k-木のスケールフリー性に関する研究

Colin Cooper† (colin.cooper@kcl.ac.uk)
Department of Computer Science, King's College, United Kingdom.

上原 隆平‡ (uehara@jaist.ac.jp)
北陸先端科学技術大学院大学 情報科学研究科

Abstract

近年，WWW やインターネットなどの非一様な構造を持つ多くの社会ネットワークを説明できるモデルとしてスケールフリーグラフが注目を集めている．本研究では，スケールフリーな k-木を生成する単純なランダムモデルを提案する．k-木とは，固定された任意の正整数パラメータ k に対する一般化された木であり，グラフマイナーの分野では基本的な概念である．本稿で提案するモデルは非常に自然で単純な規則から生成される．最初に大きさ $k+1$ の極大クリークを一様ランダムに選び，次にそのクリークの中の k 個の頂点を一様ランダムに選び，そしてこれら k 個の頂点に隣接するように新しい頂点を追加するだけである．つまり本モデルでは，新しい頂点を 1 個追加するときに一様ランダムな選択を 2 回行なうだけでよい．このとき結果として得られる k-木において，頂点の次数分布は漸近的にベキ乗則に従い，クラスタ係数は大きく，そしてグラフの直径は小さい．ランダムな選択が単純なので，こうした性質の解析も容易である．また得られた k-木の直径は実験的結果により $o(\log n)$ であることが確認できた．ただしここで n は頂点数であり，$o(1)$ 項は k の関数である．
キーワード: スケールフリーグラフ，スモールワールド，k-木

On Scale free k-trees

Colin Cooper† (colin.cooper@kcl.ac.uk)
Department of Computer Science, King's College, United Kingdom.

Ryuhei Uehara‡ (uehara@jaist.ac.jp)
School of Information Science, JAIST

Abstract

Scale free graphs have attracted attention as their non-uniform structure that can be used as a model for many social networks including the WWW and the Internet. In this short note, we propose a simple random model for generating scale free k-trees. For any fixed integer k, a k-tree consists of a generalized tree parameterized by k, and is one of the basic notions in the area of graph minors. Our model is quite simple and natural; it first picks a maximal clique of size $k+1$ uniformly at random, it then picks k vertices in the clique uniformly at random, and adds a new vertex incident to the k vertices. That is, the model only makes uniform random choices twice per vertex. Then (asymptotically) the distribution of vertex degrees in the resultant k-tree follows a power law, the k-tree achieves a large clustering coefficient, and the diameter is small. Since the random process is simple, so is the analysis of these properties. Moreover, our experimental results indicate that the resultant k-trees have extremely small diameter, proportional to $o(\log n)$, where n is the number of vertices in the k-tree, and the $o(1)$ term is a function of k.
Keywords: scale free graph, small world network, k-tree.

†Research supported by Great Britain Sasakawa Foundation Grant 3432.
‡Research supported by Grants-in-Aid for Scientific Research.

Vol.2009-AL-124 No.2
2009/5/11

1 Introduction

Small world networks are the focus of recent interest because of their potential as models for interaction networks of complex systems in real world since early works by Watts & Strogatz [7] and Barabási & Albert [1]. There are three properties of a graph G to characterize it to be called small world networks or scale free networks (see, e.g., [5]). These properties which we denote by SF, CC, SW are as follows: (SF) The degree distribution of G follows a power law distribution. That is, the number of vertices of degree i is proportional to i^α for some fixed α. It is known that α is between 2 and 3 in the real social networks. (CC) Two neighbors of any node of G are also likely to joined by an edge. More precisely, the clustering coefficient $CC(v)$ at v is defined as follows:

$$CC(v) = \frac{|\{u \sim w : u, w \in N(v)\}|}{\binom{d(v)}{2}},$$

where $u \sim w$ means that they are joined by an edge. The clustering coefficient $CC(G)$ of the graph G is the average clustering coefficient $CC(v)$ for all vertices v in G. (SW) Two nodes of G are connected by a relatively short path. Though many models for generating graphs have been proposed and investigated, there are few models that satisfy all the properties. Moreover, it is not easy to see the combinatorial structure of the graphs obtained, and analysis of the properties is rather complicated.

Recently, Miyoshi et al. propose a model of scale free graphs based on time sequential data [6]; their model, called *scale free interval graph*, employs interval graphs as basic graphs. A graph is an interval graph if and only if there is a one-to-one mapping between vertices and intervals such that two vertices are joined by an edge if and only if the corresponding intervals share a common point. In their model, each vertex in the graph corresponds to a time period, and its lifespan is determined by a simple rule: longer life tends to survive in the next generation. More precisely, if an interval has a length k at time t, it will grow to a length $k+1$ at time $t+1$ with probability $\frac{1}{\zeta(\alpha)}(k+1)^{-\alpha}$, where $\alpha > 2$ is any positive constant and $\zeta(\alpha) = \sum_{i=1}^{\infty} i^{-\alpha}$ (the Riemann's zeta function). A scale free interval graph satisfies two properties (SF) and (CC) of small world networks with high probability. The analysis of the scale free interval graphs is simpler than the other models.

In this short note, we propose a simple model to generate scale free k-trees which satisfy all three properties. In the area of graph algorithms, k-trees form a well known graph class that generalizes trees and plays an important role in graph minor area (see [2, 3] for further details). There are several equivalent definition of k-trees, and we employ one of them as follows; for any fixed positive integer k, (0) a complete graph K_k of k vertices is a k-tree, (1) for a k-tree G of n vertices, a new k-tree G' of $n+1$ vertices is obtained by adding a new vertex v incident to a clique of size k in G. We note that a complete graph K_{k+1} of $k+1$ vertices is a k-tree, which is obtained by adding a vertex to K_k.

For each time $t = 1, 2, \ldots$, our model is an algorithm that generates a sequence of k-trees of $k+t$ vertices as follows. By the definition, it is clear that $G_k(t)$ is a k-tree of $k+t$ vertices. We remark that the algorithm only makes uniform random choices twice to make one k-tree. Let $\boldsymbol{X}(t)$ be a random variable, and $X(t) = \mathbf{E}\boldsymbol{X}(t)$ be its expectation, then $\lim_{t\to\infty} X(t)/t$ is the *limiting expected proportion* of $\boldsymbol{X}(t)$. The *limiting expected clustering coefficient* $c(k)$ is defined by

$$c(k) = \lim_{t\to\infty} \mathbf{E}CC(G_k(t)).$$

Our main theorem states that the simple combination of two uniform random choices makes the graph be scale free with properties (SF) and (CC) like a scale free interval graph:

Algorithm 1: Generation of k-trees

Input : Positive integer k.
Output: A series of k-trees $G_k(1), G_k(2), \ldots$.

begin
  $t = 1$; let $G_k(t)$ be K_{k+1}; output $G_k(t)$;
  for $t = 2, 3, \ldots$ **do**
    pick $D_t = K_{k+1}$ from $G_k(t-1)$ uniformly at random with probability $\frac{1}{t}$;
    pick $f_t = K_k$ from D uniformly at random with probability $\frac{1}{k+1}$;
    let $G_k(t)$ be the graph obtained from $G_k(t-1)$ by adding a new vertex v_t incident to every vertex in f_t;
    output $G_k(t)$;
  end
end

Theorem 1 *Let $k \geq 2$. For a graph $G_k(t)$, we denote by n_i the number of vertices of degree i. Then the graph $G_k(t)$ has the following properties.*

1. *The limiting expected proportion n_i of vertices of degree $i = k + \ell - 1$ is given by*

$$n_{k+\ell-1} = \frac{(\ell-1)!k^{\ell-1}(k+1)}{\prod_{j=1\ldots\ell}((j+1)k+1)}.$$

   This expression has power law asymptotic

$$n_i \propto i^{-(2+1/k)}.$$

2. *The limiting expected clustering coefficient $c(k)$ is given by*

$$c(k) = \sum_{\ell \geq 1} \frac{\binom{k}{2} + (k-1)(\ell-1)}{\binom{(k-1)+\ell}{2}} \frac{(\ell-1)!k^{\ell-1}(k+1)}{\prod_{j=1\ldots\ell}((j+1)k+1)}.$$

   For $k \geq 2$, $c \geq 1/2$ and $c(k) \to 1$ if $k \to \infty$.

We give a short combinatorial proof of the main theorem, and note the following theorem for the finite process $G_k(t)$. We say a sequence of events $\mathcal{E}_t$ occurs *with high probability* (**whp**) if $\lim_{t\to\infty} \Pr(\mathcal{E}_t) = 1$.

Theorem 2 *The following properties hold* **whp**

1. *Let $\boldsymbol{N}(i,t)$ denote the number of vertices of degree i in $G_k(t)$. Then $\boldsymbol{N}(i,t) = tn_i(1+o(1))$ for $i \leq t^a$, where a is some positive constant.*
2. *$CC(G_k(t)) = c(k)(1 + o(1))$.*
3. *The diameter of $G_k(t)$ is $O(\log t)$.*

We also study k-trees of finite size experimentally. We show the resultant k-tree has the property (SW), and hence it achieves a small world. Precisely, the experimental results indicate that the diameter of the resultant k-tree of n vertices is proportional to $o(\log n)$, as k increases. This is an advantage of the scale free interval graphs in [6]; their model generates scale free interval graphs of n vertices width diameter $\Theta(n)$.

We assume that the reader is familiar to the notion of probability and graph theory. In this short note we prove Theorem 1, briefly discuss Theorem 2 and give experimental results supporting Theorem 1 and our hypothesis regarding diameter.

2 Proof of Theorem 1

To prove the theorem, we first show the following lemma:

Lemma 3 *Let v_t be the vertex added to $G_k(t)$ at time t. For any $t' \geq t$, let ℓ be the number of K_{k+1} that contain v_t. Then the clustering coefficient at v_t in $G_k(t')$ is*

$$CC(v_t) = \frac{\binom{k}{2} + (k-1)(\ell-1)}{\binom{(k-1)+\ell}{2}}.$$

Proof. Suppose that at time $t > 1$, we add a vertex v_t and join it to each vertex $u_1, \ldots, u_k$ in the clique f_t of size k chosen in step 5 in the clique D_t of size $k+1$ chosen in step 4. Then $G_k(t)$ contains $k+t$ vertices. We call each induced clique K_k in $G_k(t')$ a *face* of $G_k(t')$, and define the *degree* of a face f by the number of K_{k+1} containing f, that is denoted by $\deg_{t'}(f)$. At time t, we add a new clique $Q = K_{k+1}$ by joining v_t to an existing face f_t. Thus $\deg_t(f_t) = \deg_{t-1}(f_t) + 1$ since f_t is in Q.

We define *face degree* $\mathrm{Deg}_{t'}(v)$ of a vertex v by the total face degree of all faces incident with v. That is, $\mathrm{Deg}_{t'}(v) = \sum_{v \in f} \deg_{t'}(f)$. Initially, when v_t is added at time t, $\mathrm{Deg}_t(v_t) = k$ as there are k faces containing v_t, i.e., $Q = K_{k+1}$ contains k K_k subgraphs with distinguished vertex v (delete any of the k edges incident with v). Extending a face f to K_{k+1} adds one to $\deg(f)$ (since it is now in an extra K_{k+1}) and $k-1$ extra faces at v_t of face degree 1. Thus $\mathrm{Deg}_{t'}(v_t) = k\ell$, where ℓ be the number of K_{k+1} that contain v_t.

At time t', we denote the set of neighbors of v by $N_{t'}(v)$, and define $\mathrm{d_{t'}(v) = |N_{t'}(v)|}$ (that is, $\mathrm{d_{t'}}$ is the ordinary degree of v in $G_k(t)$). When v_t is added to $G_k(t)$, we have $\mathrm{d_t(v_t) = Deg_t(v_t) = k}$. Each time a face containing v_t is extended the face degree of v_t increases by k, but the vertex degree of v_t only increases by 1. Hence $\mathrm{d_{t'}(v_t) = (k-1) + \frac{Deg_{t'}(v_t)}{k}}$.

Now we define *triangle degree* Δ_v of v by the number of K_3 in the subgraph induced by $\{v\} \cup N(v)$. That is, $CC(v)$ is given by $\frac{\Delta_v}{\binom{d(v)}{2}}$. Initially, when v_t is added to $G_k(t)$ it is contained in a unique $K_{k+1}(= Q)$, and the k edges incident at v induce $\binom{k}{2}$ triangles. Suppose face $f_t = K_k$, incident with v_t, is extended to a K_{k+1} at step t'. Face f already has $k-1$ edges vu_i with $i = 1, \ldots, k-1$, each of which will form a new triangle $(v_{t'}v_t, v_{t'}u_i, v_tu_i)$ with the new vertex $v_{t'}$. Thus $\Delta_{v_t} = \binom{k}{2} + (k-1)(d_{t'}(v) - k)$. Therefore if $\mathrm{Deg}_{t'}(v_t) = k\ell$ then $d_{t'}(v_t) = (k-1) + \ell$ and $\Delta_{v_t} = \binom{k}{2} + (k-1)(\ell-1)$. Since $CC(v_t) = \frac{\Delta_{v_t}}{\binom{d(v_t)}{2}}$, the lemma follows. ∎

We now turn to the clustering coefficient of a graph $G_k(t)$, which is defined by $CC(G_k(t)) = \sum_v \frac{CC(v)}{k+t}$. Let $f_{\ell k}$ be the limiting proportion of vertices of face degree ℓk and $n_{k-1+\ell}$ the limiting proportion of vertices of degree $k-1+\ell$. Then we have $f_{\ell k} = n_{k-1+\ell}$ and hence

$$\lim_{t \to \infty} CC(G(t)) = \sum_{\ell \geq 1} f_{\ell k} \frac{\binom{k}{2} + (k-1)(\ell-1)}{\binom{(k-1)+\ell}{2}}. \tag{1}$$

Next we analyze $f_{\ell k}$:

Lemma 4

$$f_{\ell k} = \frac{(\ell-1)! k^{\ell-1}(k+1)}{\prod_{j=1\ldots\ell}((j+1)k+1)}. \tag{2}$$

Proof. For $t \geq t'$, the relationship between vertex degree and face degree of a vertex $v_{t'}$ is given by $d_t(v_{t'}) = (k-1) + \frac{\mathrm{Deg}_t(v_{t'})}{k}$. Thus it suffices to study face degree $\mathrm{Deg}_t(v_{t'})$ of the vertices $v_{t'}$ of $G_k(t)$.

Let $\boldsymbol{F}_i(G_k(t))$ be the number of vertices of face degree i in $G_k(t)$ at the end of time t, and let $F_i(t)$ be its expected value.

Recall that we make $G_k(t+1)$ from $G_k(t)$ by picking a $D_{t+1} = K_{k+1}$ uniformly at random from $G_k(t)$ with probability $\frac{1}{t}$, and then picking a face $f_{t+1} = K_k$ uniformly at random from D_{t+1} with probability $\frac{1}{k+1}$. This process in fact picks faces proportional to their degree. This can be seen as follows. Suppose face f has degree i and thus occurs in i distinct K_{k+1}. Then

$$\Pr(f \text{ is chosen}) = \frac{i}{(k+1)t}.$$

Similarly, $\Pr(\text{face incident with } v \text{ chosen}) = \frac{\mathrm{Deg}_t(v)}{(k+1)t}$.

On adding vertex v_{t+1}, the number of vertices of face degree i is updated as follows:

$$\begin{aligned}\boldsymbol{F}_i(G_k(t+1)) = \boldsymbol{F}_i(G_k(t)) + 1(i=k) + \sum_{\mathrm{Deg}_t(v)=i-k} 1(v \text{ is in chosen face}) \\ - \sum_{\mathrm{Deg}_t(v)=i} 1(v \text{ is in chosen face}),\end{aligned}$$

where $1(H)$ is the indicator for the event H. On taking expectations over the random choices made by the process on the given graph $G_k(t)$, we obtain

$$F_i(G_k(t+1)) = \boldsymbol{F}_i(G_k(t)) + \frac{(i-k)\boldsymbol{F}_{i-k}(G_k(t))}{(k+1)t} - \frac{i\,\boldsymbol{F}_i(G_k(t))}{(k+1)t} + 1(i=k).$$

On taking expectations over all processes $G_k(t)$, we obtain the following recurrences, which are valid for $i = \ell k$, $\ell \geq 1$.

$$\begin{aligned} F_k(t+1) &= F_k(t) + 1 - \frac{kF_k(t)}{(k+1)t} \\ F_i(t+1) &= F_i(t) + \frac{(i-k)F_{i-k}(t)}{(k+1)t} - \frac{iF_i(t)}{(k+1)t} \qquad (i > k). \end{aligned}$$

Now we use the following lemma on real sequences [4, Lemma 3.1]:

Lemma 5 ([4, Lemma 3.1]) *If (α_t), (β_t) and (γ_t) are real sequences satisfying the relation*

$$\alpha_{t+1} = \left(1 - \frac{\beta_t}{t}\right)\alpha_t + \gamma_t,$$

where $\lim_{t\to\infty} \beta_t = \beta > 0$ and $\lim_{t\to\infty} \gamma_t = \gamma$, then $\lim_{t\to\infty} \frac{\alpha_t}{t}$ exists and equals $\frac{\gamma}{1+\beta}$.

Using Lemma 5, we have

$$\lim \frac{F_{\ell k}(t)}{t} = \frac{(\ell-1)!k^{\ell-1}(k+1)}{\prod_{j=1\ldots\ell}((j+1)k+1)} = f_{\ell k}.$$

∎

Theorem 1(i) now follows Lemma 4, and taking the limit of equation (2) gives the claimed power law. Theorem 1(ii) follows from inserting equation (2) into relationship equation (1). It can be see directly that for $k \geq 2$, $c(k) \geq \frac{1}{2}$. For the value of $c(k)$, when $k \to \infty$, we see that

$$c(k) \to \sum_{\ell \geq 1} \frac{1}{\ell(\ell+1)} = 1.$$

3 Proof of Theorem 2

We give a brief outline of the proof. Recall that $\boldsymbol{F}_i(G_k(t))$ be the number of vertices of face degree i in $G_k(t)$ at the end of time t. The **whp** convergence of $\boldsymbol{F}_i(G_k(t))$ to $f_i t(1+o(1))$ can be established by standard methods e.g. [8]. This holds for $i \le t^a$, where a is some positive constant. This establishes that the proportion of vertices of degree i in the finite process $G_k(t)$ is close to its limiting value. The value of the clustering coefficient follows directly from this.

As regards the diameter, a crude calculation suffices to establish a **whp** upper bound of $O(\log t)$. Consider a shortest (edge) path $v_t, u_1, ..., u_i, v_0$ back from v_t to a root vertex v_0 in $G_1(t)$. As half of the K_{k+1} in $G_k(t)$ were added by time $t/2$,

$$\Pr(v_t \text{ chooses a face } f \text{ in } G_k(t/2)) = \frac{\deg_t(G_k(t/2))}{(k+1)t} \ge \frac{1}{2}.$$

Thus the expected distance to the root must be (at least) halved by the edge $v_t u_1$. Whatever the label s of $u_1 = v_s$, this halving occurs independently at the next step. This must terminate **whp** after $c \log t$ steps, for some suitably large constant c, as we now prove.

Let Z_i be an indicator variable for the event that the distance to the root halves at step i, (conditional on not being at the root), or $Z_i = 1$ identically, if we have arrived at the root. Then $\Pr(Z_i = 1) \ge 1/2$, and $S_j = Z_1 + \cdots + Z_j$ stochastically dominates the binomial random variable $B \sim Bin(j, 1/2)$. As $\Pr(B < j/4) = O(e^{-j/16})$, then after $j = c \log_2 t$ steps, where $c > 4$ we conclude **whp** that we have arrived. Thus **whp** $\textsc{diam}(G_k(t)) = O(\log t)$.

4 Experimental Results

Algorithm 1 can be implemented easily. In this section, we give experimental results for the three properties (SF, CC, SW) of scale free networks and small world networks. Although scale free property and cluster coefficients are checked on a usual PC, we used a supercomputer (SGI Altix 4700: 96 Processors with 2305GB Memory) to experiment for small world property with huge n.

Small world property

The first property is the property of a small world. This property implies that any two nodes on the network is connected by a relatively short path. The experimental results are shown in Figure 1. The figure implies that any pair of two nodes in a scale free k-tree of n vertices in our model seems to be joined by a very short path, possibly even of length less than $O(\log n / \log k)$. To observe this, we also plot the number of vertices and the value of (diameters$\times \log k$) in Figure 2. From these experimental results, we conjecture that the diameter of a random k-tree is proportional to $\Theta(\log n / \log k)$.

Scale free property

As shown in Theorem 1(1), the distribution of degrees follows power law on the resultant k-tree in asymptotically. The experimental results imply that convergence to the asymptotic degree distribution occurs rapidly. In Figure 3, we randomly generate a k-tree of $n = 100000$ vertices for $k = 3, 5$ and 10.

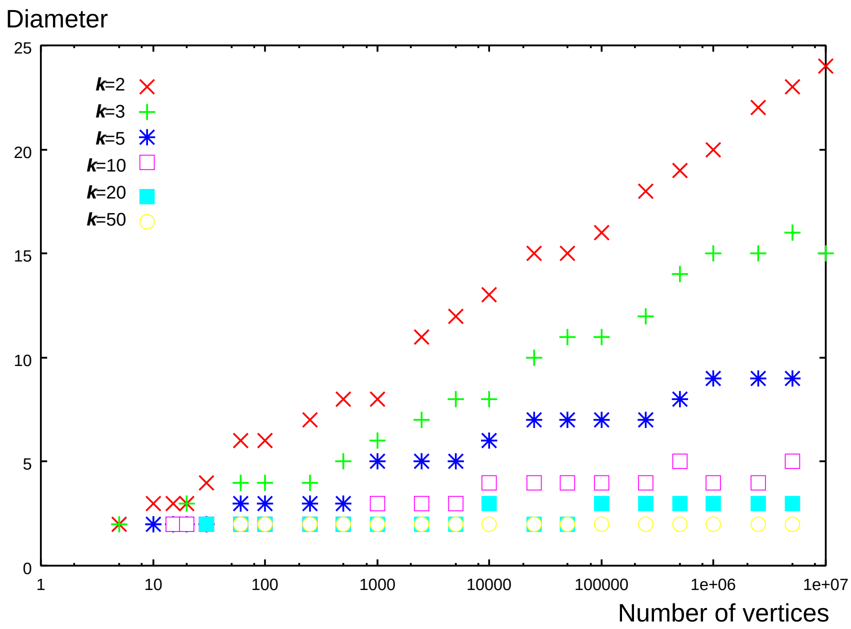

Figure 1: Diameters for scale free k-trees for $k = 2, 3, 5, 10, 20$, and 50.

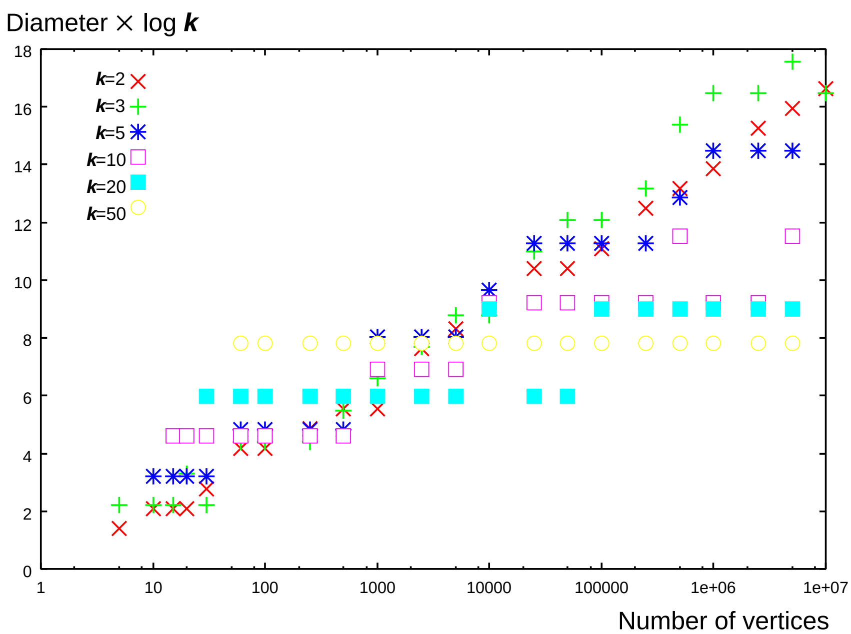

Figure 2: (Diameters$\times \log k$) for scale free k-trees for $k = 2, 3, 5, 10, 20$, and 50.

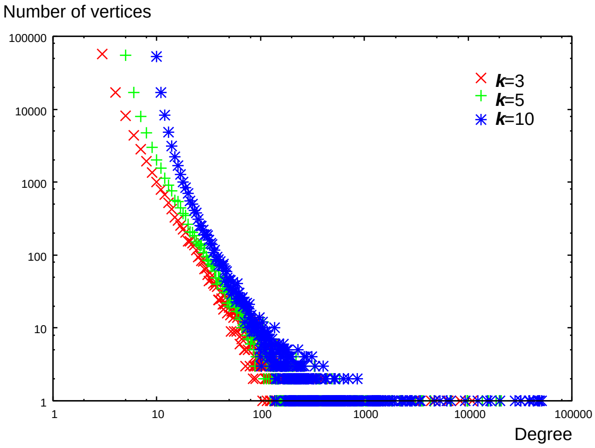

Figure 3: Degree distribution for scale free k-trees for $k = 3, 5$, and 10.

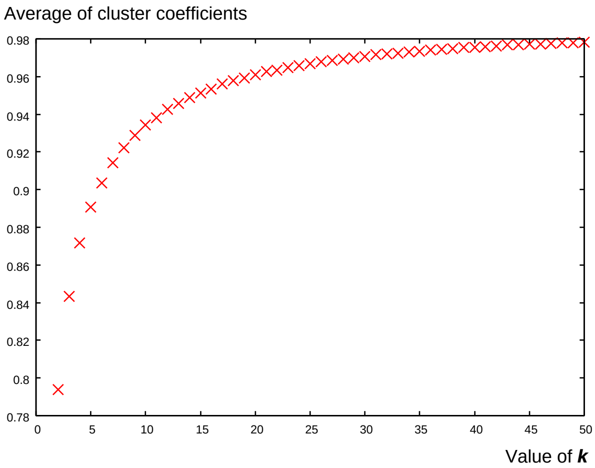

Figure 4: Average cluster coefficients.

Large cluster coefficients

As shown in Theorem 1(2), the limiting expected clustering coefficient $c(k)$ converges to 1 for sufficiently large k. In Figure 4, we generate k-tree of $n = 10000$ vertices and note the convergence to the asymptotic result.

References

[1] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

[2] H. Bodlaender. A Tourist Guide Through Treewidth. *Acta Cybernetica*, 11:1–21, 1993.

[3] H. Bodlaender. A Partial k-Arboretum of Graphs with Bounded Treewidth. *Theoretical Computer Science*, 209:1–45, 1998.

[4] F.R.K. Chung, L. Lu, *Complex Graphs and Networks*, American Mathematical Society, Providence, Rhode Island, (2006).

[5] M. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[6] T. Shigezumi, N. Miyoshi, R. Uehara, and O. Watanabe. Scale Free Interval Graphs. In *International Conference on Algorithmic Aspects in Information and Management (AAIM 2008)*, pages 202–303. Lecture Notes in Computer Science Vol. 5034, Springer-Verlag, 2008.

[7] D. J. Watts and D. H. Strogatz. Collective Dynamics of 'Small-World' Networks. *Nature*, 393:440–442, 1998.

[8] N. Wormald, The differential equation method for random graph processes and greedy algorithms, In: *Lectures on Approximation and Randomized Algorithms*, eds. M. Karoński and H. J. Prömel, PWN, Warsaw, 73–155, (1999).