Discussiones Mathematicae
Graph Theory 28 (2008) 511–533

THE SIZES OF COMPONENTS IN RANDOM CIRCLE GRAPHS

Ramin Imany-Nabiyyi

Shahid Beheshti University
Department of Statistics, Evin, Tehran, Iran

e-mail: imany@tabrizu.ac.ir

Abstract

We study random circle graphs which are generated by throwing n points (vertices) on the circle of unit circumference at random and joining them by an edge if the length of shorter arc between them is less than or equal to a given parameter d. We derive here some exact and asymptotic results on sizes (the numbers of vertices) of "typical" connected components for different ways of sampling them. By studying the joint distribution of the sizes of two components, we "go into" the structure of random circle graphs more deeply. As a corollary of one of our results we get the exact, closed formula for the expected value of the total length of all components of the random circle graph. Although the asymptotic distribution for this random characteristic is well known (see e.g. T. Huillet [4]), this surprisingly simple formula seems to be a new one.

Keywords: random interval graphs, geometric graphs, geometric probability.

2000 Mathematics Subject Classification: 05C80, 60D05.

1. Introduction

Consider a circle of unit circumference. Choose n points, as vertices, uniformly and independently on the circle. Put an edge between two vertices if and only if the length of the shorter arc between them is less than or equal to a given parameter d. We call the resulting graph a random circle graph, and we denote it by $\mathcal{CG}_{n,d}$. This model has been studied by many authors (see for example [1, 3, 4, 5, 7, 8]) and it is closely related to random interval graphs — the special case of random coincidence graphs (see Godehardt and Jaworski [2] for results for random interval graphs and Penrose [6] for random geometric graphs in the q-dimensional space, $q \geq 1$).

In this paper we use methods which correspond to the techniques used by Godehardt and Jaworski in [2]. Most of the proofs are based on the main Lemma 1 from that paper and therefore we recall it in Section 2. In Section 3, we introduce three methods of sampling a component of random circle graph and we prove some new results about its size (the number of vertices) and its length (the length of the covered arc). In particular, as a consequence of one of these results we obtain the surprisingly simple formula for the expected value of the total length of all components of a random circle graph. Although the asymptotic distribution for this random characteristic is known (see e.g. T. Huillet [4]), this exact formula seems to be a new one.

We also introduce two different methods for sampling two connected components of a random circle graph. By studying the joint distribution of the sizes of two such components we go into the structure of random circle graphs more deeply. In particular, we show that just before the connectedness, asymptotically almost surely (a.a.s.) any random circle graph has just two components, both with sizes of order $\Theta(n)$.

2. Preliminary Results

Let $\mathcal{CG}_{n,d}$ be a random circle graph on n vertices and with the distance level d. All theorems in this section, concerning properties of $\mathcal{CG}_{n,d}$, can be proven in a similar way as corresponding theorems for random interval graphs in [2] and therefore we omit their proofs here. However, let us mention that since in random circle graphs there is no border effect, some parts of the proofs may be simplified. On the other hand, since we may have a "coverage problem" (corresponding to the event that there is an arc between any two consecutive points, i.e., for any pair of vertices u, v we may have two paths in the clockwise direction — from u to v and from v to u, which, in fact, form a Hamiltonian cycle), we are facing some other difficulties as compared to the interval graphs. This is one of the reasons which makes studying circle graphs separately worthwhile, especially the exact results.

Let C_n be the number of *gaps* — arcs between consecutive vertices of length greater than d. Obviously, the event $C_n = r \geq 2$ means that the random circle graph $\mathcal{CG}_{n,d}$ has r such gaps and thus has r connected components. If $r = 1$, then $\mathcal{CG}_{n,d}$ has only one component and this component does not cover the whole circle; $r = 0$ means that there is no such gap, so $\mathcal{CG}_{n,d}$ also has only one component which now covers the whole circle. In the following theorem we give the distribution of the number of components in terms of the number of gaps. The result given in Theorem 1 is originally due to Stevens [10]. A straightforward proof follows also from the spacing approach in a similar way as for random interval graphs in [2]. In [4], Huillet gave another proof by using Steutel's calculus (see also [9]).

Theorem 1. *Let C_n be the number of gaps between the components in a random circle graph $\mathcal{CG}_{n,d}$. The discrete probability distribution of C_n is given by*

$$\Pr\{C_n = r\} = \binom{n}{r} \sum_{j=r}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n-r}{j-r} (-1)^{j+r}(1-jd)^{n-1},$$

for $r = 0, 1, \ldots, n$.

Note, once more, that in the circle model, the probability of connectedness is equal to the sum of probabilities of two events: (A) there is no gap (the whole circle is covered), (B) there is exactly one gap.

Let us recall now the main lemma from [2] which is crucial in proving most of our results. It is obvious that it can be easily transformed to the case of random circle graphs.

Lemma 1. *Let α be an arc of length y of the circle of unit circumference. Let two out of k given vertices be placed at the borders of this arc. Let $A_{k,y,d}$ be the event that $k-2$ points, corresponding to the remaining vertices and randomly drawn from circle circumference, are inside α and "join" the borders, that is, the k vertices form a connected subgraph of length y; and let $P(k,y,d) = \Pr\{A_{k,y,d}\}$. Then*

$$P(k,y,d) = \sum_{j=0}^{min\{k-1,\lfloor y/d \rfloor\}} \binom{k-1}{j} (-1)^j (y-jd)^{k-2}.$$

The next lemma will be needed in Section 4, and we will use it in the proof of Theorem 2. One can prove this lemma in a similar way as it has been done in [2]; note only that, since we have here a "coverage problem" (we may have paths between two given vertices on both "sides" of the circle), we have to subtract the coverage probability to get the result.

Lemma 2. *Let P_2 denote the probability that two given vertices are joined by a path, i.e., they are in the same connected component. Then*

$$P_2 = 2 - \frac{2}{n-1} \sum_{j=2}^{min\{n,\lfloor 1/d \rfloor+1\}} \binom{n}{j} (-1)^j (1-(j-1)d)^{n-1}$$
$$- \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j} (-1)^j (1-jd)^{n-1} .$$

Let $\mathcal{CS}_k$ and $\mathcal{CC}_k$ denote the events that k given vertices form a perfect connected subgraph ("perfect" means here the additional property that, between the realizations of the vertices — the respective points on the circle, there are no other "vertices" of the circle graph) and a connected component in a random circle graph, respectively. In a similar way as for random interval graphs, we can derive the probability of $\mathcal{CS}_k$, and consequently we can derive the probability of $\mathcal{CC}_k$. Finally, using the formula for this probability one can derive the probability distribution of the size of the connected component containing a given vertex. For the proof of the second part of the following theorem we can use Theorem 1, while the formulas for moments follow from Lemma 2 and the corresponding result for three vertices.

Theorem 2. *Let $K_n(v)$ be the size (i.e., the number of vertices) of the connected component containing a given vertex v in a random circle graph $\mathcal{CG}_{n,d}$. Then we have*

$$\Pr\{K_n(v) = k\} = k \sum_{j=0}^{min\{k-1,\lfloor 1/d \rfloor-2\}} \binom{k-1}{j} (-1)^j (1-(j+2)d)^{n-1}$$

for $k = 1, 2, \ldots, n-1$, and

$$\Pr\{K_n(v) = n\} = \Pr\{\mathcal{CS}_n\} = \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j} (-1)^j (1-jd)^{n-1}$$
$$+ n \sum_{j=0}^{min\{n-1,\lfloor 1/d \rfloor-1\}} \binom{n-1}{j} (-1)^j (1-(j+1)d)^{n-1}.$$

The first two factorial moments of $K_n(v)$ *are*

$$\begin{aligned}\mathrm{E}(K_n(v)) = &-1 - 2 \sum_{j=1}^{min\{n,\lfloor 1/d \rfloor+1\}} \binom{n}{j}(-1)^j(1-(j-1)d)^{n-1} \\ &- (n-1) \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j}(-1)^j(1-jd)^{n-1},\end{aligned}$$

$$\begin{aligned}\mathrm{E}_2(K_n(v)) = 2 &- 2(3n-1) \sum_{j=1}^{min\{n,\lfloor 1/d \rfloor+1\}} \binom{n}{j}(-1)^j(1-(j-1)d)^{n-1} \\ &- 6 \sum_{j=2}^{min\{n+1,\lfloor 1/d \rfloor+2\}} \binom{n+1}{j}(-1)^j(1-(j-2)d)^{n-1} \\ &- 2(n-1)^2 \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j}(-1)^j(1-jd)^{n-1}.\end{aligned}$$

Note that the above theorem gives us the size of a typical connected component which was sampled by choosing a vertex first (size-biased sampling). In the next section we consider another method of sampling — length-biased sampling.

3. Sampling One Component

3.1. Exact results

Each component in a random circle graph $\mathcal{CG}_{n,d}$ covers an arc on the circle; thus for each such component there is a corresponding interval (of the length equal to the length of the arc). We call the union of these arcs over all components the "covered part" of the circle. More precisely, let us choose the first (in the clockwise direction) border vertex (say y_1) of one of the components as the origin on the circle circumference and assume that our random circle graph has k components, and $0 = y_1 \le z_1 < y_2 \le z_2 < \cdots < y_k \le z_k \le 1-d$ are their border vertices. Then the sum of arcs: $[y_1, z_1] \cup [y_2, z_2] \cup \cdots \cup [y_k, z_k]$, is the covered part of the circle.

Consider a point on the circle (almost surely it will not be a vertex). Consider the following method of sampling a typical connected component of our circle graph: if the given point is in the covered part of the circle, we choose the connected component for which the corresponding arc covers the point; otherwise we choose the connected component which is the first after the given point (in the clockwise direction).

Theorem 3. *Let A be a given point on the circle and let $K_n(A)$ denote the size of the component which either covers A — or if there is no component which covers A — the size of the component which is the first after A in the clockwise direction. Then*

$$\begin{aligned}\Pr\{K_n(A)=k\} = k &\sum_{j=0}^{min\{k-1,\lfloor 1/d\rfloor-2\}} \binom{k-1}{j}(-1)^j(1-(j+2)d)^n \\ &- nd(k-1)\sum_{j=0}^{min\{k-2,\lfloor 1/d\rfloor-3\}}\binom{k-2}{j}(-1)^j(1-(j+3)d)^{n-1} \\ &+ nd\sum_{j=0}^{min\{k-1,\lfloor 1/d\rfloor-2\}}\binom{k-1}{j}(-1)^j(1-(j+2)d)^{n-1}\end{aligned}$$

for $k=2,\dots,n-1$, and

$$\Pr\{K_n(A)=1\} = (1-2d)^n + nd(1-2d)^{n-1}, \quad \Pr\{K_n(A)=n\} = \Pr\{\mathcal{CS}_n\}.$$

Proof. Let H be the event that a given point A is in the covered part, and let $2 \le k < n$ (clearly, the event H has the probability 0 for $k=1$). Let us choose a point A as the origin on the circle and assume that A is in the covered part; let x denote the distance between A and the starting point of the component which covers A, and let y be the length of that component (note that the border point of the component corresponding to x is before the point A in the clockwise direction). In order to derive the probability of the event "A is on the covered part and the size of the component which covers A is k", first we calculate it for a given x and y and then we integrate it over all possible x and y. We can choose k out of n vertices in $\binom{n}{k}$ ways, and then we can place two of them as border vertices of a component in $k(k-1)$ ways.

With probability $P(k,y,d)$ the k chosen vertices form a perfect connected subgraph, and with probability $(1-y-2d)^{n-k}$ the remaining vertices are outside the respective component. Thus

$$\Pr\{K_n(A)=k\wedge H\} = \int_0^{1-2d}\int_{1-y}^{1} k(k-1)\binom{n}{k}P(k,y,d)(1-y-2d)^{n-k}dx\,dy.$$

If A is not in the covered part and $2\le k<n$, then we consider two cases for x: x is smaller than d, or x is larger than d (note that the border point of the component corresponding to x is now after the point A in the clockwise direction). In a similar way to the previous case we obtain:

$$\Pr\{K_n(A)=k\wedge H^c\} = k(k-1)\binom{n}{k}\int_0^{d}\int_0^{1-2d} P(k,y,d)(1-y-2d)^{n-k}dy\,dx$$

$$+\,k(k-1)\binom{n}{k}\int_d^{1-d}\int_0^{1-x-d} P(k,y,d)(1-y-d-x)^{n-k}dy\,dx.$$

By using Lemma 1 and tedious but standard calculations, the proof can be completed. ∎

Theorem 3 directly implies the following corollary.

Corollary 1. *Let us assume that the event H, that a given point A is in the covered part in a random circle graph $\mathcal{CG}_{n,d}$, holds, and let $K'_n(A)$ be the size of the connected component which covers A. Then*

$$\Pr\{K'_n(A)=k\} =$$

$$=\frac{1}{\Pr\{H\}}\Bigg[(k-1)\sum_{j=0}^{min\{k-1,\lfloor 1/d\rfloor-2\}}\binom{k-1}{j}(-1)^j(1-(j+2)d)^n$$

$$-\,n(k-1)d\sum_{j=0}^{min\{k-2,\lfloor 1/d\rfloor-3\}}\binom{k-2}{j}(-1)^j(1-(j+3)d)^{n-1}\Bigg]$$

for $2\le k<n$, and

$$\Pr\{K'_n(A) = n\} =$$

$$= \frac{1}{\Pr\{H\}} \Bigg[(n-1) \sum_{j=0}^{min\{n-1,\lfloor 1/d \rfloor -1\}} \binom{n-1}{j} (-1)^j (1-(j+1)d)^n$$

$$- n(n-1)d \sum_{j=0}^{min\{n-2,\lfloor 1/d \rfloor -2\}} \binom{n-2}{j} (-1)^j (1-(j+2)d)^{n-1}$$

$$+ \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j} (-1)^j (1-jd)^{n-1} \Bigg].$$

Let $L_n(d)$ denote the length of the covered part of $\mathcal{CG}_{n,d}$ (which we call the total length of the random circle graph). By the definition of expected value, $\mathrm{E}(L_n(d)) = \Pr\{H\}$. On the other hand using the probability distribution given in Corollary 1, we can obtain the next result.

Corollary 2. *Let $L_n(d)$ be the total length of the random circle graph $\mathcal{CG}_{n,d}$, then*

$$\mathrm{E}(L_n(d)) = 1 - (1 + (n-1)d)(1-d)^{n-1}.$$

Proof. It is obvious that $\mathrm{E}(L_n(d)) = \Pr\{H\}$, and therefore after summing up probabilities given in Corollary 1 we obtain

$$\mathrm{E}(L_n(d)) = \sum_{k=1}^{n-1} (k-1) \sum_{j=0}^{min\{k-1,\lfloor 1/d \rfloor -2\}} \binom{k-1}{j} (-1)^j (1-(j+2)d)^n$$

$$- nd \sum_{k=1}^{n-1} (k-1) \sum_{j=0}^{min\{k-2,\lfloor 1/d \rfloor -3\}} \binom{k-2}{j} (-1)^j (1-(j+3)d)^{n-1}$$

$$+ (n-1) \sum_{j=0}^{min\{n-1,\lfloor 1/d \rfloor -1\}} \binom{n-1}{j} (-1)^j (1-(j+1)d)^n$$

$$- n(n-1)d \sum_{j=0}^{min\{n-2,\lfloor 1/d\rfloor -2\}} \binom{n-2}{j}(-1)^j(1-(j+2)d)^{n-1}$$

$$+ \sum_{j=0}^{min\{n,\lfloor 1/d\rfloor\}} \binom{n}{j}(-1)^j(1-jd)^{n-1}.$$

Note that

$$\sum_{j=0}^{min\{n-2,\lfloor 1/d\rfloor -2\}} (-1)^j(1-(j+2)d)^n \sum_{k=j+1}^{n-1}(k-1)\binom{k-1}{j}$$

$$- nd \sum_{j=0}^{min\{n-3,\lfloor 1/d\rfloor -3\}} (-1)^j(1-(j+3)d)^{n-1} \sum_{k=j+2}^{n-1}(k-1)\binom{k-2}{j}$$

$$= \sum_{j=0}^{min\{n-2,\lfloor 1/d\rfloor -2\}} (-1)^j(1-(j+2)d)^n \left[(j+1)\sum_{k=j+1}^{n-1}\binom{k}{j+1} - \sum_{k=j+1}^{n-1}\binom{k-1}{j}\right]$$

$$- nd \sum_{j=0}^{min\{n-3,\lfloor 1/d\rfloor -3\}} (-1)^j(1-(j+3)d)^{n-1}(j+1)\sum_{k=j+2}^{n-1}\binom{k-1}{j+1}$$

$$= \sum_{j=0}^{min\{n-2,\lfloor 1/d\rfloor -2\}} (-1)^j(1-(j+2)d)^n \left[(j+1)\binom{n}{j+2} - \binom{n-1}{j+1}\right]$$

$$- nd \sum_{j=0}^{min\{n-3,\lfloor 1/d\rfloor -3\}} (-1)^j(1-(j+3)d)^{n-1}(j+1)\binom{n-1}{j+2}$$

$$= \sum_{j=0}^{min\{n-2,\lfloor 1/d\rfloor -2\}} (-1)^j(1-(j+2)d)^n \left[((j+2)-1)\binom{n}{j+2} - \binom{n-1}{j+1}\right]$$

$$- nd \sum_{j=0}^{min\{n-3,\lfloor 1/d\rfloor -3\}} (-1)^j(1-(j+3)d)^{n-1}((j+2)-1)\binom{n-1}{j+2}$$

$$
\begin{aligned}
&= n \sum_{j=0}^{min\{n-2,\lfloor 1/d \rfloor -2\}} \binom{n-1}{j+1} (-1)^j (1-(j+2)d)^n \\
&\quad - \sum_{j=0}^{min\{n-2,\lfloor 1/d \rfloor -2\}} \binom{n}{j+2} (-1)^j (1-(j+2)d)^n \\
&\quad - \sum_{j=0}^{min\{n-2,\lfloor 1/d \rfloor -2\}} \binom{n-1}{j+1} (-1)^j (1-(j+2)d)^n \\
&\quad - n(n-1)d \sum_{j=0}^{min\{n-3,\lfloor 1/d \rfloor -3\}} \binom{n-2}{j+1} (-1)^j (1-(j+3)d)^{n-1} \\
&\quad + nd \sum_{j=0}^{min\{n-3,\lfloor 1/d \rfloor -3\}} \binom{n-1}{j+2} (-1)^j (1-(j+3)d)^{n-1} \\
&= (n-1) \sum_{j=0}^{min\{n-2,\lfloor 1/d \rfloor -2\}} \binom{n-1}{j+1} (-1)^j (1-(j+2)d)^n \\
&\quad - \sum_{j=0}^{min\{n-2,\lfloor 1/d \rfloor -2\}} \binom{n}{j+2} (-1)^j (1-(j+2)d)^n \\
&\quad - n(n-1)d \sum_{j=0}^{min\{n-3,\lfloor 1/d \rfloor -3\}} \binom{n-2}{j+1} (-1)^j (1-(j+3)d)^{n-1} \\
&\quad + nd \sum_{j=0}^{min\{n-3,\lfloor 1/d \rfloor -3\}} \binom{n-1}{j+2} (-1)^j (1-(j+3)d)^{n-1} \\
&= -(n-1) \sum_{j=0}^{min\{n-1,\lfloor 1/d \rfloor -1\}} \binom{n-1}{j} (-1)^j (1-(j+1)d)^n + (n-1)(1-d)^n \\
&\quad - \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j} (-1)^j (1-jd)^n + 1 - n(1-d)^n
\end{aligned}
$$

$$+ n(n-1)d \sum_{j=0}^{min\{n-2,\lfloor 1/d\rfloor -2\}} \binom{n-2}{j} (-1)^j (1-(j+2)d)^{n-1} - n(n-1)d(1-2d)^{n-1}$$

$$+ nd \sum_{j=1}^{min\{n-1,\lfloor 1/d\rfloor -1\}} \binom{n-1}{j} (-1)^j (1-(j+1)d)^{n-1} + n(n-1)d(1-2d)^{n-1}.$$

Therefore

$$\begin{aligned}
\mathrm{E}(L_n(d)) = {} & 1-(1-d)^n - \sum_{j=0}^{min\{n,\lfloor 1/d\rfloor\}} \binom{n}{j} (-1)^j (1-jd)^{n-1}(1-jd) \\
& + nd \sum_{j=1}^{min\{n-1,\lfloor 1/d\rfloor -1\}} \binom{n-1}{j} (-1)^j (1-(j+1)d)^{n-1} \\
& + \sum_{j=0}^{min\{n,\lfloor 1/d\rfloor\}} \binom{n}{j} (-1)^j (1-jd)^{n-1} \\
= {} & 1-(1-d)^n - nd \sum_{j=1}^{min\{n,\lfloor 1/d\rfloor\}} \binom{n-1}{j-1} (-1)^{j-1} (1-(j-1+1)d)^{n-1} \\
& + nd \sum_{j=1}^{min\{n-1,\lfloor 1/d\rfloor -1\}} \binom{n-1}{j} (-1)^j (1-(j+1)d)^{n-1} \\
= {} & 1-(1-d)^n - nd(1-d)^{n-1},
\end{aligned}$$

and the result follows immediately. ■

At the end of this subsection, we consider another sampling method. It is very similar to the previous method; the only difference is that now instead of choosing the connected component which is the first after the given point A in the clockwise direction (when the component does not cover A) we choose the *closest* connected component to A. Let $K_n^{cl}(A)$ denote the size of the closest component to the given point A. Note that for $k \geq 2$

$$
\begin{aligned}
&\Pr\{K_n^{cl}(A) = k \wedge H^c\} = \\
&= 2k(k-1)\binom{n}{k}\int_0^{d/2}\int_0^{1-2d} P(k,y,d)(1-y-2d)^{n-k}dydx \\
&\quad + 2k(k-1)\binom{n}{k}\int_{d/2}^{\frac{1-d}{2}}\int_0^{1-2x-d} P(k,y,d)(1-y-d-2x)^{n-k}dydx \\
&= k(k-1)\binom{n}{k}\int_0^{d}\int_0^{1-2d} P(k,y,d)(1-y-2d)^{n-k}dydx \\
&\quad + k(k-1)\binom{n}{k}\int_{d}^{1-d}\int_0^{1-x-d} P(k,y,d)(1-y-d-x)^{n-k}dydx
\end{aligned}
$$

and

$$
\Pr\{K_n^{cl}(A) = 1 \wedge H^c\} = 2n\int_0^{\frac{d}{2}}(1-2d)^{n-1}dx + 2n\int_{\frac{d}{2}}^{\frac{1-d}{2}}(1-2x-d)^{n-1}dx.
$$

Hence (see the proof of Theorem 3) we obtain a somehow surprising equality of probability distributions

$$
\Pr\{K_n^{cl}(A) = k\} = \Pr\{K_n(A) = k\}, \quad k = 1, 2, \ldots, n.
$$

3.2. Asymptotic results

It is not very surprising that the different sampling methods may give different distributions for the sizes of chosen components, but will they differ asymptotically, too? In this subsection we will answer this question. Let us first give asymptotic approximations for the total length of the random circle graph which follow directly from Corollary 2.

Corollary 3. *For sequences $(\mathcal{CG}_{n,d})_{n\to\infty}$ of random circle graphs*

(a) *with distance levels $d(n) = \varepsilon(n)/(n-1)$, where $\varepsilon(n) \to 0$ as $n \to \infty$ we have*

$$
\mathrm{E}(L_n(d)) = \frac{\varepsilon(n)^2}{2} + o(\varepsilon(n)^2);
$$

(b) *with distance levels* $d(n) = (c+\varepsilon(n))/(n-1)$*, where* $\varepsilon(n) \to 0$ *as* $n \to \infty$ *and* c *is a non-negative constant we have*

$$\mathrm{E}(L_n(d)) = 1 - (1 + c + \varepsilon(n))e^{-c-\varepsilon(n)}(1 + O(1/n)),$$

i.e., it tends to $1 - e^{-c}(1 + c)$ *as* $n \to \infty$*;*

(c) *with distance levels* $d(n) = \frac{\ln \omega(n)}{n-1}$*,*

$$1 - \mathrm{E}(L_n(d)) \sim \frac{\ln \omega(n)}{\omega(n)},$$

where $\omega(n) \to \infty$ *as* $n \to \infty$*.*

For sparse circle graphs, that means for $nd \to c$, the asymptotic distributions for $K_n(v)$ and $K_n'(A)$ are the same and both are negative binomial with parameters 2 and e^{-c}; however, for $K_n(A)$ (and for $K_n^{cl}(A)$), the asymptotic distribution is a mixture of two negative binomial distributions,

$$(1 + c)[k(e^{-c})^2((1 - e^{-c})^{k-1})] - c[(k - 1)(e^{-c})^2(1 - e^{-c})^{k-2}],$$

for $k = 1, 2, \ldots$ as $n \to \infty$.

For larger d, that is when $nd \to \infty$ but $n(1 - d)^n \to \infty$, or $n(1 - d)^n \to$ constant, or $n(1 - d)^n \to 0$, all random variables $K_n(v)$, $K_n(A)$, $K_n'(A)$, and $K_n^{cl}(A)$ have the same asymptotic distributions.

With distance levels $d(n) = (\log \omega(n))/n$, where $\omega(n) \to \infty$, but $\omega(n) = o(n)$ as $n \to \infty$, the normalized size $K_n(v)/\omega(n)$ converges in distribution to a random variable with density xe^{-x} for $x > 0$, i.e., to a gamma distribution with parameters 1 and 2.

With distance levels $d(n) = (\log n + c + o(1))/n$, the normalized size $K_n(v)/n$, under the condition that $K_n(v) \neq n$, converges in distribution to a random variable with density

$$\frac{e^{-2c}}{1 - e^{-e^{-c}}(1 + e^{-c})} xe^{-xe^{-c}}$$

for $x \in (0, 1)$ as $n \to \infty$.

With distance levels $d(n)$ such that $n(1-d)^n \to 0$, and under the condition that $K_n(v) \neq n$, the normalized sizes $K_n(v)/n$ converge in distribution to a random variable with the density $2x$ for $x \in (0, 1)$ as $n \to \infty$.

4. CHOOSING TWO COMPONENTS

Sampling two components leads us to results which give more information about the structure of circle graphs during their evolution when d increases.

4.1. Exact results

For two given vertices v_1 and v_2, let $U(v_1, v_2)$ denote the event that these two vertices are not in the same component in a random circle graph $\mathcal{CG}_{n,d}$. Let Q be the probability of this event. Then $Q = 1 - P_2$, where P_2 is given by Lemma 2. Let $K_n^*(v_1)$ and $K_n^*(v_2)$ be the sizes of the two components containing v_1 and v_2, respectively, under assumption that $U(v_1, v_2)$ holds, i.e.,

$$\Pr\{K_n^*(v_1) = k_1, K_n^*(v_2) = k_2\} = \Pr\{K_n(v_1) = k_1, K_n(v_2) = k_2 \mid U(v_1, v_2)\}.$$

In the next theorem we give the joint probability distribution of these random variables.

Theorem 4. *Assume that vertices v_1 and v_2 are not in the same component, and let $K_n^*(v_1)$ and $K_n^*(v_2)$ denote the sizes of the components containing v_1 and v_2, respectively. Then*

$$\Pr\{K_n^*(v_1) = k_1,\ K_n^*(v_2) = k_2\} =$$

$$= \frac{k_1 k_2 (n-k_1-k_2-1)}{(n-1)Q} \sum_{j=0}^{min\{k_1+k_2-2, \lfloor 1/d \rfloor - 4\}} \binom{k_1+k_2-2}{j} (-1)^j (1-(j+4)d)^{n-1}$$

$$+ \frac{2k_1k_2}{(n-1)Q} \sum_{j=0}^{min\{k_1+k_2-2, \lfloor 1/d \rfloor - 3\}} \binom{k_1+k_2-2}{j} (-1)^j (1-(j+3)d)^{n-1}$$

for $k_1 = 1, \dots, n-2,\ k_2 = 1, 2, \dots, n-1-k_1,$ and

$$\Pr\{K_n^*(v_1) = k,\ K_n^*(v_2) = n-k\} =$$

$$= \frac{k(n-k)}{(n-1)Q} \sum_{j=0}^{min\{n-2, \lfloor 1/d \rfloor - 2\}} \binom{n-2}{j} (-1)^j (1-(j+2)d)^{n-1}$$

for $k = 1, \ldots, n-1$, *where*

$$Q = -1 + \frac{2}{n-1} \sum_{j=2}^{min\{n,\lfloor 1/d \rfloor+1\}} \binom{n}{j} (-1)^j (1-(j-1)d)^{n-1} + \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j} (-1)^j (1-jd)^{n-1}.$$

Proof. Let us choose a given set of k_1+k_2 vertices which contains v_1 and v_2 and which is supposed to form the two desired components. It can be done in $\binom{n-2}{k_1+k_2-2}$ ways. Assume that the sum of lengths of the two components is equal to y and choose the vertex w which will be one of the border points of the component containing v_2 — it can be done in k_1+k_2-2 ways. Now, let k_1+k_2-3 vertices, i.e., all except v_1, v_2 and w, be placed at random on the interval of length y. Assuming that two additional vertices are placed at the border of the interval, we obtain by Lemma 1 that the probability that these k_1+k_2-1 vertices form a connected subgraph of length y is given by

$$P(k_1+k_2-1, y, d) = \sum_{j=0}^{min\{k_1+k_2-2,\lfloor y/d \rfloor\}} \binom{k_1+k_2-2}{j} (-1)^j (y-jd)^{k_1+k_2-3}.$$

We will be assuming that the (k_1-1)-th vertex from the left (not counting the border point), say a vertex u, is a right border point of the component containing v_1. Next, we have to find the places for three vertices v_1, v_2 and w using two empty places (at the borders) and adding one more vertex at the place occupied already by u. We are going to decompose the above connected subgraph into two components exactly at this point and the added vertex is the left border point of the second component. We have one possibility to put the vertex v_1 at the left border point of the interval and there are k_1-1 ways to put it at one of the next occupied, first k_1-1 places. In the latter case the vertex from this place and all vertices to its left side will be shifted one place to the left. The vertex v_2 can be placed together with the vertex u, i.e., v_2 is the left border point of the second connected component and the vertex w is placed at the right border of the second connected component, or we put the vertex v_2 at one of the next occupied places (k_2-2 possibilities) shifting the vertex from this place and all vertices to its right side, one place to the right, or, finally, v_2 can be placed at the right border point of the interval (in two last cases w is placed at the left border point of the second component). Next we add the interval of length u after the first component (between the respective "border" vertices of the two components). Note that if $u \geq 2d$, then we may have other components between our two connected components, while for $d \leq u \leq 2d$ it will be not possible. Integrating over all possible values of y and u gives for $k_1 = 1, \ldots, n-2$, $k_2 = 1, 2, \ldots, n-1-k_1$,

$$\Pr\{K_n^*(v_1) = k_1, K_n^*(v_2) = k_2\} =$$

$$= \frac{1}{Q}\binom{n-2}{k_1+k_2-2} k_1 k_2 (k_1+k_2-2)$$

$$\times \left[2\int_0^{1-3d}\int_d^{min\{2d,1-y-2d\}} P(k_1+k_2-1, y, d)(1-y-u-2d)^{n-k_1-k_2} du\, dy \right.$$

$$\left. + \int_0^{1-4d}\int_{2d}^{1-y-2d} P(k_1+k_2-1, y, d)(1-y-4d)^{n-k_1-k_2} du\, dy \right].$$

Standard calculations imply the result. In the same manner one can get the distribution for remaining cases to complete the proof. ■

The symmetry of the above joint distribution implies immediately that the marginal distributions are the same. To find them, we can either directly use very similar reasonings to those used in the proof of Theorem 4, or we can sum up the joint probability distribution over all k_1 (or k_2).

Corollary 4. *The marginal distribution of $K_n^*(v_1)$ and of $K_n^*(v_2)$ is*

$$\Pr\{K_n^*(v_1) = k\} = \Pr\{K_n^*(v_2) = k\}$$

$$= \frac{k(n-k)}{(n-1)Q} \sum_{j=0}^{min\{k-1,\lfloor 1/d \rfloor -2\}} \binom{k-1}{j} (-1)^j (1-(j+2)d)^{n-1}$$

for $k = 1, 2, ..., n-1$, where

$$Q = -1 + \frac{2}{n-1} \sum_{j=2}^{min\{n,\lfloor 1/d \rfloor +1\}} \binom{n}{j} (-1)^j (1-(j-1)d)^{n-1}$$

$$+ \sum_{j=0}^{min\{n,\lfloor 1/d \rfloor\}} \binom{n}{j} (-1)^j (1-jd)^{n-1}.$$

Next we will consider the joint distribution of the sizes of the component containing a given vertex and of the component which is the next one in the clockwise direction (assuming that the second variable takes the value 0 if the component does not exist).

Theorem 5. *Let v be a given vertex. The joint probability of the size $K_n(v)$ of the component containing this vertex and of the size $RK_n(v)$ of the component which the next one in the clockwise direction is given by*

$$\Pr\{K_n(v) = k_1, RK_n(v) = k_2\}$$
$$= k_1 \sum_{j=0}^{min\{k_1+k_2-2,\lfloor 1/d \rfloor -3\}} \binom{k_1+k_2-2}{j} (-1)^j (1-(j+3)d)^{n-1}$$

for $k_1, k_2 = 1, 2, \ldots, n-1, k_1 + k_2 < n$, and

$$\Pr\{K_n(v) = k,\ RK_n(v) = n-k\}$$
$$= k \sum_{j=0}^{min\{n-2,\lfloor 1/d \rfloor -2\}} \binom{n-2}{j} (-1)^j (1-(j+2)d)^{n-1}$$

for $k = 1, 2, \ldots, n-1$, and

$$\Pr\{K_n(v) = n,\ RK_n(v) = 0\} = \Pr\{\mathcal{CS}_n\}.$$

Proof. We can prove the above formulas for the joint probability distribution using the same approach as in the proof of Theorem 4. The only differences are that the interval of length u (with the same meaning as in the proof of Theorem 4) has to be a gap, so none of the vertices can be placed into it. Moreover we have exactly one given vertex and the second component does not contain any special vertex. Hence

$$\Pr\{K_n(v) = k_1,\ RK_n(v) = k_2\}$$
$$= \binom{n-1}{k_1+k_2-1}(k_1+k_2-1)(k_1+k_2-2)k_1$$
$$\times \int_0^{1-3d}\int_d^{1-y-2d} P(k_1+k_2-1, y, d)(1-y-u-2d)^{n-k_1-k_2}du\,dy$$

for $k_1 + k_2 < n$. The straightforward calculations lead us directly to the first formula in the theorem. Similarly we can obtain the other statements of the theorem. ∎

For the first random variable of the random vector $(K_n(v), RK_n(v))$, the marginal distribution is obviously the same as the distribution given in Theorem 2. We can obtain the marginal distribution of the second random variable by summing up over k_1 the joint probability distribution given in Theorem 5.

Corollary 5. *The marginal distribution of the random variable $RK_n(v)$ as introduced in Theorem 5 is given by*

$$\Pr\{RK_n(v) = k\} = \sum_{j=0}^{min\{k-1,\lfloor 1/d\rfloor-1\}} \binom{k-1}{j}(-1)^j(1-(j+1)d)^{n-1}$$
$$- \sum_{j=0}^{min\{n-1,\lfloor 1/d\rfloor-1\}} \binom{n-1}{j}(-1)^j(1-(j+1)d)^{n-1}$$

for $k = 1, 2, \ldots, n-1$, and

$$\Pr\{RK_n(v) = 0\} = \Pr\{\mathcal{CS}_n\}.$$

4.2. Asymptotic results

The evolution of random interval graphs is discussed in details in [2]. In a similar way one can describe the evolution for random circle graphs.

In this paper, we restrict ourselves to the direct consequences of our exact results - the asymptotic distributions of the sizes of connected components.

For small distances d, the coordinates of the random vector $(K^*(v_1), K^*(v_2))$ (and also of $(K_n(v), RK_n(v)))$ are asymptotically independent. It is not surprising - we could expect that since all components are small, the size of the first component has no effect on the size of the second component. By increasing d we increase the dependence, too.

A. Sparse circle graphs: $nd \to c$

In this case with distance levels $d(n) = (c+o(1))/(n-1)$, by using Theorem 4 we can easily see that the probability distribution of $(K^*(v_1), K^*(v_2))$ tends to

$$k_1 k_2 e^{-4c}(1 - e^{-c})^{k_1+k_2-2},$$

$k_1, k_2 = 1, 2, \ldots$, as $n \to \infty$, which implies that $K_n^*(v_1)$ and $K_n^*(v_2)$ are asymptotically independent, both with the same negative binomial distribution with parameters 2 and e^{-c}.

It follows from Theorem 5 that the probability distribution of $(K_n(v), RK_n(v))$ tends to

$$k_1 e^{-3c}\,(1 - e^{-c})^{k_1+k_2-2}$$

for $k_1, k_2 = 1, 2, \ldots$, as $n \to \infty$, which means that $K_n(v)$ and $RK_n(v)$ are also asymptotically independent but they have two different asymptotic distributions, the negative binomial with parameters 2 and e^{-c} and the geometrical with parameter e^{-c}, respectively.

B. The disappearance of finite components: $nd \to \infty$ but $n(1-d)^n \to \infty$

In this case the random circle graph still has components with sizes of order smaller than $\Theta(n)$ but, on the other hand, the components with finite sizes begin to vanish. Theorem 4 implies the following result.

Theorem 6. *For sequences $(\mathcal{CG}_{n,d})_{n\to\infty}$ of random circle graphs with distance levels $d(n) = (\log \omega(n))/n$, the normalized sizes $(K_n^*(v_1)/\omega(n), K_n^*(v_2)/\omega(n))$ converge in distribution to a random variable with density $x_1\,x_2\,e^{-(x_1+x_2)}$ for $x_1 > 0$, $x_2 > 0$, that is, to the product of two independent gamma distribution with parameters* 1 *and* 2, *where $\omega(n) \to \infty$, but $\omega(n) = o(n)$ as $n \to \infty$.*

Similarly Theorem 5 implies the next theorem.

Theorem 7. *For sequences $(\mathcal{CG}_{n,d})_{n\to\infty}$ of random circle graphs with distance levels $d(n) = (\log\omega(n))/n$, the normalized sizes $(K_n(v)/\omega(n), RK_n(v)/\omega(n))$ converge in distribution to a random variable with density $x_1\, e^{-(x_1+x_2)}$ for $x_1 > 0$, $x_2 > 0$, where $\omega(n) \to \infty$, but $\omega(n) = o(n)$ as $n \to \infty$.*

C. The connectivity threshold: $n(1-d)^n \to \text{constant}$

It is the most interesting case and more complicated than the previous ones, since now the random circle graph $\mathcal{CG}_{n,d}$ is connected with probability bounded away from both 0 and 1, so its properties are a mixture of properties described in the previous case and the connectivity case.

Theorem 8. *For sequences $(\mathcal{CG}_{n,d})_{n\to\infty}$ of random circle graphs with distance levels $d(n) = (\log n + c + o(1))/n$, the normalized vectors $(K_n^*(v_1)/n, K_n^*(v_2)/n)$ of component sizes converge in distribution to a random variable with joint density function given by*

$$\frac{e^{-3c}}{1-2e^c+e^{-e^{-c}}(1+2e^c)}\left[e^{-c}x_1x_2(1-x_1-x_2)e^{-(x_1+x_2)e^{-c}}+2x_1x_2e^{-(x_1+x_2)e^{-c}}\right]$$

for $x_1, x_2 \in (0,1)$, $x_1 + x_2 < 1$, and

$$\frac{e^{-2c}e^{-e^{-c}}}{1-2e^c+e^{-e^{-c}}(1+2e^c)}\,x(1-x),$$

for $x_1 = 1 - x_2 = x$ as $n \to \infty$.

This shows that under the condition that v_1 and v_2 are not in the same component, with positive probability the random circle graph $\mathcal{CG}_{n,d}$ has just two components with sizes of order $\Theta(n)$. In this case we do not have asymptotic independence, and both normalized variables converge in distribution to a random variable with density function given by

$$\frac{e^{-2c}}{1-2e^c+e^{-e^{-c}}(1+2e^c)}\,x(1-x)e^{-xe^{-c}},$$

for $x \in (0,1)$ as $n \to \infty$.

Theorem 9. *For sequences $(\mathcal{CG}_{n,d})_{n\to\infty}$ of random circle graphs with distance levels $d(n) = (\log n + c + o(1))/n$, the normalized vectors $(K_n(v)/n, RK_n(v)/n)$ of component sizes, under the assumption that $K_n(v) \neq n$, converge in distribution to a random variable with joint density function given by*

$$\frac{e^{-3c}}{1 - e^{-e^{-c}}(1 + e^{-c})}\, x_1 e^{-(x_1+x_2)e^{-c}}$$

for $0 \le x_1 + x_2 < 1$, and with density

$$\frac{e^{-2c} e^{-e^{-c}}}{1 - e^{-e^{-c}}(1 + e^{-c})}\, x$$

for $x_1 = 1 - x_2 = x$, as $n \to \infty$.

Moreover, under the assumption that $K_n(v) \neq n$, with positive probability the random circle graph $\mathcal{CG}_{n,d}$ has just two components with sizes of order $\Theta(n)$. The asymptotic distribution of $RK_n(v)/n$, under the condition that the random circle graph is not connected, is given by the following density function

$$\frac{e^{-c}}{1 - e^{-e^{-c}}(1 + e^{-c})}\left(e^{-xe^{-c}} - e^{-e^{-c}}\right)$$

for $0 < x < 1$ as $n \to \infty$, which is different from the conditional distribution of $K^*(v_2)$.

D. The connected random circle graph: $n(1-d)^n \to 0$

In this case the random circle graph is almost surely connected. Let us, however, study the structure of random circle graphs under the condition that it is not connected.

Theorem 10. *For sequences $(\mathcal{CG}_{n,d})_{n\to\infty}$ of random circle graphs with distance levels $d(n)$ such that $n(1-d)^n \to 0$, the normalized sizes $(K_n^*(v_1)/n, K_n^*(v_2)/n)$ converge in distribution to a random variable with density $6x(1-x)$ for $x \in (0,1)$, where $x_1 = 1 - x_2 = x$, and to 0 where $x_1 + x_2 \neq 1$, as $n \to \infty$.*

This means that, under the condition that v_1 and v_2 belong to different components, asymptotically almost surely the random circle graph $\mathcal{CG}_{n,d}$ has exactly two components with sizes of order $\Theta(n)$.

Theorem 11. *For sequences $(\mathcal{CG}_{n,d})_{n\to\infty}$ of random circle graphs with distance levels $d(n)$ such that $n(1-d)^n \to 0$, and under the assumption that $K_n(v) \neq n$, the normalized sizes vectors $(K_n(v)/n, RK_n(v)/n)$ converge in distribution to a random variable with density $2x$ for $x_1 = 1-x_2 = x \in (0,1)$, as $n \to \infty$.*

Again for this sampling method under the condition that $K_n(v) \neq n$, the random circle graph cannot have more than exactly two components with sizes of order $\Theta(n)$. The marginal distribution of normalized random variable $RK_n(v)/n$, under the above conditions, tends to a distribution with density

$$2(1-x)$$

for $0 < x < 1$ as $n \to \infty$.

We know that when $n(1-d)^n \to 0$ (see [2]), the probability of having more than one components tends to zero, but the above results show that the probability of having three or more components (or two components where at least one of them has not the size of order $\Theta(n)$) tends to zero faster than the probability of having two components with sizes of order $\Theta(n)$.

Acknowledgments

This work was carried out during my visit to the Department of Discrete Mathematics at the Adam Mickiewicz University, Poznań, Poland. I would like to sincerely thank everyone in the department.

References

[1] L. Fatto and A.G. Konheim, *The random division of an interval and the random covering of a circle*, SIAM Review **4** (1962) 211–222.

[2] E. Godehardt and J. Jaworski, *On the connectivity of a random interval graphs*, Random Structures and Algorithms **9** (1996) 137–161.

[3] L. Holst and J. Husler, *On the random coverage of the circle*, J. Appl. Prob. **21** (1984) 558–566.

[4] T. Huillet, *Random covering of the circle: the size of the connected components*, Adv. in Appl. Probab. **35** (2003) 563–582.

[5] H. Maehara, *On the intersection graph of random arcs on a circle*, Random Graphs' 87 (1990) 159–173.

[6] M. Penrose, Random Geometric Graphs (Oxford Studies in Probability, 2003).

[7] A.F. Siegel, *Random arcs on the circle*, J. Appl. Prob. **15** (1978) 774–789.

[8] H. Solomon, Geometric Probability (Society for Industrial and Applied Mathematics, Philadelphia, 1976).

[9] F.W. Steutel, *Random division of an interval*, Statistica Neerlandica **21** (1967) 231–244.

[10] W.L. Stevens, *Solution to a geometrical problem in probability*, Ann. Eugenics **9** (1939) 315–320.

Received 18 February 2008
Revised 4 April 2008
Accepted 7 April 2008