# Scale Free Interval Graphs [1]

Naoto Miyoshi [a] Takeya Shigezumi [a,*] Ryuhei Uehara [b]
Osamu Watanabe [a]

[a] *Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Japan.*

[b] *School of Information Science, JAIST, Japan.*

**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 a natural and simple random model for generating scale free interval graphs. The model generates a set of intervals randomly under a certain distribution, which defines a random interval graph. The main advantage of the model is its simpleness. The structure/properties of the generated graphs are analyzable by relatively simple probabilistic and/or combinatorial arguments, which is different from many other models. Based on such arguments, we show for our random interval graph that its degree distribution follows the power law, and that it has a large average cluster coefficient.

*Key words:* scale free graph, small world network, interval graphs

## 1 Introduction

Since early works by Watts & Strogatz [19] and Barabási & Albert [2], small world networks and scale free networks are the focus of recent interest because of their potential as models for interaction networks of complex systems in real

* Corresponding author.
*Email addresses:* miyoshi@is.titech.ac.jp (Naoto Miyoshi), Takeya.Shigezumi@is.titech.ac.jp (Takeya Shigezumi), uehara@jaist.ac.jp (Ryuhei Uehara), watanabe@is.titech.ac.jp (Osamu Watanabe).

[1] An early version was presented at the Algorithmic Aspects in Information and Management (AAIM 2008). This research was supported in part by JSPS Global COE program "Computationism as a Foundation for the Sciences".

world [1,18]. Some properties on a graph $G$ have been used as major properties to characterize graphs that are called complex networks. Among them, the scale free property introduced by Barabási and Albert [2] is one of the most investigated property in the last decade. $G$ is called a scale free network (SF) if the degree distribution of $G$ follows a power law distribution. There are two other major properties [19] that are often observed in complex networks; one is to have a relatively large clustering coefficient (CC) and another is to have a relatively small diameter (SW). Though many models for generating graphs with the scale free property and/or the other two properties were proposed and investigated, up to now, aside from few deterministic models, most of randomized models were based on some dynamic *recursive* construction of random graphs [1,18]. Thus it is not so easy to see combinatorial structure of obtained graphs, and analysis of their properties is rather complicated. 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 a simple random model for generating scale free interval graphs. We also give simple and clear mathematical definitions to the above somewhat vague properties for discussing them in the context of random graph models. We then show that our interval random graph $G$ satisfies the scale free property (SF). We also show that it has a large clustering coefficient (CC), that is, two neighbors of any node of $G$ are likely to have an edge between them. Unfortunately, our random graph does not satisfies the property (SW), and this point will be discussed in Concluding Remarks.

Interval graphs have many applications from scheduling to bioinformatics [10,20]. 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 $v$ have an edge in $G$ if and only if corresponding intervals $I_u$ and $I_v$ have an overlap. 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 generating intervals, and a power law distribution is used for determining intervals' lengths. Technically, our model is regarded as a *discrete immigration-death process*, where intervals are generated and terminated at integral time until a given number $n$ of intervals are generated. At each time, some number, which is determined following an independent Poisson distribution, of intervals are generated, and lengths of generated intervals are determined independently by a power law distribution.

Our random interval graph model has some natural interpretation. Each interval is regarded as a period of existence, i.e., life, of some object, and we may consider that relationships between such objects are created if there is

an overlap between their lives. The corresponding interval graph represents these relationships. One example is a graph representing a relationship among websites in a certain network community. An interval represents a time period that one website exists, i.e., from its start to the time it becomes inactive. If we may assume that some relation (e.g., having links or trackbacks) between two websites exists if and only if they coexist at some time, then the corresponding interval graph can be used as a model for the relationship among websites in this network community. Another example is a graph representing a set of activities sharing or competing for some resource such as a set of processes on computers connected to some local network. Here a life of each process can be expressed as an interval; then we may naturally assume that two processes share the local network if and only if there is an overlap between the corresponding intervals. Thus, the corresponding interval graph represents the network sharing relationship among these processes. Note also that in these examples a power law distribution of a lifespan is one of the reasonable choices. A power law distribution is derived from some simple formula (see the next section) formalizing the following rule: objects that survived long time tend to survive yet longer. This rule can be observed in several situations, e.g., the lifespan of weblogs [14], the length of data traffic on the Internet [5].

Clearly our model is too simple to be used as a model for real world networks. But due to its simplicity it is easy to introduce several modifications to adjust a model to explain some additional properties observed in the target networks, and we think that it is a good natural *basic model*.

Our random interval graph model has some variants, and we think that these variants are helpful when applying our model (or its modification) to some real world complex networks. In this paper, we first define a finite model by giving a random procedure creating a random interval graph of $n$ vertices for a given $n$. As a natural variant of this model, we may also consider the process of generating intervals during some given time period, e.g., from time 1 to $t_{\mathrm{end}}$. Let $N$ denote the number of intervals generated during this period; it can be shown that $N$ is a random variable following Poisson $(\lambda t_{\mathrm{end}})$. Then the probability of each interval representation being generated is the same as the probability that it is obtained by (i) generating $N$ intervals whose (integral) lengths are determined independently following a power law distribution, and (ii) putting them in the interval $[1, t_{\mathrm{end}}]$ so that their starting points are chosen from $\{1, ..., t_{\mathrm{end}}\}$ independently and uniformly at random. Though real networks are all finite, we will use an infinite interval graph model for showing our main results. This is because statistical properties of complex networks are analyzed asymptotically by considering the situation when $n$ goes infinity. In order to discuss such asymptotic analyses clearly and precisely, we follow an usual framework in the queueing theory and consider an infinite model; that is, we assume that the process of generating intervals starts from the infinite past and continues forever, which defines an infinite random interval graph

model.

## 2 Preliminaries

We recall some basic notions and define some notations on interval graphs. Throughout this paper, we consider only simple undirected graphs without multiedges and self loops, and we denote a graph as $G = (V, E)$, where $V$ is a set of vertices and $E$ is a set of unordered pairs $e = \{u, v\}$ of $V$ denoting edges. For any vertex $v \in V$, a vertex $u$ is called *adjacent* to $v$ if there is an edge $\{u, v\}$ in $E$. We sometimes denote by $u \sim v$ if $u$ is adjacent to $v$. The *neighborhood* of a vertex $v$ is a set $N_G(v) = \{u \in V \mid \{u, v\} \in E\}$, i.e., the set of adjacent vertices of $v$. The *degree* of $v$ is $|N_G(v)|$, which is denoted by $d_G(v)$. A sequence of distinct vertices $v_1, v_2, \ldots, v_t$ is a *path*, denoted by $(v_1, v_2, \ldots, v_t)$, if $\{v_j, v_{j+1}\} \in E$ for each $1 \leq j < t$. The *length* of a path is the number of edges on the path. For two vertices $u$ and $v$, the *distance* of these vertices, denoted by $\mathrm{dist}_G(u, v)$, is the minimum length of paths from $u$ to $v$. We define $\mathrm{dist}_G(u, v) = \infty$ if there is no path from $u$ to $v$. For a graph $G$, we define the *diameter* of the graph, $\mathrm{diam}(G) = \max_{u,v \in V} \mathrm{dist}_G(u, v)$. The graph $G$ is *connected* if $\mathrm{diam}(G) < \infty$. The subscript $G$ can be omitted if no confusion arise.

A graph $G = (V, E)$ is called an *interval graph* if there is a set of intervals $\mathcal{I} = \{I_v \mid v \in V\}$ on the real line such that for any $u$ and $v$ in $V$, $\{u, v\} \in E$ if and only if $I_u \cap I_v \neq \emptyset$. We call the set $\mathcal{I}$ of intervals an *interval representation* of the graph $G$. We write $I_u \sim I_v$ if $I_u \cap I_v \neq \emptyset$, which is equivalent to $u \sim v$ for an interval graph. For each interval $I$, we denote by $T_{\mathrm{L}}(I)$ and $T_{\mathrm{R}}(I)$ (the location of) the left and right endpoints of the interval respectively, and we use $L_I$ to denote the length of an interval $I$; hence, we have $T_{\mathrm{L}}(I) \leq T_{\mathrm{R}}(I)$, $L_I = T_{\mathrm{R}}(I) - T_{\mathrm{L}}(I)$, and $I = [T_{\mathrm{L}}(I), T_{\mathrm{R}}(I)]$. Throughout this paper, we assume that endpoints are integers. In the following, we use $[i..j]$ to denote the set of integers $\{i, i + 1, \ldots, j\}$.

### *2.1 Scale Free Network*

For characterizing graphs that can be observed as large complex networks, the property (SF) and (CC) explained in Introduction have been used, and a graph satisfying (SF) is usually called a *scale free network*. Since these properties (as well as some other major properties) are analyzed asymptotically considering the situation that the graph size $n$ goes infinity, one should be careful when defining these properties on a *finite* random graph model. In order to be precise and yet keep simplicity, we will follow in this paper the standard framework

from the queueing theory and analyze these properties on an infinite random graph model. In the following, we prepare some notions and notations for this analysis. For each property, we begin with basic definitions by considering finite graphs $G = (V, E)$ and then explain a way to discuss the property on a random infinite graph $G = (V, E)$.

(SF) Scale Free Property

Roughly speaking, by the scale free property we mean that the degree distribution of a graph follows a power law function, a function proportional to $k^{-\gamma}$ for some positive constant $\gamma$. Here we make this notion precise.

First consider a finite graph $G$ with $n$ vertices, we define the degree distribution of $G$ as the following function:

$$\delta_G(k) \;=\; \frac{|\{\, v \in V \;\mid\; d_G(v) = k \,\}|}{n}.$$

In the context of random finite/infinite graph, it is natural to consider its expectation, which is in fact equal to $\Pr[d_G(v) = k]$ shown as follows if this probability is the same for all $v \in V$ under the assumed random graph model. (Below $[\cdots]$ is used as an indicator function; that is, $[\cdots]$ is 1 if $\cdots$ holds and 0 otherwise.)

$$\begin{aligned}\mathrm{E}[\,\delta_G(k)\,] &= \frac{\mathrm{E}[\,|\{\, v \in V \;\mid\; d_G(v) = k \,\}|\,]}{n} \;=\; \frac{\mathrm{E}[\,\sum_{v\in V}[\,d_G(v) = k\,]\,]}{n} \\ &= \frac{\sum_{v\in V}\Pr[\,d_G(v) = k\,]}{n} \;=\; \Pr[\,d_G(v) = k\,].\end{aligned}$$

For the infinite random graph model that we will use for our analysis, we simply use $\Pr[d_G(v) = k]$ as the definition of the degree distribution of $G$. In other words, as we will see, we may assume that $\Pr[d_G(v) = k]$ is the same for all vertices in our infinite random graph model. Now the scale free property (on our infinite graph model) is defined to satisfy the following condition for some $\gamma > 0$ and $c > 0$ (where $v$ is any fixed vertex).

$$\text{(SF)} \qquad \lim_{k\to\infty} \frac{\Pr[\,d_G(v) = k\,]}{ck^{-\gamma}} = 1.$$

In this paper, instead of writing $\lim_{x\to\infty} f(x)/g(x) = 1$, we simply write "$f(x) \sim g(x)$ as $x \to \infty$." For example, (SF) is stated as follows.

$$\text{(SF)} \qquad \Pr[\,d_G(v) = k\,] \sim ck^{-\gamma}, \quad \text{as } k \to \infty.$$

(CC) Large Cluster Coefficient

The second property requires (again roughly speaking) that two neighbors of any node of $G$ are likely to have an edge between them. More precisely, for finite graph $G = (V, E)$, the following ratio, which we call the *cluster coefficient* of $v$, is used to discuss this property quantitatively.

$$\mathsf{CC}(v) \;=\; \frac{|\{\,\{u,w\} \in E \mid u,w \in N(v)\,\}|}{\binom{d_G(v)}{2}}$$

Recall that $d_G(v) = |N(v)|$. Here we assume that $d_G(v) \geq 2$. If $d_G(v) = 0$ or 1, i.e., if $N(v)$ is empty or $v$ has only one neighbor, we define $\mathsf{CC}(v) = 1$. The cluster coefficient of $G$, $\mathsf{CC}(G)$, is its arithmetic mean; that is, we define $\mathsf{CC}(G) = \sum_{v \in V} \mathsf{CC}(v)/|V|$.

Here again in the context of random finite/infinite graph, we consider the expectation $\mathrm{E}[\mathsf{CC}(v)]$ of $\mathsf{CC}(v)$, which is defined as follows.

$$\mathrm{E}[\,\mathsf{CC}(v)\,] \;=\; \mathrm{E}\left[\; \frac{\sum_{u,w \in V} [u \sim v \wedge v \sim w \wedge u \sim w]}{\binom{d_G(v)}{2}} \;\right]$$

Then this can be modified to the following.

$$\mathrm{E}[\,\mathsf{CC}(v)\,] \;=\; \sum_k \left( \;\Pr[d_G(v) = k] \times \frac{1}{\binom{k}{2}} \sum_{u,w \in V} \Pr[\,u \sim v \wedge v \sim w \wedge u \sim w \,|\, d_G(v) = k\,] \;\right).$$

For our random infinite graph model, we can assume that $\Pr[d_G(v) = k]$ and $\sum_{u,w \in V} \Pr[u \sim v \wedge v \sim w \wedge u \sim w | d_G(v) = k]$ are the same for all $v$. Hence from the above it follows that $\mathrm{E}[\mathsf{CC}(v)]$ is the same for all vertices $v$, and we can simply use $\mathrm{E}[\mathsf{CC}(v)]$ (for any fixed $v$) as the definition of cluster coefficient. Then we define our condition (CC) is to satisfy the following for some constant $\eta > 0$ (where $v$ is any fixed vertex).

$$\text{(CC)} \qquad \mathrm{E}[\mathsf{CC}(v)] \;\geq\; \eta.$$

### *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 process studied in the queueing theory; on the other hand, we use a power law distribution for determining the length of each interval. Here we recall basic distributions and their important properties.

We begin by explaining the Poisson distribution that is used to define our interval generating process. We say that a random variable $N$ follows the *Poisson distribution* with parameter $\lambda$ (which we denote Poisson $(\lambda)$) if it satisfies the following for any $k \geq 0$.

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

We recall below some important properties of the Poisson distribution; see, e.g., [13] for details.

Fix $k \geq 1$ and consider $k$ random variables $N_i$, $i \in [1..k]$, that follow Poisson $(\lambda)$ independently. Then the sum $N = \sum_{i=1}^{k} N_i$ also follows the Poisson distribution with parameter $k\lambda$.

Consider next the following two processes for any fixed $t \geq 1$. The first process is to generate, for each $i \in [1..t]$, a set $X_i$ of $N_i$ $i$'s, where $N_i$ follows Poisson $(\lambda)$ independently. For example, if $N_1 = 2$, $N_2 = 3$, ..., the process generates $X_1 = \{1, 1\}$, $X_2 = \{2, 2, 2\}$, and so on. Let $X$ be the multiset union of $X_1$, ..., $X_k$. The second process is defined by $N$ following Poisson $(t\lambda)$ and a die taking a value from $[1..t]$ uniformly at random. The process is to throw the die independently for $N$ times, and let $Y$ denote a multiset $\{U_1, ..., U_N\}$, where $U_j$, $j \in [1..N]$, is the outcome of the $j$th throw. It is known that these two processes define the same distribution on multisets of $[1..t]$. That is, for any multiset $S \subset [1..t]$, the following holds.

$$\Pr\left[\, X\,(= \cup_{i=1}^{t} X_i) \,= S\,\right] \;=\; \Pr\left[\, Y\,(= \{U_1, ..., U_N\}) \,= S\,\right].$$

This property guarantees the probabilistic interpretation of a variant of our model explained in Introduction.

The second distribution is one type of power law distributions that is used for specifying interval lengths. We say that a random variable $L$ on non-negative integers follows a *discrete power law distribution* with parameter $\alpha$ (which we denote $\mathcal{P}(\alpha)$) if it satisfies the following for any $k \geq 0$.

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

where $\zeta(\alpha) = \sum_{i=1}^{\infty} i^{-\alpha}$ (the Riemann's zeta function) is used for the normalization. Throughout this paper, we will consider only $\alpha > 2$. Note that a random variable $L+1$ is called as the *Zeta distribution* or the *Zipf distribution*.

We recall some basic properties of this distribution. Let $L$ be a random variable following $\mathcal{P}(\alpha)$.

First note that for any $\alpha > 2$, we have

$$\mathrm{E}[\,L\,] \;=\; \frac{\zeta(\alpha-1)}{\zeta(\alpha)} - 1. \tag{2}$$

Next we note the following relation on $\Pr[L \geq k+1 \mid L \geq k]$, which we will denote as $r_k$.

$$r_k \;=\; \Pr[L \geq k+1 \mid L \geq k] \;=\; \frac{\zeta(\alpha, k+2)}{\zeta(\alpha, k+1)}, \tag{3}$$

where $\zeta(\alpha, m) = \sum_{i=m}^{\infty} i^{-\alpha}$. Note also that this probability $r_k$ increases as $k$ increases, and this relation can be regarded as a formalization of a rule that longer intervals tend to survive yet longer.

For a random variable $X$, we denote the cumulative distribution function $F$ as

$$F(x) = \Pr[X \leq x].$$

We also denote the tail distribution function as

$$\overline{F}(x) = \Pr[X > x].$$

For two given independent random variables $X_1$ and $X_2$ following $F_1$ and $F_2$, respectively, we denote the convolution of $F_1$ and $F_2$ as

$$F_1 * F_2(x) = \Pr[X_1 + X_2 \leq x].$$

Note that the tail distribution function of the convolution of $F_1$ and $F_2$ is

$$\overline{F_1 * F_2}(x) = \Pr[X_1 + X_2 > x].$$

## 3 Scale Free Interval Graph Model

We here present our random generation model of interval graphs. We first give a concrete model for generating a graph with $n$ vertices for given parameters $n$, $\alpha$, and $\lambda$. Then we define its variant, an infinite interval graph model, which will be used for the analysis in later sections.

We use an immigration-death process [3], one of queueing models, for generating intervals, where a Poisson distribution is used to determine the number of generating intervals and a power law distribution is used to determine the

length of each generated interval. To be precise and concrete, we state our interval generation procedure as the procedure `gen_intervals` of Figure 1. For a given number $n$ (and probability parameters $\alpha$ and $\lambda$), the procedure generates $n$ intervals at discrete time $t = 1, 2, \ldots$. Two probability parameters $\alpha$ and $\lambda$ are used to determine probability distributions for generating intervals. The number of intervals generated at each time independently follows a Poisson distribution Poisson($\lambda$). On the other hand, the length or the lifespan of each generated interval independently follows $\mathcal{P}(\alpha)$. Note that our algorithm makes use of (3) for determining each lifespan following $\mathcal{P}(\alpha)$. The generation procedure terminates as soon as $n$ intervals are generated at some time $T_{\rm end}$; note that all intervals terminate at $T_{\rm end}$.

**Example 1** *Let us see the structure of our random graph for some typical parameters. For scale free networks, graphs satisfying (SF) with* $2.1 \leq \gamma \leq 3.0$ *are usually considered. As we will see in the next section, our random interval graph satisfies (SF) with* $\gamma = \alpha$. *We will also see that the smaller* $\alpha$ *gives the smaller clustering coefficient. Thus, let us consider here* $\alpha = 2.1$. *Then since* $\zeta(1.1) \simeq 10.584$ *and* $\zeta(2.1) \simeq 1.560$, *we have* $\zeta(\alpha - 1)/\zeta(\alpha) \simeq 6.784$. *Thus, the average length of intervals is* 5.784. *Also it follows from (1) that* $\Pr[L_I = 0] \simeq 0.641$ *and* $\Pr[L_I = 1] \simeq 0.150$. *As shown later, these bounds are important for bounding the cluster coefficient, and we in fact can show that the random interval graph (more precisely, its infinite graph variant) satisfies (CC) with* $\eta = 0.7120$. *On the other hand, the parameter* $\lambda$ *determines the connectivity of the generated graph. For example, by choosing* $\lambda = 3$ *we can show (Theorem 4) that the size of connected components of a generated graph is* on average $\lambda e^{\lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)}-1\right)}$, *which is about* $1.03\times10^8$ *for our choice of parameters.*

Although the procedure `gen_intervals` specifies a model, some discussion is necessary to clarify an algorithmic efficiency of our procedure. We show below that for any $\lambda \geq 1$ and $\alpha > 2$, the algorithm can be implemented to generate an interval graph representation in time linear w.r.t. $n$ on average.

We first note that for any $\lambda \geq 1$ and $\alpha > 2$, the procedure `gen_intervals`$(n, \alpha, \lambda)$ almost always terminates with $T_{\rm end} = O(n)$. More precisely, the following lemma holds for some $c_{\rm end} > 0$. (The proof, which is a standard probability analysis, is omitted here.)

**Lemma 1** *Let* $T_{\rm end}$ *be the value of* $T$ *when* `gen_intervals`$(n, \alpha, \lambda)$ *is terminated (let* $T_{\rm end} = \infty$ *if not terminated). Then for any* $\lambda \geq 1$ *and* $\alpha > 2$, *we have some* $c > 0$ *such that*

$$\Pr[\, T_{\rm end} > c_{\rm end} n \,] \;\leq\; e^{-cn}.$$

Thus, in the following, we may safely assume that $T_{\rm end} \leq c_{\rm end} n$. We also assume that for given $\lambda$ and $\alpha$, values $e_k = e^{-\lambda}\lambda^k/k!$ and $r_k = \zeta(\alpha, k+2)/\zeta(\alpha, k+1)$

are numerically computable with reasonable precision for $k$, $0 \le k \le n$ (for $e_k$) and $0 \le k \le c_{\mathrm{end}} n$ (for $r_k$). These values are precomputed and kept in a table.

The algorithm keeps currently existing intervals (their indices, starting positions, and lengths) as a doubly linked list. At each time $t$, $1 \le t \le T_{\mathrm{end}}$, the algorithm maintains this list by the following computations: (i) compute $N_t$ and add $N_t$ new intervals, and (ii) determine, for each of the existing intervals (including those generated at $t$) whether it remains in the list and then delete terminated intervals. Let $K_t$ be the number of intervals in this list at time $t$, including those newly added $N_t$ intervals. We show that both (i) and (ii) can be executed in $O(K_t)$ steps. (Precisely speaking, the algorithm needs to produce an interval representation in a standard form, i.e., as a sequence of indexed left and right endpoints in the order of their locations in the line. It is easy to see that this output task can be also done while maintaining the list in $O(K_t)$ steps at each $t$.)

Consider the task (i). For computing $N_t$ following Poisson($\lambda$), the following simple method is sufficient: generate a random number $r \in [0, 1]$ and determine $k$ such that

$$\sum_{i=0}^{k-1} e^{-\lambda} \frac{\lambda^i}{i!} \;\le\; r \;<\; \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^i}{i!}$$

holds, where $\sum_{i=0}^{-1} \cdot = 0$ conventionally, and use this $k$ as $N_t$. Note that $N_t$ can be set $n$ (and the algorithm is terminated) in the case $k \ge n$. With the precomputed table, we may be able to compute $N_t$ in $O(N_t)$ steps. Next consider (ii), that is, determining, for each interval of length $\ell \in [0..c_{end} \cdot n]$, whether the algorithm keeps it to the next time. This can be done in constant time per interval based on the precomputed value $r_\ell$. Thus, (ii) can be done in $O(K_t)$ steps, and altogether the task at time $t$ can be done in $O(K_t)$ steps.

Hence the total running time of the algorithm is $O(\sum_{t=1}^{T_{\mathrm{end}}} K_t)$. On the other hand, we have

$$\sum_{t=1}^{T_{\mathrm{end}}} K_t \;\le\; \sum_{i=1}^{n} (L_i + 1),$$

where $L_i$ denotes the length of the $i$th generated interval. Then it follows from (2) that $\mathrm{E}[\sum_{i=1}^{n} L_i + 1] \le cn$ for some constant $c > 0$. This proves the following theorem.

**Theorem 1** *For any $\lambda \ge 1$ and $\alpha > 2$, the expected running time of the procedure* `gen_intervals`$(n, \alpha, \lambda)$ *is $O(n)$.*

```
procedure gen_intervals(n, α, λ);
input:  n, α, and λ;
output: set of intervals 𝓘;
  t = 1, m = 0, 𝓘 = φ, 𝓘' = φ;
  while(m < n) {
    N_t = Poisson(λ);
    add min(N_t, n − m) intervals of length 0 to 𝓘 and 𝓘';
    for( each interval I ∈ 𝓘' ) {
      ℓ = the current length of I;
      decide it is alive at least one more step
        with probability r_ℓ = ζ(α, ℓ + 2)/ζ(α, ℓ + 1);
      if(alive) let the current length of I = l + 1;
      else(not alive) remove I from 𝓘';
    }
    m = m + N_t, t = t + 1;
  }
```

Fig. 1. procedure gen_intervals$(n, \alpha, \lambda)$

### *3.1 Infinite Interval Graph Model for Our Analysis*

The model defined above has a concrete and efficient algorithm for generating a graph with a given specified number $n$ of vertices. On the other hand, this model has some technical difficulties for discussing statistical properties, which can be avoided easily by considering its natural infinite graph variant. Thus, in the following analysis, we will consider this infinite graph model. Below we explain some of the technical difficulties, and then we introduce our infinite random graph model.

First note that under our finite graph model we cannot assume that all vertices have the same statistical properties. Suppose that vertices of $V = \{1, ..., n\}$ are assigned to intervals in the order of their generation time; then the vertex $n$ always corresponds to a length 0 interval because the generation procedure terminates as soon as the $n$th interval is generated. We can avoid this non-uniformity by simply assigning vertices uniformly at random to generated intervals. Then clearly, we can assume that the probabilities $\Pr[d_G(v) = k]$ and $\Pr[u \sim v \wedge v \sim w \wedge u \sim w | d_G(v) = k]$ are the same for all $u, v, w \in V$, which is important for deriving our simple mathematical conditions for (SF) and (CC). Unfortunately, however, this somewhat superficial solution is not enough for avoiding all technical difficulties in our analysis. Even though $\Pr[d_G(v) = k]$ is the same for all vertices $v \in V$, there is still some difference if we consider vertices assigned to, e.g., the first (leftmost) interval and the middle interval. We would like to avoid unnecessary complications due to such irregularity. Here we follow the standard framework from the queueing theory and consider a random interval graph model where the interval generation process starts

from time $t = -\infty$ and continues to time $t = +\infty$. The other points are the same as the finite random interval graph model. Under this model all vertices are statistically the same, and for example, probabilities such as $\Pr[d_G(v) = k]$ are the same for all vertices $v$ in $G$ (even if we fix some way of assigning vertices to intervals). Clearly this model is not the same as our original finite graph model; but then we may argue (separately) that the difference can be ignored if $n$ is sufficiently large.

## 4 Scale Free Property

In this section, we show that our model generates a random interval graph whose degree follows a power law distribution.

Note that our model is one of the discrete immigration-death process, which has been studied quite in depth in queueing theory, and our following analysis is derived easily from some of well known facts. Thus we omit some standard proofs here; but since somewhat a simpler proof is possible for our model, we give this simpler proof in Appendix.

As mentioned in the previous section, we consider the model where intervals are generated from time $-\infty$ and to time $+\infty$. We consider a generated interval $I$ and let it be fixed, and we analyze the number of intervals intersecting $I$, which is the degree of the vertex corresponding to $I$. For this analysis we use some random variables (see Figure 2). First recall that $T_{\mathrm{L}}(I)$ and $T_{\mathrm{R}}(I)$ denote its left and right endpoints, in other words, the starting and terminating time of $I$. We use $L_I$ to denote the length of $I$, and we use $A(I)$ to denote the number of intervals generated in $[T_{\mathrm{L}}(I)..T_{\mathrm{R}}(I)]$ except $I$ itself, precisely, those with left endpoints in $[T_{\mathrm{L}}(I)..T_{\mathrm{R}}(I)]$ except $I$. For time $t$, we define $\xi(t)$ as the number of intervals surviving at time $t$ but not including those generated at time $t$. We are mainly interested in $\xi(T_{\mathrm{L}}(I))$ because the number of intervals having an overlap with $I$ is $\xi(T_{\mathrm{L}}(I)) + A(I)$. In the following, we simply write by $\xi(I)$ for $\xi(T_{\mathrm{L}}(I))$. Thus, the target of our analysis is $\xi(I) + A(I)$.

We argue as follows. First we show that $\xi(I)$ follows the Poisson distribution $\mathrm{Poisson}\left(\lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)} - 1\right)\right)$. Note that we may assume in our model that $\xi(I)$ follows a stationary distribution. Then we show that $A(I)$ follows a power law distribution if it gets large. Finally we conclude our analysis by showing that $A(I)$ dominates $\xi(I)$ when $\xi(I) + A(I)$ is large. Our interval model has been studied in the queueing theory as one of the standard customer-service models. More specifically, it is essentially the same as the following model: (i) customers' arrival follows the Poisson($\lambda$), (ii) the number of service gates is infinite, and (iii) service time of customers follows $\mathcal{P}(\alpha)$ independently. We use known facts on this model from the literature.

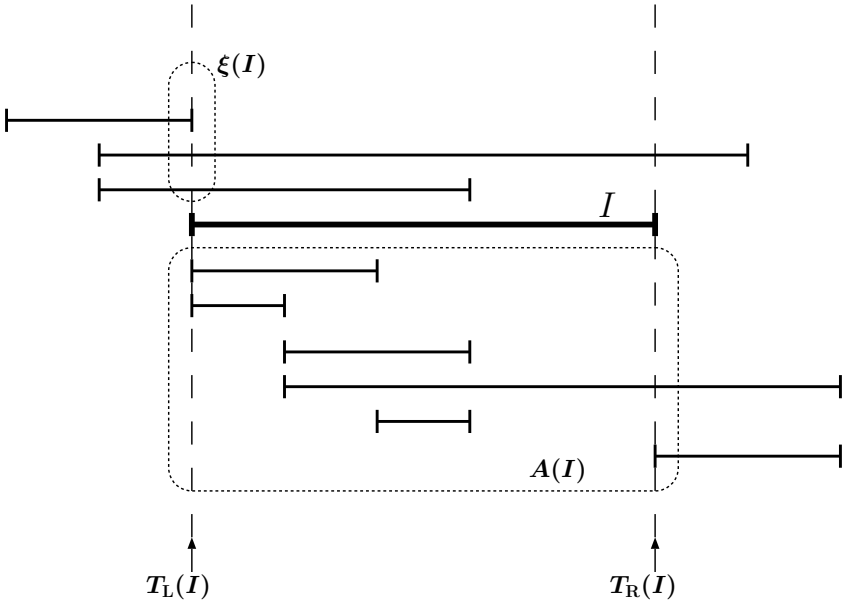


Fig. 2. An example of an interval $I$ and other intervals. There are 6 intervals at time $T_{\rm L}(I)$, 3 of them are generated at time $T_{\rm L}(I)$, and 6 intervals except $I$ start in $[T_{\rm L}(I), T_{\rm R}(I)]$. Thus, $\xi(I) = 3$ and $A(I) = 6$, and altogether 9 intervals intersects with $I$.

In [16], page 160, it is shown that the stationary distribution of the number of customers existing at time $t$ is the Poisson with parameter $\lambda\mu$ if the number of new customers follows the Poisson($\lambda$) and the average length of the service is $\mu < +\infty$. Although the result in [16] is for a homogeneous Poisson process on $\mathbf{R}_+$, the same argument works to show the same result for our "discrete" interval model. Recall that $\mu = \zeta(\alpha-1)/\zeta(\alpha) - 1$ from (2) in our model; thus, we have the following lemma. (An alternate proof of this lemma will be given in Appendix.)

**Lemma 2 [16]** *$\xi(I)$ follows Poisson* $\left(\lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)} - 1\right)\right)$.

Next we show that $A(I)$ follows a power law distribution. In fact, $A(I)$ has been studied well (see, e.g., [4], Section 8.3), and this fact is well known. We restate its derivation below. In our customer-service model, if the tail distribution of service time follows a power law distribution, then the number of customers arriving during the service also follows some power law distribution. More specifically, the following lemma is known.

**Lemma 3** *If* $\Pr[L_I > \ell] \sim c\ell^{-(\alpha-1)}$ *as* $\ell \to \infty$, *then* $\Pr[A(I) > k] \sim c\left(\frac{\lambda}{k}\right)^{\alpha-1}$ *as* $k \to \infty$.

As shown below, it is easy to see that $\Pr[L_I > \ell]$ follows a power law distribution with the exponent $\alpha - 1$; hence by using this lemma, we can show that $\Pr[A(I) > k]$ also follows a power law distribution with the same exponent.

**Fact 1**

$$\Pr[L_I > \ell] \;\sim\; \frac{1}{(\alpha-1)\zeta(\alpha)}\ell^{-(\alpha-1)}, \quad \textit{as } \ell \to \infty.$$

Proof. Note first that

$$\Pr[L_I > \ell] \;=\; \sum_{i=\ell+1}^{\infty} \Pr[L_I = i] \;=\; \frac{1}{\zeta(\alpha)} \sum_{j=\ell+2}^{\infty} j^{-\alpha}.$$

Then the fact follows from the relation

$$\frac{(\ell+2)^{-(\alpha-1)}}{\alpha-1} \;=\; \int_{\ell+2}^{\infty} t^{-\alpha}dt \;<\; \sum_{j=\ell+2}^{\infty} k^{-\alpha} \;<\; \int_{\ell+2}^{\infty}(t-1)^{-\alpha}dt \;=\; \frac{(\ell+1)^{-(\alpha-1)}}{\alpha-1}$$

and the fact that the both sides converge to $(1/(\alpha-1))\ell^{-(\alpha-1)}$. $\square$

By Lemma 3 and Fact 1, we have

$$\Pr[A(I) > k] \;\sim\; \frac{\lambda^{\alpha-1}}{(\alpha-1)\zeta(\alpha)}k^{1-\alpha}. \tag{4}$$

Next we consider the relation between $\xi(I)$ and $A(I)$. We define cumulative distribution functions $F(x)$ and $G(x)$ by

$$F(x) \;=\; \Pr[\,\xi(I) \le x\,], \;\; \text{and} \;\; G(x) \;=\; \Pr[\,A(I) \le x\,].$$

Let $\overline{F}(x)$ and $\overline{G}(x)$ be their tail distribution functions.

We would like to show that $\xi(I)$ is negligible in $\xi(I)+A(I)$. For this, we make use of a known fact on subexponential distributions. First we note that $G(x)$ is so called *subexponential* because we have the following relation;

$$\overline{G}(x) \;=\; \Pr[\,A(I) > x\,] \sim \frac{\lambda^{\alpha-1}}{(\alpha-1)\zeta(\alpha)}x^{1-\alpha},$$

and $\frac{\lambda^{\alpha-1}}{(\alpha-1)\zeta(\alpha)}$ is a constant, so $G(x)$ is a Pareto-Type distribution function (see e.g., [15] Section 2.5.2). By this fact and Theorem 2.5.2 of [15], $G$ is a subexponential distribution.

We next show that $\overline{F}(x)$ is negligible compared with $\overline{G}(x)$. More specifically, the following lemma holds.

**Lemma 4**

$$\overline{F}(x)/\overline{G}(x) \to 0, \quad \textit{as } x \to \infty. \tag{5}$$

Proof. Let us recall that $k! \geq \left(\frac{k}{3}\right)^k$ and let us $c = \frac{\lambda^{\alpha-1}}{(\alpha-1)\zeta(\alpha)}$ and $\mu = \frac{\zeta(\alpha-1)}{\zeta(\alpha)} - 1$. Since $\lim_{x\to\infty} \overline{F}(x) = 0$ and $\lim_{x\to\infty} \overline{G}(x) = 0$, applying the L'Hôpital's rule, we obtain;

$$\begin{aligned}
\lim_{x\to\infty} \frac{\overline{F}(x)}{\overline{G}(x)} &= \lim_{x\to\infty} \frac{\sum_{i=x}^{\infty} e^{-\lambda\mu} \frac{(\lambda\mu)^i}{i!}}{cx^{1-\alpha}} < \lim_{x\to\infty} \frac{e^{-\lambda\mu} \sum_{i=x}^{\infty} \left(\frac{3\lambda\mu}{i}\right)^i}{cx^{1-\alpha}} \\
&< \lim_{x\to\infty} \frac{e^{-\lambda\mu} \int_x^{\infty} \left(\frac{3\lambda\mu}{t}\right)^t dt}{cx^{1-\alpha}} \\
&= \lim_{x\to\infty} \frac{-e^{-\lambda\mu} \left(\frac{3\lambda\mu}{x}\right)^x}{c(1-\alpha)x^{-\alpha}} \qquad \text{( L'Hôpital's rule)} \\
&= \frac{e^{-\lambda\mu}(3\lambda\mu)^{\alpha}}{c(\alpha-1)} \lim_{x\to\infty} \left(\frac{3\lambda\mu}{x}\right)^{x-\alpha} = 0
\end{aligned}$$

□

Here we introduce a distribution for $\xi(I) + A(I)$; that is, define $H(k)$ by

$$H(k) \;=\; \Pr[\,\xi(I) + A(I) \leq k\,].$$

Note that this is the degree distribution that we want to analyze. We can also express $H(k)$ as the convolution of $F$ and $G$ as follows.

$$\overline{H}(k) \;=\; \Pr[\,\xi(I) + A(I) > k\,] \;=\; \overline{F * G}(k).$$

Now we make use of the following relation, derived as a special case of a well known fact (see, e.g., [15] Lemma 2.5.2).

**Lemma 5** *Let $Q$ and $R$ be any cumulative distributions on any reasonable domain, e.g., the set of nonnegative integers. If $R$ is subexponential and $\overline{Q}(x)/\overline{R}(x) \to 0$ as $x \to \infty$, then we have*

$$\frac{\overline{Q * R}(x)}{\overline{R}(x)} \to 1, \quad \text{as } x \to \infty.$$

Since our $\overline{F}$ and $\overline{G}$ satisfy the conditions of the lemma, we have $\overline{F * G}(k)/\overline{G}(k) \to 1$ as $k \to \infty$. On the other hand, we have $\overline{F * G}(k) = \overline{H}(k)$ and $\overline{G}(k)$ follows a power low distribution with exponent $\alpha - 1$ (from Equation 4). Precisely, we have

$$\overline{H}(k) \;=\; \Pr[\,\xi(I) + A(I) > k\,] \;\sim\; \frac{\lambda^{\alpha-1}}{(\alpha-1)\zeta(\alpha)} k^{1-\alpha}$$

as $k \to \infty$.

**Theorem 2** *The degree distribution* $\Pr[\, d_G(v) = k \,]$ *satisfies (SF) with* $\gamma = \alpha$. *That is, the following holds.*

$$\Pr[\, d_G(v) = k \,] \sim \frac{\lambda^{\alpha-1}}{\zeta(\alpha)} k^{-\alpha}, \quad \text{as } k \to \infty.$$

Proof. In this proof, we use the L'Hôpital's rule and consider the following $f(x)$ for real number $x$. Let $c' = \frac{\lambda^{\alpha-1}}{(\alpha-1)\zeta(\alpha)}$, $c = c'(\alpha-1)$ and $f(x) = \frac{c'\left((x-1)^{1-\alpha} - x^{1-\alpha}\right)}{cx^{-\alpha}}$. Since $\Pr[d_G(v) = k] = \overline{H}(k-1) - \overline{H}(k) = c'(k-1)^{1-\alpha} - c'k^{1-\alpha}$, it suffices for the theorem to show that $f(x) \to 1$ as $x \to \infty$.

Both $(1 + \frac{1}{x-1})^{\alpha-1} - 1$ and $\frac{1}{x}$ are continuous and converge to 0 for $x \to \infty$, using the L'Hôpital's rule, we have

$$\begin{aligned}
&\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{c'\left((x-1)^{1-\alpha} - x^{1-\alpha}\right)}{cx^{-\alpha}} = \lim_{x\to\infty} \frac{(x-1)^{1-\alpha} - x^{1-\alpha}}{(\alpha-1)x^{-\alpha}} \\
&= \lim_{x\to\infty} \frac{1}{\alpha-1} \frac{(1+\frac{1}{x-1})^{\alpha-1} - 1}{\frac{1}{x}} \\
&= \lim_{x\to\infty} \frac{1}{\alpha-1} \frac{(\alpha-1)(1+\frac{1}{x-1})^{\alpha-2}(x-1)^{-2}}{x^{-2}} \qquad \text{(L'Hôpital's rule)} \\
&= \lim_{x\to\infty} \left(1 + \frac{1}{x-1}\right)^{\alpha-2} \frac{1}{1 - \frac{2}{x} + \frac{1}{x^2}} = 1.
\end{aligned}$$

□

## 5 Clustering Coefficient

We show that our random interval graph has a large cluster coefficient; more specifically, we show that the condition (CC) holds with a large constant for a reasonable range of parameter $\alpha$. (Parameter $\lambda$ can be any number satisfying $\lambda \geq 1$.) For example, for the case $\alpha = 2.1$, from our analysis, we can show that $\mathrm{E}[\mathsf{CC}(v)] \geq 0.7120$ when we choose $\lambda = 3$.

For any vertex $v$ and its corresponding interval $I$, we first observe the following two basic facts.

**Fact 2**

$$L_I = 0 \;\Rightarrow\; \mathsf{CC}(v) = 1.$$

Proof. Since $L_I = 0$, all neighbor of $v$ are alive at the time $T_{\mathrm{L}}(I) = T_{\mathrm{R}}(I)$. Thus, any two of them are overlapping at the time $T_{\mathrm{L}}(I)$. □

**Fact 3** *(1) For any even number $d \geq 2$, we have*

$$L_I = 1 \wedge d_G(v) = d \;\Rightarrow\; CC(v) \geq \frac{d-2}{2(d-1)}.$$

*(2) For any odd number $d \geq 2$, we have*

$$L_I = 1 \wedge d_G(v) = d \;\Rightarrow\; CC(v) \geq \frac{d-1}{2d}.$$

*Note that these lower bounds are positive unless $d = 2$.*

Proof. Let us assume $L_I = 1$ and $d_G(v) = d$. $\mathsf{CC}(v)$ becomes the smallest when following three conditions stand. (See Figure 3.)

- $\xi(T_{\mathrm{L}}(I)) + N_{T_{\mathrm{L}}(I)} - 1 = \frac{d}{2}$.
- No intervals other than $I$ survive at time $T_{\mathrm{L}}(I)$.
- $N_{T_{\mathrm{R}}(I)} = \frac{d}{2}$.

In the above, $\xi(T_{\mathrm{L}}(I))$ denotes the same as in Section 4, $N_{T_{\mathrm{L}}(I)}$ and $N_{T_{\mathrm{R}}(I)}$ denotes the number of intervals generated at time $T_{\mathrm{L}}(I)$ and $T_{\mathrm{R}}(I)$, respectively. If $d$ is odd, those conditions become $\xi(T_{\mathrm{L}}(I)) + N_{T_{\mathrm{L}}(I)} - 1 = \frac{d-1}{2}$ and $N_{T_{\mathrm{R}}(I)} = \frac{d+1}{2}$, or vise versa.

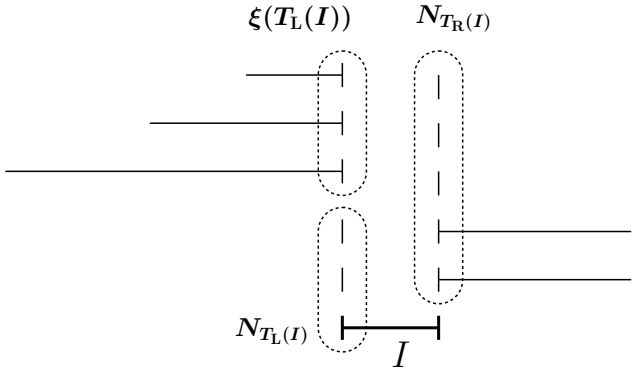


Fig. 3. An example which gives the smallest $\mathsf{CC}(v)$ when $L_I = 1$ and $d_G(v) = 10$.

In this case, $\mathsf{CC}(v) = \frac{2\binom{d/2}{2}}{\binom{d}{2}} = \frac{d-2}{2(d-1)}$ if $d$ is even, and $\mathsf{CC}(v) = \frac{\binom{(d-1)/2}{2}+\binom{(d+1)/2}{2}}{\binom{d}{2}} = \frac{d-1}{2d}$ if $d$ is odd. □

Since $\frac{d-1}{2d} > \frac{d-2}{2(d-1)}$, we have a lower bound for any $d$;

$$L_I = 1 \wedge d_G(v) = d \;\Rightarrow\; CC(v) \geq \frac{d-2}{2(d-1)}.$$

For the sake of simplicity, we use this somewhat loose lower bound in the following analysis. Using this lower bound, we have

$$\Pr\left[\mathsf{CC}(v) \geq \frac{d-2}{2(d-1)} \wedge d_G(v) = d \wedge L_I = 1\right] = \Pr\left[d_G(v) = d \wedge L_I = 1\right]. \tag{6}$$

**Theorem 3** *For any vertex $v$ and its corresponding interval $I$, we have*

$$\mathrm{E}[\mathsf{CC}(v)] > \Pr[L_I = 0] + \Pr[L_I = 1] \times e^{-C_\lambda}\left(1 + C_\lambda + \frac{1}{2}\sum_{d\geq 2}\frac{d-2}{d-1}\frac{C_\lambda^k}{d!}\right)$$

*where* $C_\lambda = \lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)} + 1\right) - 1$.

Proof. We show by a case analysis on the length $I$ and the degree of $v$. If $L_I = 0$, $\mathsf{CC}(v)$ is 1, and if $L_I = 1$, $\mathsf{CC}(v)$ has a lower bound given by Fact 3.

$$\begin{aligned}
\mathrm{E}[\mathsf{CC}(v)] &= \sum_x x \Pr[\mathsf{CC}(v) = x] \\
&> \sum_x x \Pr[\mathsf{CC}(v) = x \wedge L_I = 0] + \sum_x x \Pr[\mathsf{CC}(v) = x \wedge L_I = 1] \\
&= \Pr[\mathsf{CC}(v) = 1 \wedge L_I = 0] + \sum_x x \Pr[\mathsf{CC}(v) = x \wedge L_I = 1] \\
&= \Pr[L_I = 0] + \sum_x x \Pr[\mathsf{CC}(v) = x \wedge L_I = 1].
\end{aligned}$$

In the above, we used Fact 2, which says that $\mathsf{CC}(v) = 1$ if $L_I = 0$.

We now consider the term $\sum_x x \Pr[\mathsf{CC}(v) = x \wedge L_I = 1]$. By the definition of $\mathsf{CC}(v)$, if the degree of $v$ is 0 or 1, we have $\mathsf{CC}(v) = 1$. Hence we have

$$
\begin{aligned}
&\sum_x x \Pr[\mathsf{CC}(v) = x \wedge L_I = 1] \\
&= \sum_x x \left( \sum_{d \geq 0} \Pr[\mathsf{CC}(v) = x \wedge d_G(v) = d \wedge L_I = 1] \right) \\
&= \sum_{d \geq 0} \left( \sum_x x \Pr[\mathsf{CC}(v) = x \wedge d_G(v) = d \wedge L_I = 1] \right) \\
&= \Pr[d_G(v) = 0 \text{ or } 1 \wedge L_I = 1] \\
&\quad + \sum_{d \geq 2} \left( \sum_x x \Pr[\mathsf{CC}(v) = x \wedge d_G(v) = d \wedge L_I = 1] \right) \\
&> \Pr[d_G(v) = 0 \text{ or } 1 \wedge L_I = 1] \\
&\quad + \sum_{d \geq 2} \left( \frac{d-2}{2(d-1)} \Pr\left[ \mathsf{CC}(v) \geq \frac{d-2}{2(d-1)} \wedge d_G(v) = d \wedge L_I = 1 \right] \right) \\
&= \Pr[d_G(v) = 0 \text{ or } 1 \wedge L_I = 1] + \sum_{d \geq 2} \left( \frac{d-2}{2(d-1)} \Pr[d_G(v) = d \wedge L_I = 1] \right).
\end{aligned} \tag{7}
$$

We used Equation (6) in the above.

Recall that the degree of a vertex whose corresponding interval, say $I$, has length 1 can be represented as the sum of $\xi(T_{\mathrm{L}}(I)) + N_{T_{\mathrm{L}}(I)} - 1 + N_{T_{\mathrm{R}}(I)}$. By the analysis in Section 4, the degree of a vertex corresponding to $I$ follows the Poisson distribution with parameter $C_\lambda = \lambda \left( \frac{\zeta(\alpha-1)}{\zeta(\alpha)} - 1 \right) + \lambda - 1 + \lambda = \lambda \left( \frac{\zeta(\alpha-1)}{\zeta(\alpha)} + 1 \right) - 1$. That is, we have

$$
\begin{aligned}
\Pr[d_G(v) = d \wedge L_I = 1] &= \Pr[L_I = 1] \times \Pr[d_G(v) = d \mid L_I = 1] \\
&= \Pr[L_I = 1] \times e^{-C_\lambda} \frac{C_\lambda^d}{d!}.
\end{aligned}
$$

The last part of equation (7) is:

$$
\begin{aligned}
&\Pr[d_G(v) = 0 \text{ or } 1 \wedge L_I = 1] + \sum_{d \geq 2} \left( \frac{d-2}{2(d-1)} \Pr[d_G(v) = d \wedge L_I = 1] \right) \\
&= \Pr[L_I = 1] \times \left[ e^{-C_\lambda} + e^{-C_\lambda} C_\lambda + \sum_{d \geq 2} \left( \frac{d-2}{2(d-1)} e^{-C_\lambda} \frac{C_\lambda^d}{d!} \right) \right] \\
&= \Pr[L_I = 1] \times e^{-C_\lambda} \left( 1 + C_\lambda + \frac{1}{2} \sum_{d \geq 2} \frac{d-2}{d-1} \frac{C_\lambda^k}{d!} \right)
\end{aligned}
$$

Thus, the clustering coefficient of $v$ has the following lower bound:

$$\mathrm{E}[\mathsf{CC}(v)] > \Pr[L_I = 0] + \Pr[L_I = 1] \times e^{-C_\lambda} \left(1 + C_\lambda + \frac{1}{2} \sum_{d \geq 2} \frac{d-2}{d-1} \frac{C_\lambda^k}{d!}\right).$$

□

**Remark 1** *We proved this theorem by using only (1) of Fact 3, i.e.,* $\mathsf{CC}(v) > \frac{d-2}{2(d-2)}$*, which holds for all* $d$*. By using both bounds (1) and (2) (depending whether* $d$ *is even or odd), we can obtain the following slightly better bound:*

$$\mathrm{E}[\mathsf{CC}(v)]$$
$$> \Pr[L_I = 0] + \Pr[L_I = 1] \times e^{-C_\lambda} \left\{1 + C_\lambda + \sum_{i=1}^{\infty} \frac{i}{2i+1} \frac{C_\lambda^{2i+1}}{(2i+1)!} \left(1 + \frac{C_\lambda}{2i+2}\right)\right\}.$$

**Example 2** *Following Example 1, consider the random interval graph generation with parameter* $\alpha = 2.1$ *and* $\lambda = 3$*. Then for any* $v \in V$*, since interval length follows a power law distribution of (1), the probability* $p_0 = \Pr[|I| = 0]$ *is* 0.641 *and the probability* $p_1 = \Pr[|I| = 1]$ *is* 0.150*. We also have* $C_\lambda = 22.352$*, and by the above theorem with some arithmetic calculations, we have* $\mathrm{E}[\mathsf{CC}(v)] \geq 0.7120$*.*

## 6 Concluding Remarks

In this paper, we propose a random interval graph model and show that our random graph satisfies the scale free property; we also showed that it has a large cluster coefficient. Though we consider only a discrete time model in this paper, we can generalize our model and analysis for time-continuous models. For more details, see [11] and [12].

Unfortunately, another major property, the small world property introduced by Watts and Strogatz [19], does not seem to hold for our random interval graph. The Small World Property [2] is on the distance between any pair of vertices in a graph. In the literature, it is defined as a condition requiring for $G(V, E)$ that the average distance between pairs of vertices in $V$ is $O(\log |V|)$. However, if the graph is not connected, i.e., there is a pair of vertices with infinite distance, the average distance would become infinite. In the literature, two approaches have been usually taken for avoiding this situation. One is to consider a model that creates almost always connected graphs, and another is that the average distance is taken among the pairs in a same connected component.

[2] In the literature, e.g., [19], sometimes a small world network refers to a graph satisfying two of these conditions (CC) and this.

Our model seems inapproporiate for both approaches. First, the expected size of a connected component our model is finite even in our *infinite* random graph model. In fact, we can bound expected connected component size as follows.

**Theorem 4** *Let $G$ be a random interval graph generated by* `gen_intervals`$(n, \alpha, \lambda)$. *Then the average size of its connected components is* $\lambda e^{\lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)}-1\right)}$.

Proof. We state our analysis in terms of $\xi(t)$; recall that $\xi(t)$ is the number of intervals that surviving at time $t$ not including those generated at time $t$. Consider any time $t_0$ such that $\xi(t_0) = 0$, which means that all intervals at time $t_0 - 1$ are dead on time $t_0 - 1$. Let $T_1$ denote a random variable such that $\xi(t_0) = 0$, $\xi(t') > 1$ for $t_0 + 1 \leq t' < t_0 + T_1$, and $\xi(t_0 + T_1) = 0$. That is, $T_1$ is a time period between $t_0$ and the next time $\xi(t) = 0$ occurs. Thus, the size of a connected component can be expressed by $\sum_{t=t_0}^{t_0+T_1-1} N_t$. Using Wald's Equation (see, e.g., [13], p.300), we have

$$\mathrm{E}\left[\sum_{t=t_0}^{t_0+T_1-1} N_t\right] = \mathrm{E}\left[T_1\right] \cdot \mathrm{E}\left[N_t\right].$$

Note that $\mathrm{E}[N_t]$ is $\lambda$ for any $t$; hence it suffices for the theorem to show that

$$\mathrm{E}[T_1] \;=\; \frac{1}{\Pr[\xi(t) = 0]} \;=\; e^{\lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)}-1\right)}. \tag{8}$$

For any time $\tau$, let $k = \sum_{t=1}^{\tau}[\xi(t) = 0]$ where $[\cdots]$ is 1 if $\cdots$ occurs and 0 if otherwise, and let $T_0(k)$ be a $k$-th time such that $\xi(t) = 0$. We can easily observe that $T_0(k) \leq \tau < T_0(k+1)$, and hence $\frac{T_0(k)}{k} \leq \frac{\tau}{k} < \frac{T_0(k+1)}{k}$. Since $\Pr[\xi(t) = 0] > 0$ for any $t$, we have $k \to \infty$ as $\tau \to \infty$. Thus, by the law of large numbers, $\frac{T_0(k)}{k} \to \mathrm{E}[T_1]$ and $\frac{T_0(k+1)}{k} \to \mathrm{E}[T_1]$ as $\tau \to \infty$. So, we have $\frac{\tau}{k} \to \mathrm{E}[T_1]$ as $\tau \to \infty$. Using the law of large numbers again, we also have

$$\frac{\tau}{k} \;=\; \frac{\tau}{\sum_{t=1}^{\tau}[\xi(t) = 0]} \to \frac{1}{\Pr[\xi(t) = 0]} \quad \text{as } \tau \to \infty.$$

Hence we obtain equation (8). □

Hence, as stated in Example 1, for $\alpha = 2.1$ and $\lambda = 3$, the average component size is at most $1.03 \times 10^8$, which is large but still constant independent from $n$.

Second, consider any connected component of a given infinite random interval graph. Let $m$ denote its size. Our computer experiment shows that the avarage distance on a connected component is quite likely $\Theta(m)$.

There may be several ways to modify our model so that an obtained graph

also satisfies the small world property. But in order for proposing a reasonable one, further investigation seems necessary, and we leave it for our future work.

Another important subject is the problem of fitting our model to some observed networks. For this purpose, our model should be used as a basic model and we need to again consider some modifications. Due to the simplicity of our model, we may be able to consider several ways to modify its probability setting to create a model appropriate to observed networks. For example, instead of introducing an edge between any pair of vertices whose corresponding intervals overlap, we may consider a model where an edge is introduced with a certain probability between such vertices. Once we fix some appropriate model, the next and another interesting problem is to develop some algorithmic method to measure the closeness of an observed network to the model. For this one might want to consider some statistical analysis, and in fact, many statistical techniques have been developed for this purpose, e.g., [6]. Yet it would be nice if we have some combinatorial/algorithmic ways. For example, if a given observed network is indeed an interval graph, we can easily (i.e., in linear time) compute an interval representation of the graph [8]. If that graph is an "almost" interval graph, we may be able to fit our model to the graph by adding or deleting some edges. Unfortunately, however, it is well known fact that minimizing the number of edges added to the given graph to obtain an interval graph, which is called a minimum interval completion problem, is NP-hard in the worst case [7]. But this problem is fixed parameter tractable [17] and some heuristic approach might still work for solving the problem on average. Designing such heuristics is again our important future work.

## 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] D. R. Cox and V. Isham. *Point Processes*. Chapman & Hall, 1980.

[4] J. Grandell. *Mixed Poisson Processes*. Chapman & Hall, 1997.

[5] F. Hernandez-Campos, J.S. Marron, G. Samorodnitsky, and F.D. Smith, Variable heavy tailed durations in internet traffic: Part I, Understanding heavy tails, in *Proc. of the 10th IEEE Int'l Sympos. on Modeling, Analysis and Simulation of Computer and Telecommunications Systems* (MASCOTS 2002), IEEE, 43–50, 2002.

[6] D. R. Hunter,S. M. Goodreau and M. S. Handcock. Goodness of Fit of Social Network Models. *Journal of the American Statistical Association*, 103(481):248–258, 2008.

[7] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1979.

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

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

[10] T.A. McKee and F.R. McMorris. *Topics in Intersection Graph Theory*. SIAM, 1999.

[11] N. Miyoshi, T. Shigezumi, R. Uehara, and O. Watanabe. Scale Free Interval Graphs. *Dept. of Math. and Comp. Sciences Tokyo Institute of Technology Research Reports (Series C)*, series C-255, 2008.
http://www.is.titech.ac.jp/research/research-report/C/C-255.pdf.

[12] N. Miyoshi, M. Ogura, T. Shigezumi, and R. Uehara. Scale-free Interval Graphs Generated by Immigration-death Processes. *International Workshop on Applied Probability (IWAP 2008)*.

[13] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[14] S. Morimoto, *Scale Free Graphs Based on Time Sequential Data* (in Japanese), (availabe from `http://hdl.handle.net/10119/8101`), Master Thesis, School of Information Science, Japan Advanced Inst. of Tech., 2009.

[15] T. Rolski, H. Schmidli, V. Schmidt, J. Teugels. *Stochastic processes for insurance and finance*. J. Wiley, Chichester, 1999.

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

[17] Y. Villanger, P. Heggernes, C. Paul, and J. A. Telle. Interval Completion Is Fixed Parameter Tractable. *SIAM Journal on Computing, 38(5):2007-2020, 2009*.

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

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

[20] P. Zhang,E. A. Schon, S. G. Fischer, E. Cayanis, J. Weiss, S. Kistler and P. E. Bourne. An algorithm based on graph theory for the assembly of contigs in physical mapping of DNA. *Bioinformatics* 10(3):309–317, 1994.

## Appendix

We here present another proof of Lemma 2.

**Lemma 2 [16]** *For any $t$, $\xi(t)$ follows Poisson $\left(\lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)}-1\right)\right)$.*

Consider the time $t$ of the generation process of our infinite graph model. Some intervals exist at time $t$ and each of them has their current length $\geq 0$. Recall that the $r_\ell$ (for $\ell \geq 0$) is the probability such that an interval having current length $\ell$ at time $t$ will survive at time $t+1$. $r_\ell$ is derived from equation (3).

Let $\rho_\ell^t$ be the number of intervals which are alive and have current length $\ell$ at time $t$. As the time $t$ proceeds, $\rho_{\ell+1}^{t+1}$ depends only on $\rho_\ell^t$ because some of intervals of $\rho_\ell^t$ will survive at time $t+1$ with probability $r_\ell$ and others end at time $t$. From this observation, we obtain this formula for $\ell \geq 0$:

$$\Pr[\rho_{\ell+1}^{t+1}=k] = \sum_{m=k}^{\infty} \binom{m}{k} r_\ell^k (1-r_\ell)^{m-k} \Pr[\rho_\ell^t = m]. \tag{9}$$

Since $\rho_0^t$ is the number of intervals starting at time $t$, $\Pr[\rho_0^t = k] = e^{-\lambda}\frac{\lambda^k}{k!}$. Let us consider the stationary distribution $\pi_\ell$ such that $\pi_\ell(k) = \lim_{t\to\infty}\Pr[\rho_\ell^t = k]$. For the stationary distribution $\pi_\ell$, applying the dominated convergence theorem since $\pi_l(m) \leq 1$, the equation (9) becomes

$$\pi_{\ell+1}(k) = \sum_{m=k}^{\infty} \binom{m}{k} r_\ell^k (1-r_\ell)^{m-k} \pi_\ell(m) \tag{10}$$

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

We will show the following lemma as the solution of the equation (10).

**Lemma A1** *Let us denote $P_\ell = \prod_{j=0}^{\ell-1} r_j$ for $\ell \geq 1$ and $P_0 = 1$. The stationary distribution $\pi_\ell$ follows Poisson $(\lambda P_\ell)$;*

$$\pi_\ell(k) = e^{-\lambda P_\ell}\frac{(\lambda P_\ell)^k}{k!}.$$

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

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

Note that we used $r_\ell P_\ell = P_{\ell+1}$ in the above. □

We finally obtained the Lemma 2.

**Lemma 2** **[16]** *For any $t$, $\xi(t)$ follows Poisson* $\left(\lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)}-1\right)\right)$.

Proof. Since $\pi_\ell$s are independent and follow Poisson $(\lambda P_\ell)$, the sum $\xi(T) = \sum_{\ell=1}^{\infty}\pi_\ell$ also follows the Poisson distribution with parameter $\sum_{\ell=1}^{\infty}\lambda P_\ell = \lambda\left(\frac{\zeta(\alpha-1)}{\zeta(\alpha)}-1\right)$. □