ISSN 1342-2812

Research Reports on Mathematical and Computing Sciences

Scale Free Interval Graphs

Naoto Miyoshi, Takeya Shigezumi, Ryuhei Uehara and Osamu Watanabe

Mar 2008, C–255

Department of Mathematical and Computing Sciences
Tokyo Institute of Technology

SERIES C: Computer Science

Abstract

Scale free graphs have attracted attention by their non-uniform structure that can be used as a model for various social and physical networks. In this paper, we propose two natural and simple random models for generating scale free interval graphs. These models generate a set of intervals randomly, which defines a random interval graph. The main advantage of those models are its simpleness. The structure/properties of the generated graphs are analyzable by relatively simple probabilistic and/or combinatorial arguments, which is different from the most of the other models for which we need to approximate the processes by certain differential equations. We indeed show that the distribution of degrees follows power law, and it achieves large cluster coefficient.
Keywords: scale free graph, small world network, interval graphs.

1 Introduction

Since early works by Watts & Strogatz [9] and Barabási & Albert [2], small world networks are the focus of recent interest because of their potential as models for the interaction networks of complex systems in real world [1, 8]. There are three major properties that a small world network and/or a scale free network has (see, e.g., [6]): (SF) the node connectivities follow a scale free power law distribution, (CC) two neighbors of a node are also connected by an edge with high probability, and (SW) any two nodes are connected by a short path through a very few nodes called hubs.

Up to now, many models have been proposed and their properties have been investigated. Aside from few deterministic models, most of the randomized models are based on some dynamic *recursive* construction of random graphs. Thus, the analysis of certain properties of the obtained graphs becomes rather complicated, and it is not so easy to see the combinatorial structure of the obtained graphs. Typically, for example, in order to obtain a formula for the distribution of degrees (for showing the property (SF) mentioned above), one has to approximate the process by some differential equations and solve them. Therefore, although many random graph models have been proposed, we think that it is yet important to introduce some random graph model that can be easier to analyze by somewhat standard probabilistic/combinatorial methods. This is important in particular for designing and analyzing algorithms for scale free networks.

In this paper, we propose two simple random models for generating scale free interval graphs. Interval graphs have many applications from scheduling to bioinformatics. A graph $G = (V, E)$ is an interval graph if and only if G has an interval representation $\mathcal{I}$ such that each vertex v corresponds to an interval I_v and two vertices u and v are adjacent in G if and only if corresponding intervals I_u and I_v share a common interval on $\mathcal{I}$. For defining a random interval graph model, we introduce a way to randomly generate an interval representation $\mathcal{I}$; some standard random process is used for choosing intervals' starting points, and a power law distribution is used for determining intervals' lengths. This model has the following intuitive reasoning: Each interval is regarded as a period of existence, i.e., life, of some object (or creature), and relationships are created between these objects who have an overlap of lives. A power law distribution of a lifespan is derived from the simple rule "longer intervals tend to survive yet longer" (since experience is the best teacher).

Technically we consider a random model for generating interval representations. For combinatorial analysis, it is easier to assume that all intervals start at integer points and their lengths are integers. Thus, we adapt *the immigration and death process* for randomly choosing intervals' starting points as integers; this model has been studied well in the queuing theory as the infinite server model. We use a power law distribution on integers for determining lengths of generated intervals.

Although our interval model is defined as a random process, it is also possible to consider random interval distributions in a static way. For example (under the condition that n intervals are generated in a given period) we may assume that the starting points of these intervals are uniformly distributed in the period. Thus, the probabilistic/combinatorial structure of the model gets more clear, and we may be able to use various techniques for analyzing the obtained graphs. In fact, by relatively standard methods, we show that the obtained random interval graphs satisfy two properties of the scale free networks, namely, (SF) and (CC).

2 Preliminaries and Related works

We first introduce the notions for a (undirected) graph $G = (V, E)$ of which each *edge* $e = \{u, v\}$ in $E \subseteq V^2$ has no ordering. We only consider simple graphs without multiedges and self loops. The *neighborhood* of a vertex

v in V is the set $N(v) = \{u \in V \mid \{u, v\} \in E\}$, and the *degree* of v is $|N(v)|$ denoted by $\deg_G(v)$. The subscript G can be omitted if no confusion can arise. We sometimes denote by $v \sim u$ if $u \in N(v)$. For a vertex v in V, the edge $\{v, v\}$ is called *self loop*. An edge $\{u, v\}$ is called *multiedge* if E contains two or more $\{u, v\}$s. A graph G is *simple* if G contains neither self loops nor multiedges. Hereafter, we assume that G is simple unless otherwise stated. For a vertex set $U \subseteq V$, the vertex induced graph $G' = (U, F)$ of $G = (V, E)$, which is denoted by $G[U]$, is defined by $F = \{\{u, v\} \mid u, v \in U \text{ and } \{u, v\} \in E\}$. Given a graph $G = (V, E)$, its *complement* is defined by $\overline{E} = \{\{u, v\} \mid \{u, v\} \notin E\}$, and denoted by $\overline{G} = (V, \overline{E})$. A vertex set I is an *independent set* if $G[I]$ contains no edges, and then the graph $\overline{G}[I]$ is said to be a *clique*. A sequence of distinct vertices $v_1, v_2, \ldots, v_\ell$ is a *path*, denoted by $(v_1, v_2, \ldots, v_\ell)$, if $\{v_j, v_{j+1}\} \in E$ for each $1 \le j < \ell$. The *length* of a path is the number of edges on the path. For two vertices u and v, the *distance* of the vertices, denoted by $d(u, v)$, is the minimum length of the paths joining u and v. We define $d(u, v) = \infty$ if u is not reachable to v. The graph G is *connected* if $d(u, v) < \infty$ for each pair of vertices.

A graph (V, E) with $V = \{v_1, v_2, \ldots, v_n\}$ is an *interval graph* if there is a finite set of intervals $\mathcal{I} = \{I_{v_1}, I_{v_2}, \ldots, I_{v_n}\}$ on the real line such that $\{v_i, v_j\} \in E$ if and only if $I_{v_i} \cap I_{v_j} \neq \emptyset$ for each i and j with $0 < i, j \le n$. We call the set $\mathcal{I}$ of intervals an *interval representation* of the graph. For each interval I, we denote by $R(I)$ and $L(I)$ the right and left endpoints of the interval, respectively (hence we have $L(I) \le R(I)$ and $I = [L(I), R(I)]$). For any interval representation $\mathcal{I}$ and a point p, $\xi[p]$ denotes the set of intervals that contain the point p. We denote by $I_{v_i} \sim I_{v_j}$ if $I_{v_i} \cap I_{v_j} \neq \emptyset$, which means same as $v_i \sim v_j$ for an interval graph, and denote the length of an interval I by $|I|$.

In this paper, we focus on discrete and continuous interval representations. In the discrete interval representation model, each interval I has two integer endpoints $L(I)$ and $R(I)$, and each interval is closed interval with minimum length 0. The discrete model seems the most natural and simple one. However, sometimes, it is (intuitively) better to assume that the minimum length of an interval is 1. In this case, we may use another (but equivalent) interval model that consists of open intervals of length at least one. In the following, we use $[i..j]$ to denote the set of integers $\{i, i+1, \ldots, j\}$. In the continuous model, each interval I has two real endpoints $L(I)$ and $R(I)$, and each interval is closed interval with minimum length $x_{\min}$. We will describe about the length of an interval and the minimum length $x_{\min}$ in the rest of this section.

2.1 Scale free graph

Many social networks can be modeled as a scale free graph such that the degrees of the graph follow a scale free power law distribution [6]. More precisely, given a random distribution on some family of graphs, we consider the following condition for a random graph under this distribution: (SF) the probability that a vertex v has $\deg(v) = k$ is proportional to $k^{-\gamma}$ for some positive constant γ. We call such a random graph (more precisely, a random graph distribution) satisfying this condition *scale free*. Two other properties are required for the notion of *small world*. The first one is about "clustering coefficient", which characterizes the probability that two neighbors of a node are adjacent. The second one is the average (or longest) distance between any pair of vertices in the graph. In this paper we consider the first property and leave the second one for our future topic.

We explain a condition for the small world property on the clustering coefficient. For a vertex $v \in V$, *clustering coefficient of* v, denoted by $\mathsf{CC}(v)$, is defined by:

$$\frac{|\{\{u, w\} \in E \mid u, w \in N(v)\}|}{\binom{\deg(v)}{2}}$$

The *clustering coefficient of* $G = (V, E)$, denoted by $\mathsf{CC}(G)$, is defined by the arithmetical mean of the clustering coefficient of v in V. By definition, we immediately have the following:

$$\mathsf{CC}(G) = \frac{1}{|V|} \sum_{v \in V} \mathsf{CC}(v) = \frac{1}{|V|} \left(\sum_{v \in V} \sum_{u, w \in V \setminus \{v\}, u \neq w} \Pr[u \sim w \mid u \sim v \text{ and } w \sim v] \right).$$

As a desired property of small world graphs, for a given random distribution on some family of graphs, the following condition has been proposed: (CC) for some constant $c > 0$, $\mathsf{CC}(G)$ under the distribution is larger than c.

2.2 Probability Distributions

Our random interval graph model is defined based on a random interval generation model, a way of generating intervals randomly. To determine each interval's starting point, we use some random processes studied in the queuing theory; on the other hand, we use power law distribution for determining the length of each interval. As mentioned in Introduction, we consider two versions. A time-discrete version where both starting points and lifespans are integers, and a more general continuous version. Here for each version, we recall basic distributions and their important properties.

Time-Discrete Version: Bernoulli(p), Poisson(λ) and $\mathcal{P}(\alpha)$

In the time-discrete version, we use the Poisson distribution for specifying the distribution of intervals' starting points. First, we begin by explaining the Poisson distribution that is used to define our interval generating process.

The *Poisson distribution* with parameter λ, Poisson(λ) is given by;

$$\Pr[Y = k] = e^{-\lambda}\frac{\lambda^k}{k!}$$

where Y is the random variable following the Poisson distribution. The *Binomial distribution* with parameter (m, p), written as $B(m, p)$, is considered as the the number of heads in m trials of coin flips with a biased coin. Let us consider a coin which lands on heads with probability p and m trials of coin flips. Then, the probability of the number of heads, say X, equals to k is;

$$\Pr[X = k] = \binom{m}{k} p^k (1-p)^{m-k}.$$

The Poisson distribution can be considered as the limiting case of this probability as the expected number of heads, pm, remains fixed. So, the Poisson distribution is used as an approximation of the Binomial distribution.

We use the Poisson distribution in our interval generating process. We summarize below some important properties of the Poisson distribution.

Let us consider random variables t_i ($i = 1, 2, \ldots, k$) such that t_i follows Poisson(λ) independently. Then, the sum $\sum_{i=1}^{k} t_i$ is also follows the Poisson distribution with parameter $k\lambda$. For any given $t_1, \ldots, t_k$, let $X = \{x_1, x_2, \ldots, x_{\sum_{i=1}^{k} t_i}\}$ be a multiset subset of $[1..k]$ such that $x_j = 1$ for $j = 1, \ldots, t_1$, $x_j = 2$ for $j = t_1 + 1, \ldots, t_1 + t_2$, and so on. Under the condition of $\sum_{i=1}^{k} t_i = n$, we can show that those multiset X of size n occurs uniformly at random. Thus,we have for any multiset subset S of $[1..k]$ of size n and for any n uniformly and independently chosen elements $U_1, U_2, \ldots, U_n$ of $[1..k]$,

$$\Pr[\{x_1, \ldots, x_n\} = S \mid \sum_{i=1}^{k} t_i = n\,] = \Pr[\{U_1, \ldots, U_n\} = S\,]. \tag{1}$$

Second, for specifying the distribution of intervals' lengths, we explain a power law distribution.

We say that a random variable L on non-negative integers follows a *discrete power law distribution* with parameter α (which we denote $\mathcal{P}(\alpha)$) if it satisfies the following.

$$\Pr[L = k] = \frac{1}{\zeta(\alpha)}(k+1)^{-\alpha}, \quad (k \geq 0)$$

where $\zeta(\alpha) = \sum_{i=1}^{\infty} i^{-\alpha}$ is the Riemann's zeta function. Here we note the following property for this random variable L following $\mathcal{P}(\alpha)$.

$$p_k = \Pr[L \geq k+1 \mid L \geq k] = \frac{\zeta(\alpha, k+2)}{\zeta(\alpha, k+1)} \tag{2}$$

where $\zeta(\alpha, n) = \sum_{i=n}^{\infty} i^{-\alpha}$ is the generalized zeta function. This probability, say p_k, increases as k increasing. This gives the simple rule as mentioned in Section 1; "longer intervals tend to survive yet longer".

Continuous Version: $\mathcal{E}(\lambda)$ and $\mathcal{P}(\alpha, x_{\min})$

In the time-continuous version, we use the Poisson process for specifying the distribution of intervals' starting points and use a power law distribution for specifying the length of intervals.

The *exponential distribution* with parameter λ, $\mathcal{E}(\lambda)$, has the following density function $f_{\mathrm{EXP}_\lambda}(x)$.

$$f_{\mathrm{EXP}_\lambda}(x) = \lambda e^{-\lambda x} \quad (x \geq 0).$$

Let X following the exponential distribution. The exponential distribution has the "memoryless property" such that $\Pr[X > s + t \mid X > s] = \Pr[X > t]$.

Let us consider that random variables $t_i (i = 1, 2, \ldots)$ such that t_1 and $t_i - t_{i-1} (2 \le i)$ follows $\mathcal{E}(\lambda)$ independently. Then, $\{t_1, t_2, \ldots, \}$ is called the *Poisson process*. For any given $T_0 \ge 0$ and $T > 0$, let us denote $N(T)$ as the number of t_is between T_0 and $T_0 + T$, i.e., $N(T) = j - i + 1$ such that $t_{i-1} < T_0 \le t_i < \ldots < t_j \le T_0 + T < t_{j+1}$. It is a well known fact that if t_1 and $t_i - t_{i-1} (2 \le i)$ follows $\mathcal{E}(\lambda)$ independently, $N(T)$ follows the Poisson distribution with parameter λT for any T_0. There is also a well known fact (See [7],) that if $\{t_i, \ldots t_{i+n}\}$ satisfies $N(T) = n$ for some T_0, $\{t_i, \ldots, t_{i+n}\}$ can be treated as the uniform distribution on $[T_0, T_0 + T]$. Precisely, for any multiset subset S of $[T_0, T_0 + T]$ of size n and for any uniformly and independently chosen n elements $U_1, \ldots, U_n$ of $[T_0, T_0 + T]$,

$$\Pr[\{t_1, \ldots, t_n\} = S \mid N(T) = n] = \Pr[\{U_1, \ldots, U_n\} = S]. \tag{3}$$

Same as the discrete model, we use a power law distribution to specify the distribution of intervals' length. The continuous version of a power law distribution has two parameters. We say that a random variable L on a real value at least $x_{\min}(> 0)$ follows a *continuous power law distribution* with parameter α and $x_{\min}$ (which we denote $\mathcal{P}(\alpha, x_{\min})$) if its density function $f_{\mathrm{POW}_{\alpha, x_{\min}}}(x)$ satisfies the following.

$$f_{\mathrm{POW}_{\alpha, x_{\min}}}(x) = C_\alpha x^{-\alpha}$$

where $C_\alpha = (\alpha - 1) x_{\min}^{\alpha-1}$ is the normalizing constant. Note that for an random variable X following some continuous distribution, we write $\Pr[X = x]$ to denote the density function $f(x)$ of the continuous distribution.

3 New Model of Scale Free Interval Graphs

To convert an interval representation to an interval graph, there are well known algorithms running in time, e.g. $O(|V| + |E|)$[5], we here present two algorithms which output an interval representaion.

3.1 Discrete interval generation model

We also present a discrete version of the random generation of interval graphs. We use the birth and death process to generate an interval representation. The birth and death process is one of the waiting queue model such as the customers arrive independently to other customers and there exists infinite number of gates for service.

In our model, we set a clock $T = 1, 2, \ldots$ and put intervals on each time using the Poisson distribution. The algorithm for our model, *put-intervals-D* (λ, α, n), is shown in Algorithm 1. In this algorithm, the variable T is the clock for the arriving time, t_T holds the number of intervals begin at time T, and the sub-procedure **RAND_Poisson(λ)** is a random procedure returning an integer according to Poisson(λ). To decide the length of an interval, we use the sub-procedure **RAND_Pow(α)** such that returns an integer according to $\mathcal{P}(\alpha)$.

Actually, this is an approximated approach to the model below. Consider a coin such that lands on heads with probability p. Flip the coin m times at each time step T and if the coin lands on head, we put an interval starting at time T. The number of heads on time T follows $B(m, p)$, and we can approximate it with the Poisson distribution, Poisson(λ), if m tends to infinity as the expected number of heads, $pm = \lambda$, remains fixed.

The complexity of this algorithm depends on the parameter λ. Let T_{end} be the final value of T. Since the expected number of intervals born in time T is λ, the expected value of T_{end} is $\frac{n}{\lambda}$. The sub-procedure **RAND_Poisson(λ)** is called T_{end} times and the sub-procedure **RAND_Pow(α)** is called n times. The total expected time complexity is

$$O\left(\frac{1}{\lambda} n \mathbf{Time(Poisson(}\lambda\mathbf{))Time(Power(}\alpha\mathbf{))}\right)$$

where **Time(Poisson(λ))** and **Time(Power(α))** are time complexities of sub-prosedures.

We here consider values of $\zeta(\alpha)$ and how it is related intervals' lifespan for a typical value of the parameter α.

Example 1 *It has been usually claimed that typical scale-free networks satisfy (SF) with* $\alpha = 2.1 \simeq 2.8$. *Since our later analysis shows that the smaller* α *gives the smaller clustering coefficient in our model, we consider* $\alpha = 2.1$ *for our example. Then, since* $\zeta(2.1) \simeq 1.560$, *we have* on average $n_0 \simeq 0.641n$, $n_1 \simeq 0.150n$, *and* $n_2 \simeq 0.064n$, *where* n_i *denotes the number of vertices such that corresponding interval has length i, and n denotes the number of vertices.*

Algorithm 1: *put-intervals-D* (λ, α, n)

input : Parameters λ, α, and n.
output: A set of intervals $\mathcal{I}$.
begin
  $T = 1$, $i = 1$, $\mathcal{I} = \phi$;
  while $i \leq n$ **do**
    t_T = **RAND_Poisson(**λ**)**;
    put t_T intervals on T, i.e.,
    for $j = i$ *to* $i + t_T - 1$ **do**
      set $L(I_j) = T$;
      set the length of I_j, l_j = **RAND_Pow(**α**)**;
      set $R(I_j) = T + l + j$;
    end
    $i = i + t_T$;
    proceed the clock T to $T + 1$;
  end
  output $\mathcal{I}$.
end

3.2 Continuous interval generation model

In the continuous model, The birth and death process are represented as follows. Let $\{L(I_i) \mid i = 1, \ldots, n\}$ be the set of the birth time of intervals. Let $t_1 = L(I_1)$ and $t_i = L(I_i) - L(I_{i-1})$ $(2 \leq i \leq n)$. t_i follows $\mathcal{E}(\lambda)$ and the length of the interval I_i, say l_i, follows $\mathcal{P}(\alpha, x_{\min})$. Let sub-procedures **Rand_Pow(**$\alpha, x_{\mathbf{min}}$**)** and **Rand_Expo(**λ**)** be returning a random real number according to $\mathcal{P}(\alpha, x_{\min})$ and $\mathcal{E}(\lambda)$, respectively. Note that **Rand_Pow(**$\alpha, x_{\mathbf{min}}$**)** and **Rand_Pow(**α**)** are different distributions. The random generation of an interval graph is done by this procedure, *put-intervals-C*$(\lambda, \alpha, x_{min}, n)$, shown in Algorithm 2.

Algorithm 2: *put-intervals-C* $(\lambda, \alpha, x_{min}, n)$

input : Parameters λ, α, $x_{\min}$ and n.
output: A set of intervals $\mathcal{I}$.
begin
  $T = 0$, $i = 1$, $\mathcal{I} = \phi$;
  while $i \leq n$ **do**
    /* decide the left endpoint of the interval */
    $L(I_i) = T$;
    /* decide the lifetime(service time) of this interval */
    l_i = **Rand_Pow(**$\alpha, x_{\mathbf{min}}$**)**;
    /* decide the right endpoint of this interval */
    $R(I_i) = T + l_i$;
    add I_i to $\mathcal{I}$;
    t_i = **Rand_Expo(**λ**)**;
    $T = T + t_i$;
    $i = i + 1$;
  end
  output $\mathcal{I}$.
end

For a contrast to the discrete model, we note the following fact. For a fixed time period, say $[T_0, T_0 + T]$, the number of birth in this period, $N(T)$ follows the Poisson(λT) independent to T_0.

The time complexity is:

$$O\left(n\mathbf{Time(Expo(}\lambda\mathbf{))Time(Power(}\alpha, x_{\mathbf{min}}\mathbf{)))}\right)$$

where **Time(Expo(**λ**))** and **Time(Power(**$\alpha, x_{\mathbf{min}}$**))** are time complexities of sub-prosedures.

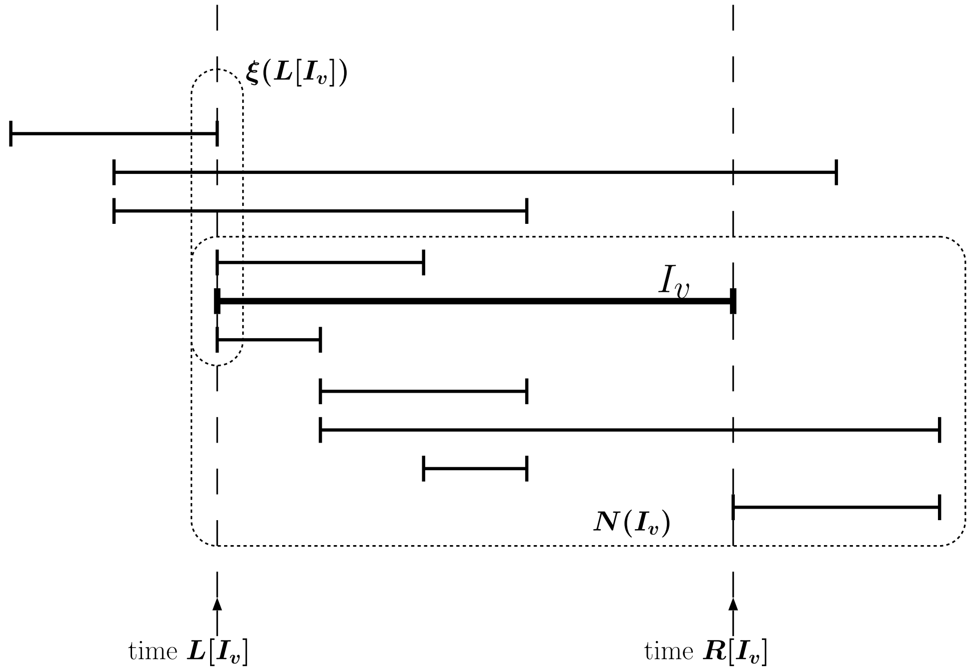

Figure 1: An example of the degree of v: $\deg(v) = 9$. There are 6 intervals on time $L[I_v]$, 6 intervals are put on time $[L[I_v]..R[I_v]]$ and 3 intervals starts at time $L[I_v]$.

4 Scale Free Property

In this section, we will show that both versions of a scale free interval graph has the degree sequence following a power law distribution.

4.1 Scale free property for the discrete model

To consider the degree of a vertex, let us define $\xi(T)$ for time T and $A(I)$ for an interval I. $\xi(T)$ is the number of intervals which exist on time T in our algorithm 1. $A(I)$ represents the number of intervals whose left endpoints are put on $[L[I]..R[I]]$. It is easy to see that the degree of a vertex v is the sum of $\xi(L[I_v])$ and $A(I_v)$ minus $t_{L[I_v]}$. (See figure 1). $\xi(L[I_v]) - t_{L[I_v]}$ means the number of intervals which exist on time $L[I_v]$ and started before $L[I_v]$. We will analyze the stationary distribution $\pi(k)$ of $\xi(T)$ (i.e., $\pi(k) = \lim_{T\to\infty} \Pr[\xi(T) = k]$) and $A(I)$. First, we will show that $\xi(T) - t_T$ follows Poisson($\lambda \frac{\zeta(\alpha-1)-1}{\zeta(\alpha)}$) in the steady state. Second, we will show that $\Pr[A(I) = k]$ follows a power law distribution for large k. Third, we conclude with the fact that a power law distribution dominates the Poisson distribution for large degrees. In the rest of this section, we use the $f(x) \sim g(x)$ notation to approximate $f(x)$ by $g(x)$. Precisely, "$f(x) \sim g(x)$ as $x \to \infty$" stands for "$\lim_{x\to\infty} f(x)/g(x) = 1$".

Consider the time T of the procedure *put-intervals*. Some intervals exist and each of them has their current length ≥ 0. Since the length of an interval follows $\mathcal{P}(\alpha)$, the probability of survive depends on the current length of the interval. Let the p_i (for $i \geq 0$) be the probability such that an interval whose current length is i at time T will survive at time $T + 1$. Since we consider that the length of an interval follows $\mathcal{P}(\alpha)$, p_i is derived from equation (2).

Let ρ_i^T be the number of intervals which are alive and have current length i at time T. As the time T will proceed, ρ_{i+1}^{T+1} is depends only on ρ_i^T because some of intervals of ρ_i^T will be alive at time $T + 1$ with probability p_i and others die at time T. From this observation, we obtain this formula:

$$\Pr[\rho_{i+1}^{T+1} = k] = \sum_{m=k}^{\infty} \binom{m}{k} p_i^k (1 - p_i)^{m-k} \Pr[\rho_i^T = m] \tag{4}$$

for $i \geq 0$. Since ρ_0^T is the number of intervals born at time T, $\Pr[\rho_0^T = k] = e^{-\lambda} \frac{\lambda^k}{k!}$. Let us consider the stationary distribution π_i such that $\pi_i(k) = \lim_{T\to\infty} \Pr[\rho_i^T = k]$. For the stationary distribution π_i, the equation (4) becomes

$$\pi_{i+1}(k) = \sum_{m=k}^{\infty} \binom{m}{k} p_i^k (1-p_i)^{m-k} \pi_i(m) \tag{5}$$

and $\pi_0(k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

We will show the following lemma as the solution of the equation (5).

Lemma 2 *Let us denote $P_i = \prod_{j=0}^{i-1} p_j$ for $i \geq 1$ and $P_0 = 1$. The stationary distribution π_i follows Poisson(λP_i);*

$$\pi_i(k) = e^{-\lambda P_i} \frac{(\lambda P_i)^k}{k!}.$$

Proof. The proof is done by induction. For $i = 0$, $\pi_0(k) = e^{-\lambda P_0} \frac{(\lambda P_0)^k}{k!}$. Assume it holds for $i \leq k$, i.e., $\pi_i(k) = e^{-\lambda P_i} \frac{(\lambda P_i)^k}{k!}$. The stationary distribution is:

$$\begin{aligned}
\pi_{i+1}(k) &= \sum_{m=k}^{\infty} \frac{m!}{(m-k)!k!} p_i^k (1-p_i)^{m-k} \pi_i(m) = \sum_{m=k}^{\infty} \frac{m!}{(m-k)!k!} p_i^k (1-p_i)^{m-k} e^{-\lambda P_i} \frac{(\lambda P_i)^m}{m!} \\
&= e^{-\lambda P_i} \frac{p_i^k}{k!} (\lambda P_i)^k \sum_{m=k}^{\infty} \frac{(1-p_i)^{m-k}}{(m-k)!} (\lambda P_i)^{m-k} = e^{-\lambda P_i} \frac{(\lambda p_i P_i)^k}{k!} \sum_{m'=0}^{\infty} \frac{\{\lambda P_i (1-p_i)\}^{m'}}{m'!} \\
&= e^{-\lambda P_i} \frac{(\lambda P_{i+1})^k}{k!} e^{\lambda P_i (1-p_i)} = e^{-\lambda P_i + \lambda P_i - \lambda p_i P_i} \frac{(\lambda P_{i+1})^k}{k!} = e^{-\lambda P_{i+1}} \frac{(\lambda P_{i+1})^k}{k!}.
\end{aligned}$$

Note that we used $p_i P_i = P_{i+1}$ in the above. ∎

We now can obtain the stationary distribution of $\xi(T)$, however, what we want to know is the distribution of $\xi(T) - t_T$ since the degree of a vertex is $(\xi(L[I]) - t_{L[I]}) + A(I)$. we conclude the first part of the degree analysis with the following lemma.

Lemma 3 *Let $\xi(T)$ be the number of intervals which are alive at time T and t_T be the number of intervals starting at time T. Then $\xi(T) - t_T$ follows Poisson($\lambda \frac{\zeta(\alpha-1)-1}{\zeta(\alpha)}$) in the steady state.*

Proof. Since π_is are independent and follow Poisson(λP_i), the sum $\xi(T) - \pi_0 = \sum_{i=1}^{\infty} \pi_i$ also follows the Poisson distribution with parameter $\sum_{i=1}^{\infty} \lambda P_i$. Recall that $P_i = \prod_{j=0}^{i-1} p_j = \frac{\zeta(\alpha, i+1)}{\zeta(\alpha)}$,

$$\begin{aligned}
\sum_{i=1}^{\infty} \lambda P_i &= \lambda \sum_{i=1}^{\infty} \frac{\zeta(\alpha, i+1)}{\zeta(\alpha)} = \frac{\lambda}{\zeta(\alpha)} \sum_{i=2}^{\infty} \zeta(\alpha, i) = \frac{\lambda}{\zeta(\alpha)} \sum_{i=2}^{\infty} \sum_{j=i}^{\infty} \frac{1}{j^\alpha} \\
&= \frac{\lambda}{\zeta(\alpha)} \left[\left\{ \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \frac{1}{4^\alpha} + \frac{1}{5^\alpha} + \dots \right\} + \left\{ \frac{1}{3^\alpha} + \frac{1}{4^\alpha} + \frac{1}{5^\alpha} + \dots \right\} + \left\{ \frac{1}{4^\alpha} + \frac{1}{5^\alpha} + \dots \right\} + \dots \right] \\
&= \frac{\lambda}{\zeta(\alpha)} \sum_{i=2}^{\infty} \frac{i}{i^\alpha} = \frac{\lambda}{\zeta(\alpha)} \sum_{i=2}^{\infty} \frac{1}{i^{\alpha-1}} = \lambda \frac{\zeta(\alpha-1)-1}{\zeta(\alpha)}.
\end{aligned}$$

∎

Second, we will show that $A(I)$ follows a power law distribution. Recall that for any T_0 and $T \geq 0$, the number of intervals which start on time $[T_0..T_0+T]$ follows the Poisson distribution with parameter $\lambda(T+1)$. Let us suppose that the interval I has length l. The number of intervals starting on $[L[I]..R[I]](= [L[I]..L[I]+l])$ is;

$$\Pr[A(I) = k \mid |I| = l] = e^{-\lambda(l+1)} \frac{\{\lambda(l+1)\}^k}{k!} \tag{6}$$

Since the length of an interval follows $\mathcal{P}(\alpha)$,

$$\begin{aligned}
\Pr[A(I) = k] &= \sum_{l=0}^{\infty} e^{-\lambda(l+1)} \frac{\{\lambda(l+1)\}^k}{k!} \Pr[|I| = l] = \sum_{l=0}^{\infty} e^{-\lambda(l+1)} \frac{\{\lambda(l+1)\}^k}{k!} \frac{1}{\zeta(\alpha)} (l+1)^{-\alpha} \\
&= \sum_{l=1}^{\infty} e^{-\lambda l} \frac{(\lambda l)^k}{k!} \frac{1}{\zeta(\alpha)} l^{-\alpha} = \frac{\lambda^k}{\zeta(\alpha) k!} \sum_{l=1}^{\infty} e^{-\lambda l} l^{k-\alpha} = \frac{\lambda^k}{\zeta(\alpha) k!} \sum_{l=0}^{\infty} e^{-\lambda l} l^{k-\alpha}
\end{aligned}$$

For the last formula , we show the following lemma.

Lemma 4

$$\frac{\lambda^k}{k!}\sum_{l=0}^{\infty} l^{k-\alpha}\, e^{-\lambda l} \sim \lambda^{\alpha-1}\, k^{-\alpha} \quad \textit{as } k \to \infty.$$

Proof. Let $f(k) = \frac{\lambda^k}{k!}\sum_{l=0}^{\infty} l^{k-\alpha}\, e^{-\lambda l}$ and suppose $k > \alpha$. We then have

$$f(k) \geq \frac{\lambda^k}{k!}\int_0^\infty x^{k-\alpha}\, e^{-\lambda x}\, \mathrm{d}x - \frac{\lambda^\alpha}{k!}\,(k-\alpha)^{k-\alpha}\, e^{-(k-\alpha)},$$
$$f(k) \leq \frac{\lambda^k}{k!}\int_0^\infty x^{k-\alpha}\, e^{-\lambda x}\, \mathrm{d}x + \frac{\lambda^\alpha}{k!}\,(k-\alpha)^{k-\alpha}\, e^{-(k-\alpha)}.$$

Let the integral part $g(k) = \frac{\lambda^k}{k!}\int_0^\infty x^{k-\alpha}\, e^{-\lambda x}\, \mathrm{d}x$ and the rest $h(k) = \frac{\lambda^\alpha}{k!}\,(k-\alpha)^{k-\alpha}\, e^{-(k-\alpha)}$. We show that $g(k) \sim \lambda^{\alpha-1}k^{-\alpha}$ as $k \to \infty$ and $h(k) = o(k^{-\alpha})$ as $k \to \infty$. Changing the variable to $u = \lambda\, x$, we have

$$g(k) = \frac{\lambda^{\alpha-1}}{k!}\int_0^\infty u^{k-\alpha}\, e^{-u}\, \mathrm{d}u = \frac{\lambda^{\alpha-1}}{k!}\,\Gamma(k-\alpha+1),$$

where Γ denotes the Gamma function $\Gamma(s) = \int_0^\infty u^{s-1}\, e^{-u}\, \mathrm{d}u$. We use the following properties of Gamma function;

(i) $\Gamma(s+1) = s\,\Gamma(s)$,

(ii) $\Gamma(s) = \lim_{k\to\infty} \dfrac{k!\, k^s}{s\,(s+1)\cdots(s+k)}$ $(s \neq 0, -1, -2, \cdots)$.

Then, we have for $\alpha \neq 0, 1, 2, \ldots$,

$$g(k) \times k^\alpha = \lambda^{\alpha-1}\,\frac{(k-\alpha)\cdots(1-\alpha)(-\alpha)}{k!\, k^{-\alpha}}\,\Gamma(-\alpha) \to \lambda^{\alpha-1} \quad \text{as } k \to \infty.$$

When α is a nonnegative integer, we have

$$g(k) = \lambda^{\alpha-1}\,\frac{(k-\alpha)!}{k!} \sim \lambda^{\alpha-1}\, k^{-\alpha} \quad \text{as } k \to \infty.$$

Finally, for the term $h(k)$, applying Stirling's formula $k! \sim \sqrt{2\pi}\, k^{k+1/2}\, e^{-k}$ as $k \to \infty$, we have

$$h(k) \times k^\alpha = \lambda^\alpha\,\frac{k^k\, e^{-(k-\alpha)}}{k!}\left(1 - \frac{\alpha}{k}\right)^{k-\alpha} \sim \frac{\lambda^\alpha}{\sqrt{2\pi}}\, k^{-1/2} \to 0 \quad \text{as } k \to \infty.$$

∎

Applying Lemma 4 , we obtain

$$\Pr[A(I) = k] \sim \frac{1}{\zeta(\alpha)}\lambda^{\alpha-1}k^{-\alpha}.$$

Third, we will show that we can neglect the effect of $\xi(L[I]) - t_{L[I]}$ if the degree of v is large enough. We present the following lemma.

Lemma 5 *Let $\bar{F}$ and $\bar{G}$ be the tail probability of a power law distribution with parameter α and the Poisson distribution with parameter $\lambda\mu$, respectively. In precise, using the constant c, $\bar{F}(k) = c\sum_{i=k}^{\infty} i^{-\alpha}$ and $\bar{G}(k) = \sum_{i=k}^{\infty} e^{-\lambda\mu}\frac{(\lambda\mu)^k}{k!}$. Then we have;*

$$\frac{\bar{G}(k)}{\bar{F}(k)} \to 0 \quad \textit{as } k \to \infty.$$

Proof. Let us recall that $k! \geq \left(\frac{k}{3}\right)^k$ and if $\lim_{x\to\infty} f(x) = 0$ and $\lim_{x\to\infty} g(x) = 0$, then $\lim_{x\to\infty}\frac{g(x)}{f(x)} = \lim_{x\to\infty}\frac{\frac{d}{dx}g(x)}{\frac{d}{dx}f(x)}$.

$$\lim_{k\to\infty}\frac{\bar{G}(k)}{\bar{F}(k)} < \lim_{k\to\infty}\frac{e^{-\lambda\mu}\sum_{i=k}^{\infty}\left(\frac{3\lambda\mu}{i}\right)^i}{c\sum_{i=k}^{\infty} i^{-\alpha}} < \lim_{k\to\infty}\frac{e^{-\lambda\mu}\int_k^\infty\left(\frac{3\lambda\mu}{x}\right)^x \mathrm{d}x}{c\int_k^\infty x^{-\alpha}\mathrm{d}x} = \frac{e^{-\lambda\mu}}{c}\lim_{x\to\infty}\frac{\left(\frac{3\lambda\mu}{x}\right)^x}{x^{-\alpha}} = 0$$

∎

By Lemma 5 with $\mu = \frac{\zeta(\alpha-1)-1}{\zeta(\alpha)}$, for sufficiently large degree vertices, $A(I_v)$ dominates $\xi[L[I_v]] - t_{L[I_v]}$ on the degree distribution. We now conclude with the following result.

Theorem 6 *A scale free interval graph generated according to our discrete model has the degree sequence following $\mathcal{P}(\alpha)$ for large degrees.*

4.2 Scale-free property for the continuous model

Same as the discrete model, we will analyze the distribution of the stationary distribution $\pi(k)$ and the number of intervals which are born on an interval I, $A(I)$.

For a continuous model, there is a well known fact on the analysis of the infinite server queuing model with Poisson process input [7]. The stationary distribution of the number of customers in service follows the Poisson distribution.

Theorem 7 **[7]** *Let us suppose that the number of servers are infinite, and let us denote the number of customers being served at time T, $\xi(T)$. Let the inter-arrival times $t_i = L(I_{i+1}) - L(I_i)$ ($i = 1, 2, \ldots; L(I_1) = 0$) are identically distributed and independent random variables with the exponential distribution. If the average of the service time μ is finite, then, the stationary distribution $\pi(k) = \lim_{T\to\infty} \Pr[\xi(T) = k]$ follows the Poisson distribution with parameter $\lambda\mu$*

We can regard the left endpoint of an interval as the arrival of a customer and regard the length of an interval as the service time of a customer.

Same as the discrete model, let us denote $A(I_i)$ the number of birth on the interval I_i, in other words, the number of arrivals of customers on $[L(I_i), R(I_i)]$. $A(I_i)$ depends on the length of the interval such that follows $\mathcal{P}(\alpha)$. We will show that $\Pr[A(I_i) = k]$ follows a power law distribution for large enough k, and $A(I_i)$ is the main term of the degree of an interval using Lemma 5.

For $A(I_i)$, we will show the following lemma.

Lemma 8 $\Pr[A(I_i) = k] \to ck^{-\alpha}$ *as $k \to \infty$ for some constant c.*

Proof.

$$\begin{aligned}\Pr[A(I_i) = k] &= \int_0^\infty \Pr[A(I_i) = k \mid l_i = t] \Pr[l_i = t] \mathrm{d}t = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} \frac{\alpha - 1}{x_{\min}^{1-\alpha}} t^{-\alpha} \mathrm{d}t \\ &= \frac{(\alpha - 1)\lambda^\alpha}{x_{\min}^{1-\alpha} k!} \int_0^\infty (\lambda t)^{k-\alpha} e^{-\lambda t} \mathrm{d}t = \frac{(\alpha - 1)\lambda^{\alpha-1}}{x_{\min}^{1-\alpha} k!} \int_0^\infty u^{k-\alpha} e^{-u} \mathrm{d}u = \frac{(\alpha - 1)\lambda^{\alpha-1}}{x_{\min}^{1-\alpha}} \frac{1}{k!} \Gamma(k - \alpha + 1).\end{aligned}$$

We used that changing the variable $u = \lambda t$. We will show that

$$\frac{1}{k!}\Gamma(k - \alpha + 1) \to k^{-\alpha} \quad \text{as } k \to \infty.$$

Using the property of the Gamma function shown in the analysis of the discrete model, for $\alpha \neq 0, 1, 2, \ldots$,

$$\frac{1}{k!}\Gamma(k - \alpha + 1) = k^{-\alpha} \frac{(k - \alpha) \cdots (1 - \alpha)(-\alpha)}{k! k^{-\alpha}} \Gamma(-\alpha) \to k^{-\alpha} \quad \text{as } k \to \infty.$$

When α is a nonnegative integer, we have

$$\frac{1}{k!}\Gamma(k - \alpha + 1) = \frac{(k - \alpha)!}{k!} \to k^{-\alpha} \quad \text{as } k \to \infty.$$

∎

By Lemma 5, we conclude with the following result.

Theorem 9 *A scale free interval graph generated according to our continuous model has the degree sequence following $\mathcal{P}(\alpha)$ for large degrees.*

5 Clustering coefficient

In this section, we will show the constant lower bound of the expected value of the clustering coefficient.

5.1 The clustering coefficient analysis for the continuous model

We analyze the expected value of the clustering coefficient of a scale free interval graph. First, we show that there are many *short* intervals and, second, the expected value of the clustering coefficient of those short intervals are large. We call that a vertex is *r-short* if the length of the interval corresponding to the vertex is less than $rx_{\min}$. Since the length of an interval follows a continuous power law distribution, a vertex is r-short with probability;

$$\Pr[L < rx_{\min}] = 1 - \int_{rx_{\min}}^{\infty} (\alpha-1)x_{\min}^{\alpha-1}t^{-\alpha}\mathrm{d}t = 1 - x_{\min}^{\alpha-1}(rx_{\min})^{1-\alpha} = 1 - r^{1-\alpha}.$$

The expected number of short intervals is $\left(1 - r^{1-\alpha}\right)n$. For a fixed r, let us consider about the clustering coefficient CC(v) for an r-short vertex v. Hereafter, we refer to the length of the interval I_v corresponding to a vertex v as l_v and let us denote the set Pair($N(v)$) be all of pairs chosen from $N(v)$. Let us use the [ ] notation for an event A which means that $[A] = 1$ if A occurs and $[A] = 0$ otherwise. The clustering coefficient of the vertex v is:

$$\begin{aligned}\mathsf{CC}(v) = \frac{|\mathrm{Pair}(N(v)) \cap E|}{|\mathrm{Pair}(N(v))|} &= \frac{1}{|\mathrm{Pair}(N(v))|} \sum_{\{u,w\}\in\mathrm{Pair}(N(v))} [\{u,w\} \in E] \\ &= \frac{1}{|\mathrm{Pair}(N(v))|} \Pr_{\{u,w\}\in_U \mathrm{Pair}(N(v))}[\{u,w\} \in E] = \frac{1}{|\mathrm{Pair}(N(v))|} \Pr_{\{u,w\}\in_U V}[\{u,w\} \in E \mid \{u,w\} \in \mathrm{Pair}(N(v))].\end{aligned}$$

The last probability is,

$$\begin{aligned}&\Pr[\{u,w\} \in E \mid \{u,w\} \in \mathrm{Pair}(N(v))] \\ =\ &\int_{x_{\min}}^{\infty}\int_{x_{\min}}^{\infty} \Pr[\{u,w\} \in E \mid \{u,v\} \in E, \{w,v\} \in E, l_u = s, l_w = t]\, f_{\mathrm{POW}_{\alpha,x_{\min}}}(s) f_{\mathrm{POW}_{\alpha,x_{\min}}}(t)\mathrm{d}s\mathrm{d}t. \qquad (7)\end{aligned}$$

The key lemma for three intervals is the following, which is independent of the degree distribution:

Lemma 10 *Let three intervals I_v, I_u, I_w have fixed length l_v, s, t, respectively. We have;*

$$\Pr[I_u \sim I_w \mid I_u \sim I_v, I_w \sim I_v] = \frac{(s+t)l_v + st}{(l_v+s)(l_v+t)}.$$

Proof. Let us consider an axis such that the left endpoint of I_v, $L(I_v)$ is set to 0.

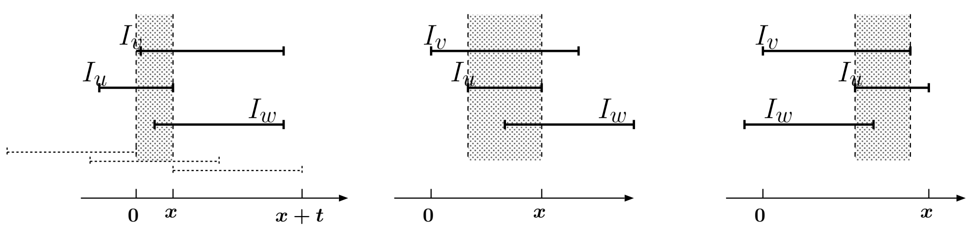

Figure 2: Three cases for the condition $s < l_v$.

Under the condition $I_u \sim I_v$ and $I_w \sim I_v$, $R(I_u)$ and $R(I_w)$ should be in $[0, l_v + s]$ and $[0, l_v + t]$ respectively. Under this condition, we can regard the distribution of $R(I_u)$ and $R(I_w)$ as the uniform distribution by equation (3).

First, for the case $s < l_v$, as shown in Figure 2, if $R(I_u)$ is in $[0, l_v]$, $R(I_w)$ should be in $[0, R(I_u) + t]$. Summing up the three cases in Figure 2, the above probability is;

$$\begin{aligned}&\Pr[I_u \sim I_w \mid I_u \sim I_v, I_w \sim I_v, l_u = s, l_w = t] \\ =\ &\int_0^s \Pr[R(I_w) \in [0, t+x] | R(I_w) \in [0, l_v + t]]\Pr[R(I_u) = x]\mathrm{d}x \\ &+ \int_s^{l_v} \Pr[R(I_w) \in [x-s, t+x] | R(I_w) \in [0, l_v + t]]\Pr[R(I_u) = x]\mathrm{d}x \\ &+ \int_{l_v}^{l_v+s} \Pr[R(I_w) \in [x-s, t+l_v] | R(I_w) \in [0, l_v + t]]\Pr[R(I_u) = x]\mathrm{d}x \\ =\ &\int_0^s \frac{x+t}{l_v+t}\frac{1}{l_v+s}\mathrm{d}x + \int_s^{l_v} \frac{s+t}{l_v+t}\frac{1}{l_v+s}\mathrm{d}x + \int_{l_v}^{l_v+s} \frac{l_v+t-x+s}{l_v+t}\frac{1}{l_v+s}\mathrm{d}x = \frac{st+(s+t)l_v}{(l_v+s)(l_v+t)}.\end{aligned}$$

For the case $s > l_v$, we can show the same equation in a similar way. ∎

We here obtained that the equation (7) is;

$$\int_{x_{\min}}^{\infty}\int_{x_{\min}}^{\infty}\Pr[\{u,w\}\in E \mid \{u,v\}\in E, \{w,v\}\in E, l_u = s, l_w = t]f(s)f(t)\mathrm{d}s\mathrm{d}t$$

$$= \int_{x_{\min}}^{\infty}\int_{x_{\min}}^{\infty}\frac{st+(s+t)l_v}{(l_v+s)(l_v+t)}\{C_\alpha s^{-\alpha}C_\alpha t^{-\alpha}\}\mathrm{d}s\mathrm{d}t = C_\alpha^2\int_{x_{\min}}^{\infty}\int_{x_{\min}}^{\infty}\left\{1-\frac{l_v^2}{(l_v+s)(l_v+t)}\right\}s^{-\alpha}t^{-\alpha}\mathrm{d}s\mathrm{d}t$$

$$= 1 - l_2^v C_\alpha^2\int_{x_{\min}}^{\infty}\frac{1}{(l_v+s)s^\alpha}\left[\int_{x_{\min}}^{\infty}\frac{1}{(l_v+t)t^\alpha}\mathrm{d}t\right]\mathrm{d}s.$$

For the term of integration on t, we use $x_{\min} \le l_v$,

$$\int_{x_{\min}}^{\infty}\frac{1}{(l_v+t)t^\alpha}\mathrm{d}t \le \int_{x_{\min}}^{l_v}\frac{1}{(2t)t^\alpha}\mathrm{d}t + \int_{l_v}^{\infty}\frac{1}{(2l_v)t^\alpha}\mathrm{d}t = \frac{x_{\min}^{-\alpha}}{2\alpha} + \frac{l_v^{-\alpha}}{2\alpha(\alpha-1)} \le \frac{x_{\min}^{-\alpha}}{2\alpha}\left(1+\frac{1}{\alpha-1}\right) = \frac{x_{\min}^{-\alpha}}{2(\alpha-1)}.$$

Using $l_v < rx_{\min}$, the lower bound of the equation (7) is;

$$1 - l_2^v C_\alpha^2\int_{x_{\min}}^{\infty}\frac{1}{(l_v+s)s^\alpha}\left[\int_{x_{\min}}^{\infty}\frac{1}{(l_v+t)t^\alpha}\mathrm{d}t\right]\mathrm{d}s \ge 1 - l_2^v C_\alpha^2\left\{\frac{x_{\min}^{-\alpha}}{2(\alpha-1)}\right\}^2 > 1 - \frac{r^2}{4}.$$

We now obtained that for r-short vertex v,

$$\Pr[\{u,w\}\in E \mid \{u,w\}\in \mathrm{Pair}(N(v))] > 1 - \frac{r^2}{4}.$$

Let V_r be the set of r-short vertices. The lower bound of the clustering coefficient of the graph G is;

$$\mathsf{CC}(G) = \frac{1}{|V|}\sum_{v\in V}\mathsf{CC}(v) = \frac{1}{|V|}\left[\sum_{v\in V_r}\mathsf{CC}(v) + \sum_{v\in V\setminus V_r}\mathsf{CC}(v)\right] \ge \frac{1}{|V|}\sum_{v\in V_r}\mathsf{CC}(v) > \frac{1}{|V|}|V_r|\left(1-\frac{r^2}{4}\right).$$

Since the expected size of V_r is $\left(1-r^{1-\alpha}\right)|V|$, the lower bound of the expected value of CC(G) is;

$$E[\mathsf{CC}(G)] > E\left[\frac{1}{|V|}|V_r|\left(1-\frac{r^2}{4}\right)\right] = \left(1-r^{1-\alpha}\right)\left(1-\frac{r^2}{4}\right).$$

For example, $\alpha = 2.1$ (same as Example 1), let $r = 1.447$, $E[\mathsf{CC}(G)] > 0.159$. Note that it is independent to the size of the graph.

5.2 The clustering coefficient analysis for the discrete model

Same as the continuous model, we will show that there are many short vertices and they have a large clustering coefficient. For given $G = (V, E)$, we partition V into $V_0, V_1, \ldots$ such that V_i contains vertices corresponding to intervals of length i. Let n_i be the number of vertices in V_i, and n the number of vertices in V. Our goal is to show how to compute a lower bound of the clustering coefficient of a scale free interval graph since it depends on the distribution of n_i. Typically, we have the following lower bound.

Example 11 *By Example 1, we assume that the expected values of* n_0, n_1, *and* n_2 *are* $0.641n$, $0.150n$, *and* $0.064n$, *respectively. Then the expected value of the clustering coefficient of* G, $\mathsf{CC}(G)$, *is at least* 0.7713.

We have the key lemma same as Lemma 10. It is independent of the degree distribution:

Lemma 12 *Let* I_u, I_v, I_w *be any three intervals placed randomly. We assume that the positions of the intervals are independent, and the universal interval is long enough. Then,* $\Pr[I_u \sim I_w \mid I_u \sim I_v \text{ and } I_w \sim I_v] = \dfrac{l_u l_v + l_v l_w + l_w l_u + l_u + l_v + l_w + 1}{(l_u + l_v + 1)(l_w + l_v + 1)}$.

Proof. To simplify, we shift the whole intervals and fix $L(I_v) = 0$ and $R(I_v) = l_v$. Then $R(I_u)$ takes i in $[0..l_v + l_u]$ with conditional probability $\frac{1}{l_u+l_v+1}$ given $I_u \sim I_v$. Similarly, $R(I_w)$ takes j in $[0..l_v+l_w]$ with conditional probability $\frac{1}{l_w+l_v+1}$ given $I_w \sim I_v$.

We first assume that $l_w \le l_v \le l_u$. Then, for each i in $[0..l_v + l_u]$, we have the following cases; $|I_u \cap I_v| = i$ for $0 \le i < l_v$, $|I_u \cap I_v| = l_v$ for $l_v \le i \le l_u$, and $|I_u \cap I_v| = l_u + l_v - i$ for $l_u < i \le l_u + l_v$. That is, we have $l_v + l_u + 1$ different cases that I_u intersects with I_v, and each of them occurs with the same probability.

Now, we turn to consider the cases that I_w intersects with $I_u \cap I_v$. The number of cases that I_w intersects with I_v is $l_v + l_w + 1$. Among them, when $0 \le i < l_v$, I_w intersects with $I_u \cap I_v$ for each $j = R(I_w)$ with j in $[0..i + l_w]$. If $l_v \le i \le l_u$, I_w always intersects with $I_u \cap I_v = I_v$. The case $l_u < i \le l_u + l_v$ is symmetric. Hence, taking average, we have

$$\begin{aligned}\Pr[I_u \sim I_w \mid I_u \sim I_v \text{ and } I_w \sim I_v] \\ = \frac{1}{l_v + l_w + 1}\left(2 \times \sum_{i=0}^{l_v-1} \frac{i + l_u + 1}{l_v + l_u + 1} + 1 \times (l_w - l_v + 1)\right) = \frac{l_u l_v + l_v l_w + l_w l_u + l_u + l_v + l_w + 1}{(l_u + l_v + 1)(l_w + l_v + 1)}.\end{aligned}$$

In the other two cases ($l_v < l_w, l_u$, and $l_w, l_u < l_v$), we can analyze in a similar way, and obtain equations which imply the same results. ∎

Hereafter, we denote by $f(l_v; l_u, l_w) = \frac{l_u l_v + l_v l_w + l_w l_u + l_u + l_v + l_w + 1}{(l_u + l_v + 1)(l_w + l_v + 1)}$. It is easy to check that for any fixed positive integer l_v, $f(l_v; l_u, l_w)$ is a nondecreasing function for l_u and l_w. We also note that $f(0; l_u, l_w) = 1$ for any l_u and l_w, which means that any two intervals I_u and I_w intersecting with I_v of length 0, I_u and I_w share a common interval I_v, which is a point.

Now, we turn to the computation of the lower bound of the expected value of $\mathsf{CC}(G)$. We denote by $\mathsf{CC}(V_i)$ the expected value of the clustering coefficient of a vertex in V_i. Then we have $\mathsf{CC}(G) = \frac{1}{n}\sum_{i=0,1,\ldots} n_i \mathsf{CC}(V_i)$. In this section, our goal is to give a good lower bound of $\mathsf{CC}(V_i)$. In our model, first few V_is are influential. Hence we can give a good lower bound by analyzing them.

Lemma 13 *We have* $\mathsf{CC}(V_0) = 1$, *and* $\mathsf{CC}(V_1) > (63n^2 - 9n_0^2 - n_1^2 - 18n_0 n - 6n_1 n - 6n_0 n_1 - 183n + 51n_0 + 15n_1 + 112)/(72(n-2)(n-1))$.

Proof. Let v be any vertex in V_i, $V' = V \setminus \{v\}$, and $V'_i = V_i \setminus \{v\}$. Then, by definition, $\mathsf{CC}(V_i)$ is computed by $\mathsf{CC}(V_i) = \frac{1}{|V'|(|V'|-1)} \sum_{u \in V'} \sum_{w \in V' \setminus \{u\}} f(i; l_u, l_w)$. When $i = 0$, it is easy to have $\mathsf{CC}(V_0) = 1$. We assume $i = 1$. We then divide into six cases;

(1) Case $u, w \in V_0$ occurs $n_0(n_0 - 1)$ times with coefficient $f(1; 0, 0) = 1/2$. We have a partial summation $\sum_{u \in V_0} \sum_{w \in (V_0) \setminus \{u\}} f(1; 0, 0) = \frac{1}{2} n_0(n_0 - 1)$.

(2) Case $u \in V_0$ and $w \in V'_1$ (or vice versa) occurs $2n_0(n_1 - 1)$ times with coefficient $f(1; 0, 1) = 2/3$. They give $2\sum_{u \in V_0} \sum_{w \in V'_1} f(1; 0, 1) = \frac{4}{3} n_0(n_1 - 1)$.

(3) Case $u \in V_0$ and $w \in V \setminus (V_0 \cup V_1)$ (or vice versa) occurs $2n_0(n - n_0 - n_1)$ times. In the case, since f is nondecreasing function, the coefficient is at least $f(1; 0, 2)$. Thus we have a lower bound $2\sum_{u \in V_0} \sum_{w \in V \setminus (V_0 \cup V_1)} f(1; 0, 2) = \frac{3}{2} n_0(n - n_0 - n_1)$.

(4) Case $u, w \in V'_1$ occurs $(n_1 - 1)(n_1 - 2)$ times with coefficient $f(1; 1, 1) = 7/9$. Hence we have a partial summation $\frac{7}{9}(n_1 - 1)(n_1 - 2)$.

(5) Case $u \in V'_1$ and $w \in V \setminus (V_0 \cup V_1)$ (or vice versa) occurs $(n_1 - 1)(n - n_0 - n_1)$ times. Similarly, a lower bound of the coefficient is given by $\frac{5}{3}(n_1 - 1)(n - n_0 - n_1)$.

(6) Case $u, w \in V \setminus (V_0 \cup V_1)$ occurs $(n - n_0 - n_1)(n - n_0 - n_1 - 1)$ times. Coefficient is bounded by $\frac{7}{8}(n - n_0 - n_1)(n - n_0 - n_1 - 1)$.

Now, we have

$$\begin{aligned}\mathsf{CC}(V_1) &= \frac{1}{|V'|(|V'|-1)} \sum_{u \in V'} \sum_{w \in V' \setminus \{u\}} f(1; l_u, l_w) \\ &> \frac{1}{(n-1)(n-2)}\left(\frac{1}{2} n_0(n_0 - 1) + \frac{4}{3} n_0(n_1 - 1) + \cdots + \frac{7}{8}(n - n_0 - n_1)(n - n_0 - n_1 - 1)\right)\end{aligned}$$

which equals to $(63n^2 - 9n_0^2 - n_1^2 - 18n_0 n - 6n_1 n - 6n_0 n_1 - 183n + 51n_0 + 15n_1 + 112)/(72(n-2)(n-1))$. ∎

By the equation of Lemma 13, we have a lower bound of $\mathsf{CC}(G)$ for fixed α. For example, letting $n_0 = 0.641n$ and $n_1 = 0.150n$ (see Example 1), we have $\mathsf{CC}(G) = \frac{1}{n}\sum_{i=0,1,\ldots} n_i \mathsf{CC}(V_i) > \frac{1}{n}\sum_{i=0,1} n_i \mathsf{CC}(V_i) = \frac{1}{n}(n_0 +$

$\frac{46.2647n^2-148.059n+112}{72(n-1)(n-2)}n_1) = \frac{1}{n}(0.641n + \frac{452n^2-1395n+1008}{648(n-1)(n-2)}0.150n) = \frac{0.737385(n-1.99726)(n-1.02891)}{(n-1)(n-2)} \sim 0.7374$. In Lemma 13, we only consider three sets V_0, V_1, and $V \setminus (V_0 \cup V_1)$. We can repeat the idea once more, and obtain a better lower bound:

Lemma 14 *We have* $\mathsf{CC}(V_1) > (1656n^2 - 324n_0^2 - 64n_1^2 - 9n_2^2 - 432n_0n - 192n_1n - 72n_2n - 288n_1n_0 - 108n_2n_0 - 48n_2n_1 - 4776n + 1476n_0 + 576n_1 + 201n_2 + 2800)/(1800(n-1)(n-2))$ *and* $\mathsf{CC}(V_2) > (500n^2 - 100n_0^2 - 25n_1^2 - 4n_2^2 - 200n_0n - 100n_1n - 40n_2n - 100n_1n_0 - 40n_2n_0 - 20n_2n_1 - 1460n + 540n_0 + 245n_1 + 92n_2 + 921)/(600(n-1)(n-2))$.

Proof. We now consider four sets V_0, V_1, V_2 and $V \setminus (V_0 \cup V_1 \cup V_2)$. The computations are straightforward and tediously, and hence omitted. ∎

Using the equations of Lemma 14, we have a better lower bound of $\mathsf{CC}(G)$ for fixed α. For example, letting $n_0 = 0.641n$, $n_1 = 0.150n$, and $n_2 = 0.064n$, we have $\mathsf{CC}(G) = \frac{1}{n}\sum_{i=0,1,\dots} n_i\mathsf{CC}(V_i) > \frac{1}{n}\sum_{i=0,1,2} n_i\mathsf{CC}(V_i) = \frac{1}{n}\left(n_0 + \frac{1178.5n^2-3730.62n+2800}{1800(n-1)(n-2)}n_1 + \frac{301.125n^2-1071.22n+921}{600(n-1)(n-2)}n_2\right) = \frac{0.771328(n-1.99648)(n-1.04782)}{(n-1)(n-2)} \sim 0.7713$.

6 Concluding remarks

In this paper, we have proposed the scale free interval graph model, and analyzed that it has power law degree distribution and large clustering coefficient. Actually, we had considered the time-continuous model which is almost same as the time-discrete model introduced in Section **??**. For the time-continuous model, we also showed following results.

Theorem 15 *A scale free interval graph generated according to our time-continuous model has the degree sequence following $\mathcal{P}(\alpha)$ for large degrees.*

Theorem 16 *For $\alpha = 2.1$ (same as Example 1), A scale free interval graph generated according to our time-continuous model has the expected clustering coefficient of G is at least 0.159. Note that it is independent of the size of the graph.*

However, our model seems to not satisfy the third property (SW). The property (SW) is that any two nodes are joined by short path, which is estimated by average or longest distance between any two nodes in G. Our experimental results showed that the average distance and the diameter of the graph are both linear in n. We leave for future works that proposing scale free interval graph model that has the property (SW).

Acknowledgment

This research was supported in part by JSPS Global COE program ”Computationism as a Foundation for the Sciences”.

References

[1] A. Barabási. *Linked: The New Science of Networks*. Perseus Books Group, 2002.

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

[3] C. Cooper and A. Frieze. The Cover Time of Two Classes of Random Graphs. In *Proc. 16th Ann. ACM-SIAM Symp. on Discrete Algorithms*, pages 961–970. ACM, 2005.

[4] D. R. Cox and V. Isham. Point Processes. Chapman & Hall, 1980.

[5] N. Korte and R. H. Möhring. An incremental linear-time algorithm for recognizing interval graphs. SIAM Journal on Computing, 18(1):68–81, 1989.

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

[7] L. Takács. *Introduction to the Theory of Queues*. Oxford University Press, 1962.

[8] D. J. Watts. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, 2004.

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

[10] R. W. Wolff. Poisson Arrivals See Time Averages. *Operations Research*, 30:223-231, 1982.