2}$$

where $a_k = k - 1$, $b_k = k(k - 2)$, and $c_k = k - \frac{k^2-1}{n}$.

Note that the above expression only depends on the degree of v in $G^k(n)$. It follows that, given $\deg_{G^k(n)}(v) = d$, the conditional probability for v to be connected to v_{n+1} is

$$f_d^k(n) \overset{\text{def}}{=} \mathbb{P}\left[v \text{ is connected to } v_{n+1} \mid \deg_{G^k(n)}(v) = d\right] = \frac{a_k d - b_k}{c_k n} \tag{2.3}$$

where $a_k = k - 1$, $b_k = k(k - 2)$, and $c_k = k - \frac{k^2-1}{n}$. We see that even though there is no explicit preferential-attachment mechanism employed, Eq. (2.3) shows that the construction scheme does have a similar effect.

2.2. The advantages of random k-trees

The k-tree construction scheme can be viewed as a very rigid copying mechanism; In step $n + 1$, the new vertex v_{n+1} is connected to an existing vertex v_i selected uniformly at random from $\{v_1, v_2, \ldots, v_n\}$ and to a subset of $k - 1$ vertices selected uniformly at random without replacement from the k neighbors that are connected to v_i in step i.

As has been discussed in Section 1, in almost all the existing preferential-attachment models and copying models, there is an essential assumption that old vertices to be connected to a new vertex are selected independently. The random k-tree model studied in the current paper is unique in that these vertices are selected in a highly correlated manner. This captures in a better way the phenomenon that neighboring vertices are more likely to be connected to a new vertex, which is exactly what the copying mechanism tries to model. In addition, the random k-tree has by construction a treewidth k — a structural feature of algorithmic significance that none of the existing models has a mechanism to control.

3. The degree distribution of random k-trees

This section is devoted to proving that for the k-tree process, the proportion of vertices of degree d follows asymptotically a power law $d^{-\gamma}$ with exponent $\gamma = 1 + \frac{k}{k-1}$. Throughout the discussion, we assume that k is a fixed constant. In the following, we use $X_d(n)$ to denote the random variable for the total number of vertices of degree d in $G^k(n)$, and write

$$\alpha_d \triangleq \frac{\Gamma(3 + \frac{2}{k-1})}{\Gamma(1 + \frac{1}{k-1})} \frac{\Gamma(d - \frac{k(k-2)}{k-1})}{\Gamma(d - \frac{k(k-2)}{k-1} + \frac{k}{k-1} + 1)}$$

which, by Stirling's approximation, is approximately

$$e^{-(1 + \frac{k}{k-1})} d^{-(1 + \frac{k}{k-1})}$$

for large d.

Denote by $\mathcal{F}_n = \sigma(G^k(n), n \geq 1)$ the σ-algebra generated by the k-tree process up to time n. We use I_A to denote the indicator function of an event A. To ease the presentation, we use $I_d(i, n)$ to denote the indicator function of the event that the degree of the vertex v_i in $G^k(n)$ is d, i.e.,

$$I_d(i, n) = \begin{cases} 1, & \deg_{G^k(n)}(v_i) = d \\ 0, & \text{otherwise.} \end{cases}$$

The following simple observation will be used to deal with the case $d = k$.

Lemma 3.1. *For any vertex v and $n \geq k+1$, $\deg_{G^k(n)}(v) \geq k$. Furthermore, for $n \geq k+2$ any k-clique in $G^k(n)$ contains at most one vertex with $\deg_{G^k(n)}(v) = k$.*

Proof. The first claim that $\deg_{G^k(n)}(v) \geq k$ follows from the fact that when a new vertex is added, it is connected to the k vertices of the selected k-clique.

We use induction to prove the second claim. First, consider the base case of $n = k+2$. Recall that $G^k(k+2)$ is obtained by connecting a new vertex v_{k+1} to the vertices of a k-clique in the initial $(k+1)$-clique. We see that in $G^k(k+2)$ there are exactly two vertices of degree k, namely the vertex v_{k+1} and one of the vertices in $\{v_1, \ldots, v_k\}$ that is not connected to v_{k+1}. Therefore, no k-clique in $G^k(k+2)$ contains more than one vertex of degree k, and thus the second claim holds for the base case of $n = k+2$.

Assume that the second claim holds for $G^k(n)$. Consider the graph $G^k(n+1)$ obtained from $G^k(n)$. Note that by adding a new vertex v_{n+1} to $G^k(n)$ and connecting it to the vertices of a k-clique in $G^k(n)$, exactly k new k-cliques are created each of which has v_{n+1} as its only vertex of degree k. By the assumption that the second claim holds for $G^k(n)$, no k-clique in $G^k(n+1)$ contains more than one vertices of degree k. This completes the induction step and the second claims follows. ∎

The next theorem shows that the expected degree sequence of the k-tree process obeys a power law distribution.

Theorem 3.1. *Let $\mathbb{E}[X_d(n)]$ be the expected number of vertices with degree d in the random k-tree $G^k(n)$. There exists a constant $N = N(k)$ (independent of d) such that for any $n > N$,*

$$|\mathbb{E}[X_d(n)] - \alpha_d n| \leq C \tag{3.4}$$

where $C = C(k)$ is a constant that is independent of d and n.

The above result is proved by first establishing a recurrence for the expected number $\mathbb{E}[X_d(n)]$ of vertices with a given degree, and then showing that $\mathbb{E}[X_d(n)]$ can be asymptotically approximated by $\beta_d n$ where the sequence $\{\beta_d\}$ is the unique solution to the following simple recurrence relation

$$\beta_d = \frac{a_k(d-1) - b_k}{a_k d - b_k + k} \beta_{d-1}, \quad \beta_k = \frac{1}{2}. \tag{3.5}$$

Recall that to construct the graph $G^k(n+1)$ from $G^k(n)$, a new vertex added to the graph will be connected to all the vertices of a randomly-selected k-clique. This creates a high correlation between the degree of the vertices. A recurrence is still possible due to the fact that the conditional probability for a vertex v to have a degree d in $G^k(n+1)$ given $G^k(n)$ only depends on the degree of v in $G^k(n)$. A detailed account is given in the following proof.

Proof (*Proof of Theorem 3.1*). To begin with, consider the base case $d = k$. Due to Lemma 3.1, we have

$$X_k(n+1) = \begin{cases} X_k(n), & \text{if a k-clique containing a vertex of degree k is selected,} \\ X_k(n) + 1, & \text{otherwise.} \end{cases} \tag{3.6}$$

Let A be the event that a k-clique containing a vertex of degree k is selected in step $n+1$ and let I_A be its indicator function. We have

$$X_k(n+1) = X_k(n)I_A + (X_k(n) + 1)I_{A^c}$$

where A^c is the complement of A.

By Lemma 3.1, a k-clique contains at most one vertex of degree k. It follows that the conditional expectation of I_A (which is equal to the conditional probability of A) is equal to $X_k(n)$, the total number of degree-k vertices in $G^k(n)$, times the conditional probability that a degree-k vertex is selected to be connected to v_{n+1}, i.e.,

$$\mathbb{E}[I_A | \mathscr{F}_n] = f_k^k(n) X_k(n). \tag{3.7}$$

Therefore, by the basic properties of conditional expectation in theory of probability, we have

$$\begin{aligned} \mathbb{E}[X_k(n+1)|\mathscr{F}_n] &= \mathbb{E}[X_k(n)I_A + (X_k(n)+1)I_{A^c}|\mathscr{F}_n] \\ &= \mathbb{E}[X_k(n)I_A|\mathscr{F}_n] + \mathbb{E}[(X_k(n)+1)I_{A^c}|\mathscr{F}_n] \\ &= \mathbb{E}[I_A|\mathscr{F}_n]X_k(n) + \mathbb{E}[I_{A^c}|\mathscr{F}_n](X_k(n)+1) \end{aligned} \tag{3.8}$$

where the last equality is due to the fact that $X_k(n)$ is measurable with respect to $\mathscr{F}_n$ (i.e., in the context of discrete probability space, $X_k(n)$ is a function of $G^k(n)$).

Combining Eqs. (3.7) and (3.8), we have

$$\begin{aligned} \mathbb{E}[X_k(n+1)|\mathscr{F}_n] &= f_k^k(n)X_k(n)X_k(n) + (1 - f_k^k(n)X_k(n))(X_k(n)+1) \\ &= 1 + (1 - f_k^k(n))X_k(n). \end{aligned} \tag{3.9}$$

By the mathematical definition, $\mathbb{E}\left[X_k(n+1)|\mathcal{F}_n\right]$ itself is a random variable measurable with respect to $\mathcal{F}_n$. Recall, from the probability theory, that the unconditional expectation of the conditional expectation of a random variable is equal to the unconditional expectation of the random variable itself. So, we have

$$\mathbb{E}\left[\mathbb{E}\left[X_k(n+1)|\mathcal{F}_n\right]\right] = \mathbb{E}\left[X_k(n+1)\right].$$

Therefore, by taking expectations on both sides of Eq. (3.9), we get the following recurrence

$$\mathbb{E}\left[X_k(n+1)\right] = 1 + (1 - f_k^k(n))\mathbb{E}\left[X_k(n)\right]. \tag{3.10}$$

Solving the above recurrence (3.10) with $\mathbb{E}\left[X_k(k+2)\right] = 2$ gives us

$$\mathbb{E}\left[X_k(n)\right] = \frac{1}{2}n + O(1). \tag{3.11}$$

We now consider the general case of $d > k$. Recall that $I_d(i, n)$ is the indicator function of the event $\{\deg_{G^k(n)}(v_i) = d\}$. The total number of vertices of degree d in $G^k(n)$ can thus be written as $X_d(n) = \sum_{i=1}^{n} I_d(i, n)$. By the additive property of conditional expectation, we have

$$\mathbb{E}\left[X_d(n+1)|\mathcal{F}_n\right] = \sum_{i=1}^{n+1} \mathbb{E}\left[I_d(i, n+1)|\mathcal{F}_n\right]. \tag{3.12}$$

Due to the way in which $G^k(n)$ is constructed, the vertex v_i has degree d in $G^k(n+1)$ if and only if one of the following two situations occurs:

(1) The degree of v_i in $G^k(n)$ is d, and v_i is not selected to be connected to v_{n+1}; or
(2) The degree of v_i in $G^k(n)$ is $d - 1$, and v_i is selected to be connected to v_{n+1}.

Therefore, letting B be the event that v_i is selected to be connected to v_{n+1}, we have

$$I_d(i, n+1) = I_B I_{d-1}(i, n) + I_{B^c} I_d(i, n). \tag{3.13}$$

We claim that

$$\mathbb{E}\left[I_B I_{d-1}(i, n)|\mathcal{F}_n\right] = f_{d-1}^k(n) I_{d-1}(i, n). \tag{3.14}$$

We prove the claim by the mathematical definition of conditional expectation. Consider any event $A \in \mathcal{F}_n$. (Recall that $I_{d-1}(i, n)$ is the indicator function of the event $\{\deg_{G^k(n)}(v_i) = d - 1\}$.) We have

$$\begin{aligned}
\mathbb{E}\left[I_B I_{d-1}(i, n) I_A\right] &= \mathbb{P}\left[B \cap \{\deg_{G^k(n)}(v_i) = d - 1\} \cap A\right] \\
&= \mathbb{P}\left[B \mid \{\deg_{G^k(n)}(v_i) = d - 1\} \cap A\right] \mathbb{P}\left[\{\deg_{G^k(n)}(v_i) = d - 1\} \cap A\right] \\
&= \mathbb{P}\left[B \mid \{\deg_{G^k(n)}(v_i) = d - 1\}\right] \mathbb{P}\left[\{\deg_{G^k(n)}(v_i) = d - 1\} \cap A\right] \\
&= f_{d-1}^k(n)\mathbb{E}\left[I_{d-1}(i, n) I_A\right],
\end{aligned}$$

where the second last equality is due to the fact that the event $\{\deg_{G^k(n)}(v_i) = d\}$ completely determines the (conditional probability of) the event B. The claim then follows from the mathematical definition of conditional expectation.

Similarly, we have

$$\mathbb{E}\left[I_{B^c} I_d(i, n)|\mathcal{F}_n\right] = (1 - f_d^k(n)) I_d(i, n). \tag{3.15}$$

Combining Eqs. (3.13)–(3.15), we see that for any $i < n + 1$,

$$\mathbb{E}\left[I_d(i, n+1)|\mathcal{F}_n\right] = f_{d-1}^k(n) I_{d-1}(i, n) + (1 - f_d^k(n)) I_d(i, n). \tag{3.16}$$

Also note that for $i = n + 1$, by the construction of $G^k(n)$ we have

$$\mathbb{E}\left[I_d(n+1, n+1)|\mathcal{F}_n\right] = 0$$

for any $d > k$. Summing over i on both sides of Eq. (3.16) and based on Eq. (3.12), we have

$$\mathbb{E}\left[X_d(n+1)|\mathcal{F}_n\right] = f_{d-1}^k(n) X_{d-1}(n) + (1 - f_d^k(n)) X_d(n). \tag{3.17}$$

Recall that the unconditional expectation of the condition expectation of a random variable is equal to the unconditional expectation of the random variable itself. Taking unconditional expectations on both sides of Eq. (3.17), we get the following recurrence equation for the expected number of vertices of degree d:

$$\mathbb{E}\left[X_d(n+1)\right] = f_{d-1}^k(n)\mathbb{E}\left[X_{d-1}(n)\right] + (1 - f_d^k(n))\mathbb{E}\left[X_d(n)\right]. \tag{3.18}$$

Using the recurrence equation (3.18) and the base case equation (3.11), we now prove that $|\mathbb{E}\left[X_d(n)\right] - \beta_d n|$ is asymptotically upper bounded by a constant, where the sequence $\{\beta_d\}$ is the unique solution to the following simple recurrence equation

$$\beta_d = \frac{a_k(d-1) - b_k}{a_k d - b_k + k}\beta_{d-1}, \quad \beta_k = \frac{1}{2}. \tag{3.19}$$

Let $\epsilon_d^n = \mathbb{E}\left[X_d(n)\right] - \beta_d n$. For the base case $d = k$, we have from Eq. (3.11) that $\epsilon_k^n = O(1)$. For the general case $d > k$, we have from Eq. (3.18) that

$$\epsilon_d^{n+1} = f_{d-1}^k(n)\epsilon_{d-1}^n + f_{d-1}^k(n)\beta_{d-1}n + (1 - f_d^k(n))\epsilon_d^n + (1 - f_d^k(n))\beta_d n - (n+1)\beta_d. \tag{3.20}$$

By the definition of β_d (Eq. (3.19)), we see that

$$\begin{aligned}
f_{d-1}^k(n)\beta_{d-1}n + (1 - f_d^k(n))\beta_d n - (n+1)\beta_d &= \frac{a_k(d-1) - b_k}{c_k}\beta_{d-1} - \frac{a_k d - b_k}{c_k}\beta_d - \beta_d \\
&= \frac{a_k d - b_k + k}{c_k}\beta_d - \frac{a_k d - b_k}{c_k}\beta_d - \beta_d \quad \text{(using (3.19))} \\
&= \frac{k - c_k}{c_k}\beta_d \\
&= \frac{k^2 - 1}{c_k n}\beta_d \quad \text{(since } c_k = k - (k^2 - 1)/n).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
|\epsilon_d^{n+1}| &\leq f_{d-1}^k(n)|\epsilon_{d-1}^n| + (1 - f_d^k(n))|\epsilon_d^n| + \frac{k^2 - 1}{nc_k}\beta_d \\
&\leq (1 + f_{d-1}^k(n) - f_d^k(n))\max(|\epsilon_{d-1}^n|, |\epsilon_d^n|) + \frac{k^2 - 1}{nc_k}\beta_d.
\end{aligned} \tag{3.21}$$

From (3.19), we see that $\beta_d \leq \beta_{d-1}$ for any $d > k$. Since by (3.20) $\epsilon_k^n = O(1)$ and since

$$f_{d-1}^k(n) - f_d^k(n) = -\frac{a_k}{nc_k} < 0$$

by the definition of $f_d^k(n)$, we can use (3.21) to prove by induction that there exists a constant $N = N(k) > 0$ independent of d such that for any $n > N$, $|\epsilon_d^n|$ is bounded by a constant $C = C(k)$ independent of d and n, and therefore

$$\mathbb{E}\left[X_d(n)\right] = \beta_d n + O(1). \tag{3.22}$$

To complete the proof of Theorem 3.1, we see from the definition of β_d that

$$\begin{aligned}
\beta_d &= \prod_{l=k}^{d}\frac{a_k(l-1) - b_k}{a_k l - b_k + k} = \prod_{l=k}^{d}\frac{l - 1 - \frac{b_k}{a_k}}{l - \frac{b_k}{a_k} + \frac{k}{a_k}} \\
&= \frac{\Gamma(k - \frac{b_k}{a_k} + \frac{k}{a_k} + 1)}{\Gamma(k - \frac{b_k}{a_k})}\frac{\Gamma(d - \frac{b_k}{a_k})}{\Gamma(d - \frac{b_k}{a_k} + \frac{k}{a_k} + 1)} \\
&= \frac{\Gamma(3 + \frac{2}{k-1})}{\Gamma(1 + \frac{1}{k-1})}\frac{\Gamma(d - \frac{b_k}{a_k})}{\Gamma(d - \frac{b_k}{a_k} + \frac{k}{a_k} + 1)}
\end{aligned} \tag{3.23}$$

which by Stirling's approximation is approximately $e^{-(1+k/a_k)}d^{-(1+k/a_k)}$ for large d. ∎

Next, we show that $X_d(n)$, the number of vertices of degree d, concentrates on its expectation, which together with Theorem 3.1, establishes the power law degree distribution of the k-tree process.

Theorem 3.2. *Let $X_d(n)$ be the total number of vertices of degree d in $G^k(n)$. For any $\lambda > 0$, we have*

$$\mathbb{P}\left[|X_d(n) - \mathbb{E}\left[X_d(n)\right]| > \lambda\right] \leq e^{-\frac{\lambda^2}{8kn}}. \tag{3.24}$$

Proof. Consider the martingale $\{Z_i \triangleq \mathbb{E}\left[X_d(n)|\mathcal{F}_i\right], i \geq k + 1\}$ and the associated martingale difference sequence $\{Z_{i+1} - Z_i, i \geq k + 1\}$. If we can show that

$$|Z_{i+1} - Z_i| \leq 2k,$$

then an application of Azuma's Inequality (see, e.g. Theorem 7.4.2 of [4]) gives (3.24).

For each i, let $\mathcal{C}_i$ be the collection of size-$(k+1)$ vertex subsets of the form $\{v_{i_1}, \ldots, v_{i_k}, v_i\}$ where $i_1 < i_2 < \cdots < i_k < i$. $\mathcal{C}_i$ is the collection of the possible $(k+1)$-cliques in $G^k(i)$ that contain v_i as one of their vertices. We call v_i the head of a subset $C \in \mathcal{C}_i$ and write $head(C) = v_i$.

Now consider the probability space $(\Omega, \mathbb{P}[\cdot])$ defined over the product space $\Omega = \prod_{i=k+1}^{n} \mathcal{C}_i$. A sample point $h = \{h_i\} \in \Omega$ is said to be a *realization* of a k-tree if for any $k + 1 \leq i \leq n$, the vertex subset $h_i = \{v_{i_1}, \ldots, v_{i_k}, v_i\} \in \mathcal{C}_i$ is such that $\{v_{i_1}, \ldots, v_{i_k}\}$ is a subset of h_j for some $j < i$.

Let $\Omega_0 \subset \Omega$ be the set of sample points that are realizations of a k-tree. The probability measure $\mathbb{P}[\cdot]$ is defined as follows. It has Ω_0 as its support and for each $h \in \Omega_0$,

$$\mathbb{P}[h] \stackrel{\text{def}}{=} \prod_{i=k+1}^{n-1} \frac{1}{k+1} \frac{1}{i-k}.$$

The reason for $\mathbb{P}[h]$, where $h = (h_i)$, to be defined as in the above is explained as follows. Let G be the k-tree on the vertex set $\{v_1, \ldots, v_n\}$ such that for each $k + 2 \leq i \leq n$, when v_i is added, it is connected to a subset of k vertices from some h_j where $k + 1 \leq j < i$. The probability that the random k-tree $G^k(n)$ obtained according to our construction is equal to G is

$$\mathbb{P}\left[G^k(n) = G\right] = \prod_{i=k+1}^{n-1} \frac{1}{k+1} \frac{1}{i-k},$$

where the term $\frac{1}{k+1}\frac{1}{i-k}$ is the conditional probability (given $G^k(i-1)$) that a specific size-k vertex subset in a specific h_j is selected to be connected to v_{i+1}.

For any $g \in \Omega_0$, writing $X_d(n, g)$ for the total number of vertices of degree d in the k-tree realized by g, we have

$$Z_i(h) = \mathbb{E}\left[X_d(n, g)|g_j = h_j, \forall k + 1 \leq j \leq i\right].$$

The following argument is motivated by a similar one used in [9]. Let $h = \{h_{k+1}, \ldots, h_n\} \in \Omega_0$ be a sample point and $H_{i+1}[h] \subset \Omega_0$ be the collection of sample points that agree with h for $k + 1 \leq j \leq i + 1$, i.e.,

$$H_{i+1}[h] = \{g \in \Omega_0 | g_j = h_j, \forall k + 1 \leq j \leq i + 1\}.$$

Consider a size-$(k + 1)$ vertex set $C \in \mathcal{C}_{i+1}$ such that $C \neq h_{i+1}$. Define $H_{i+1}[C]$ to be the collection of the sample points $g \in \Omega_0$ such that

$$\begin{cases} g_j = h_j, & \forall k + 1 \leq j \leq i \\ g_{i+1} = C. \end{cases} \tag{3.25}$$

We claim that there is a one-to-one correspondence between the elements of $H_{i+1}[h]$ and $H_{i+1}[C]$.

Assume that $h_{i+1} = \{v_{j_1}, \ldots, v_{j_k}, v_{i+1}\}$ and $C = \{v_{l_1}, \ldots, v_{l_k}, v_{i+1}\}$. The claimed one-to-one correspondence can be shown by the mapping defined as follows. For each $1 \leq p \leq k$, define $\sigma_{i+1}(v_{j_p}) = v_{l_p}$. For each $g \in H_{i+1}[h]$, define $\sigma(g) = g' \in H_{i+1}[C]$ as

(1) $g'_j = h_j$ for any $j \leq i$;
(2) $g'_{i+1} = C$; and
(3) for each $j > i + 1$, g'_j is a size-$(k + 1)$ vertex subset defined as
   (a) if g_j doesn't contain the vertex v_{i+1}, then $g'_j = g_j$, and
   (b) if g_j contains v_{i+1}, then g'_j is obtained by replacing each vertex $v \in g_j \cap h_{i+1}$ with $\sigma_{i+1}(v)$.

For any $g \in H_{i+1}[h]$, since the only vertices whose vertex degree might have been changed by the mapping σ are those in $\{v_{j_1}, \ldots, v_{j_k}\}$ and $\{v_{l_1}, \ldots, v_{l_k}\}$, we have

$$|X_d(n, g) - X_d(n, \sigma(g))| \leq 2k.$$

Since the probability measure $\mathbb{P}[\cdot]$ assigns equal probability to the sample points,

$$|Z_{i+1} - Z_i| \leq 2k, \quad \forall h \in \Omega_0,$$

holds due to the definition of conditional expectation:

$$Z_{i+1} = \mathbb{E}[X(n, h)|\mathcal{F}_{i+1}] = \sum_{g \in H_{i+1}[h]} X(n, g)\mathbb{P}\left[g|g_j = h_j, j \leq i + 1\right]$$

and

$$Z_i = \mathbb{E}[X(n, h)|\mathcal{F}_i] = \sum_{C \in \mathcal{C}_{i+1}} \sum_{g \in H_{i+1}[C]} X(n, g)\mathbb{P}\left[g|g_j = h_j, j \leq i\right].$$

This completes the proof. ∎

4. Concluding remarks

In this paper, we have shown that a simple evolving graph model based on the notion of *k*-trees has a power law degree distribution with high probability. Due to its simplicity and unique structures, we think this model of evolving graphs provides a useful alternative in the modeling, analysis, and simulations of complexity networks such as the web graphs that have attracted much attention [7]. The fact that a power law degree distribution exists in such a graph models with quite unique characteristics also serves as a further cautionary note, reinforcing the viewpoint that a power law degree distribution should not be regarded as the only important feature of a complex network, as has been previously argued in [1,16,17].

A partial *k*-tree is a subgraph of a *k*-tree. To enrich the modelling power of the class of models, it is desirable to have a natural model of random partial *k*-trees. It is tempting to think of the following model based on the construction of the *k*-tree process: For each vertex v in $G^k(n)$, delete randomly-selected $(1 - b) * k$ of its *k* out-edges for some $0 < b < 1$. In [12], we claimed that a model of random partial *k*-trees obtained in this way has a power law degree distribution $d^{-(1+\frac{k}{b*(k-1)})}$, which turns out to be flawed. A few alternatives have since then been investigated, resulting in very unnatural random models. We leave it as an open question the existence of a natural evolution model for the partial *k*-trees.

References

[1] D. Achlioptas, A. Clauset, D. Kempe, C. Moore, On the bias of traceroute sampling, or power-law degree distribution in regular graphs, in: Proceedings of the 37th ACM Symposium on Theory of Computing, STOC'05, 2005, pp. 694–703.
[2] W. Aiello, F. Chung, L. Lu, Random evolution of massive graphs, in: Proceedings of the 42nd Annual Symposium on Foundations of Computer Science, 2001, pp. 510–519.
[3] R. Albert, A. Barabási, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (47) (2002).
[4] N. Alon, J.H. Spencer, The Probabilistic Method, Wiley, 2000.
[5] G. Bebek, P. Berenbrink, C. Cooper, T. Friedetzky, J. Nadeau, S. Sahinalp, The degree distribution of the generalized duplication model, Theoretical Computer Science 369 (2006) 234–249.
[6] B. Bollobas, O. Riordan, J. Spencer, G. Tusnady, The degree sequence of a scale-free random graph process, Random Structures and Algorithms 18 (2001) 279–290.
[7] D. Chakrabarti, C. Faloutsos, Graph mining: Laws, generators, and algorithms, ACM Computing Survey 38 (1) (2006) 2–69.
[8] F. Chung, L. Lu, T. Dewey, D. Galas, Duplication models for biological networks, Journal of Computational Biology 10 (2003) 677–687.
[9] C. Cooper, A. Frieze, A general model of web graphs, Random Structures and Algorithms 22 (2003) 311–335.
[10] S. Dorogovtsev, J. Mendes, Evolution of networks, Advances in Physics 51 (2002) 1079–1187.
[11] S. Dorogovtsev, J. Mendes, A. Samukhin, Structure of growing networks with preferential linking, Physical Review Letters 85 (2000) 4633.
[12] Y. Gao, C. Hobson, Random *k*-tree as a model for complex networks, in: The Fourth Workshop on Algorithms and Models for the Web-Graph, WAW2006, 2006.
[13] J. Jordan, The degree sequences and spectra of scale-free random graphs, Random Structures and Algorithms 29 (2) (2006) 226–242.
[14] T. Kloks, Treewidth: Computations and Approximations, Springer-Verlag, 1994.
[15] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, E. Upfal, Stochastic models for the web graph, in: Proceedings of the 41st IEEE Symp. on Foundations of Computer Science, 2000, pp. 57–65.
[16] L. Li, D. Alderson, J. Doyle, W. Willinger, Towards a theory of scale-free graphs: Definition, properties, and implications, Internet Mathematics 2 (4) (2005) 431–523.
[17] M. Mitzenmacher, The future of power law research, Internet Mathematics 2 (4) (2005) 525–534.