

## Random rectangular graphs

Ernesto Estrada and Matthew Sheerin

*Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, United Kingdom*

(Received 22 January 2015; revised manuscript received 21 March 2015; published 21 April 2015)

A generalization of the random geometric graph (RGG) model is proposed by considering a set of points uniformly and independently distributed on a rectangle of unit area instead of on a unit square  $[0,1]^2$ . The topological properties of the *random rectangular graphs* (RRGs) generated by this model are then studied as a function of the rectangle sides lengths  $a$  and  $b = 1/a$ , and the radius  $r$  used to connect the nodes. When  $a = 1$  we recover the RGG, and when  $a \rightarrow \infty$  the very elongated rectangle generated resembles a one-dimensional RGG. We obtain here analytical expressions for the average degree, degree distribution, connectivity, average path length, and clustering coefficient for RRG. These results provide evidence that show that most of these properties depend on the connection radius and the side length of the rectangle, usually in a monotonic way. The clustering coefficient, however, increases when the square is transformed into a slightly elongated rectangle, and after this maximum it decays with the increase of the elongation of the rectangle. We support all our findings by computational simulations that show the goodness of the theoretical models proposed for RRGs.

DOI: [10.1103/PhysRevE.91.042805](https://doi.org/10.1103/PhysRevE.91.042805)

PACS number(s): 89.75.-k, 02.10.Ox

### I. INTRODUCTION

The use of graphs for representing physical systems is becoming ubiquitous in many areas of theoretical and applied physics [1]. We can mention the use of graphs in statistical mechanics and condensed matter physics, for solving Feynman integrals as well as in the study of quantum phenomena [1,2]. More recently, the use of graphs has been very broadened by their application in the analysis of complex systems [3–5]. In this case, those graphs receive the name of complex networks, due to the fact that they represent the skeleton of complex interconnected systems. In this case, networks are used to study a variety of physical scenarios, ranging from social and infrastructural, to biological and ecological ones. Here, we will use the terms graphs and networks interchangeably. When graphs are used to represent real-world physical systems it is necessary to have at our disposal some null model that allows us to evaluate which properties of the system have arisen from their connectivity pattern. In this sense, the common election is the use of random graphs. These are graphs with the same number of nodes and edges as the one under study, but in which the connection between the nodes is made randomly and independently [6,7]. There are several of these random models of great usability in current network theory, such as the Erdős-Rényi [8], the Barabási-Albert [9], or the Watts-Strogatz [10] model to mention just three.

In many real-world scenarios the networks emerge under certain geometrical constraints. This is the case of the so-called spatial networks [11], which include infrastructural networks such as road networks, airport transportation networks, etc. [11], and certain biological networks such as brain networks or the networks representing the proximity of cells in a biological tissue (see Ref. [3]). The list also includes the networks of patches and corridors in a landscape [12], the networks of galleries in animal nests [13,14], and the networks of fractures in rocks [15], among others. The classical election of a random graph used to represent these systems are the so-called random geometric graphs [16,17]. Here the term random geometric graph (RGG) is reserved for the case in which the nodes of the graph are distributed randomly and

independently in a unit square and two nodes are connected if they are inside a disk of a given radius centered at one of the nodes. Other graphs in which the edges are constructed by using different geometric rules will be named here generically as random proximity graphs.

RRGs have found important applications in the area of wireless communication devices [18–20], such as mobile phones, wireless computing systems, wireless sensor networks, etc. This was indeed the first application in mind when Gilbert proposed the very first RGG model [21]. RRGs have also found applications in areas such as modeling of epidemic spreading in spatial populations, which may include cases such as the spreading of worms in a computer network, viruses in a human population, or rumors in a social network [22–26]. RRGs have been used to describe how cities have been evolving under the geometric constraints imposed by their geographic locations [27]. For a wider perspective on the applications of spatial graphs the reader is referred to Ref. [11].

In all the previously mentioned real-world scenarios, the shape of the location in which the nodes of the graph are distributed may play a fundamental role in the topological and dynamical properties of the resulting graphs. That is, it is intuitive to think that the connectivity, distance, clustering, and other fundamental topological properties of the graphs are affected if we, for instance, elongate the unit square in which the points are distributed. Here, we develop a model that generalizes the RGG by allowing the embedding of the nodes in a unit rectangle instead of a unit square. Our main goal is to investigate how the elongation of a unit square influences the topological properties of the graphs generated by the model. These generalized graphs will be named here the *random rectangular graphs* (RRGs). In this work we concentrate on the influence of the length of the rectangle on the topological properties of the graphs emerging on them, such as their average degree, connectivity, degree distribution, average path length, and clustering coefficient. In particular, we find analytical expressions and bounds for all of them and provide computational evidence of the goodness of these approaches for relatively large RRGs.

**II. DEFINITION OF THE MODEL**

The RGG is defined by distributing uniformly and independently  $n$  points in the unit  $d$ -dimensional cube  $[0, 1]^d$  [16]. Then, two points are connected by an edge if their (Euclidean) distance is at most  $r$ , which is a given fixed number known as the connection radius.

Let us now define a unit hyperrectangle as the Cartesian product  $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d]$ , where  $a_i, b_i \in \mathbb{R}, a_i \leq b_i$  and  $1 \leq i \leq d$ . Hereafter, we will restrict ourselves to the two-dimensional (2D) case, which corresponds to a rectangle of unit area, which we will call the unit rectangle. Now, the RRG is defined by distributing uniformly and independently  $n$  points in the unit rectangle  $[a, b]$  and then connecting two points by an edge if their (Euclidean) distance is at most  $r$ . It is evident that the only change we have introduced here is to consider a rectangle of unit area instead of the analogous square. The rest of the construction process remains the same as for the RGG. This means that  $\text{RRG} \rightarrow \text{RGG}$  as  $(a/b) \rightarrow 1$ . In this sense we can say that the RRG is a generalization of the RGG. In Fig. 1 we illustrate an RGG and an RRG constructed with the same number of nodes and edges.

An interesting question is what happens at the other extreme, when  $a \rightarrow \infty$ . In this case we have that  $b \rightarrow 0$ , which means that the  $n$  points are uniformly and independently distributed on the straight line. Let us now consider a disk of radius  $r > 0$  centered at each of these points and let us

connect every point to the other points that lie inside its disk. Thus, the resulting graph resembles a one-dimensional RGG, that is a graph created by placing the  $n$  points uniformly and independently on the interval  $[0, 1]$  and then connecting pairs of nodes if they are at a (Euclidean) distance smaller than or equal to a certain connection radius  $r$  (see, for instance, Refs. [28–30]).

**III. TOPOLOGICAL PROPERTIES OF RRGs**

**A. Average degree**

We start the study of the topological properties of RRGs by considering an analytical expression for the average degree  $\bar{k}$ . We remind that the degree of a node is the number of edges connected to it. The average degree is a property not only interesting by itself but as we will see in the next sections of this paper it is fundamental to understand other topological parameters of RRGs.

To start with, let us consider that for a given node, there are  $n - 1$  nodes distributed in the rest of the rectangle. Define  $A_p$  to be the area within the radius  $r$  of a point  $p$  that lies within the rectangle. Since the nodes are uniformly and independently distributed, the expected node degree of a node  $v_i$  is  $\mathbf{E}(k_i) = (n - 1)A_i/(ab)$ , where  $A_i$  is taken for the point where node  $v_i$  is located. This is because dividing the nodes between the area within distance  $r$  and the rest of the rectangle gives rise to the Binomial distribution  $B[n - 1, A_i/(ab)]$ . Averaging this over

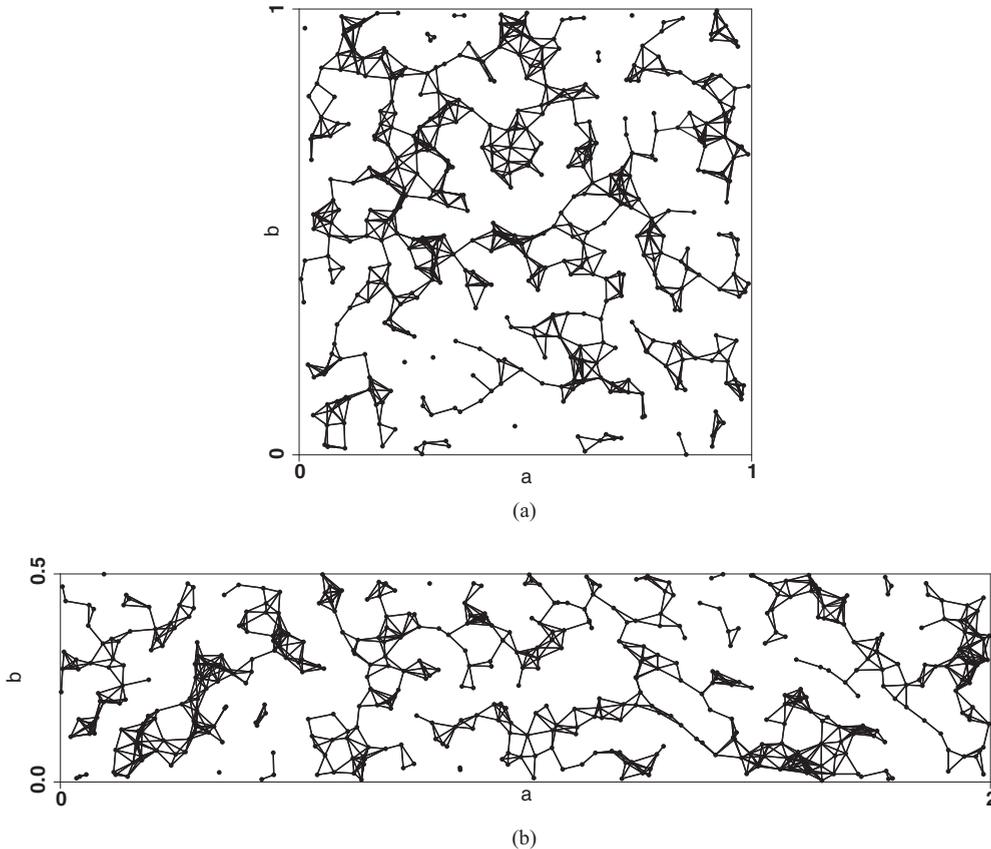


FIG. 1. Illustration of two random rectangular graphs with  $a = 1$  (top), which corresponds to a random geometric graph on a unit square and with  $a = 2$  ( $b = 0.5$ ) (bottom). Both graphs are built with 500 nodes and 1750 edges.

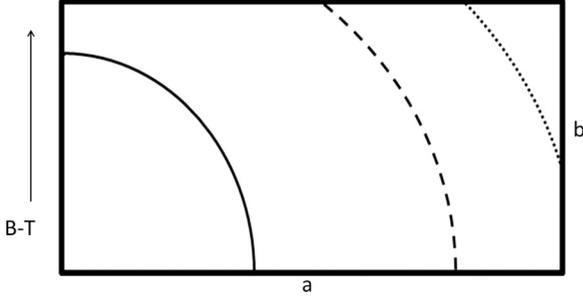


FIG. 2. Illustration of three different quarter circles in the rectangle corresponding to  $0 \leq r \leq b$  (solid line),  $b \leq r \leq a$  (broken line), and  $a \leq r \leq \sqrt{a^2 + b^2}$  (dotted line). The direction used for displacing the circles is represented as bottom-top (B-T) with an arrow in the graphic.

all possible node locations (i.e., the points in the rectangle) gives

$$\mathbf{E}\bar{k} = \frac{\int_p \{(n-1)A_p/(ab)\}}{ab} = \frac{(n-1) \int_p A_p}{(ab)^2}. \quad (1)$$

Let  $f(a,b,r)$  be the area within radius  $r$  of a point which lies in the rectangle, integrated over all points, i.e.,  $f(a,b,r) = \int_p A_p$ . Based on preliminary computational results (not shown) obtained for the average degree we consider here the following three regions:  $0 \leq r \leq b$ ,  $b \leq r \leq a$  and  $a \leq r \leq \sqrt{a^2 + b^2}$ , recalling that  $a \geq b$ . We call these cases 1, 2, and 3, respectively. Thus, the function  $f(a,b,r)$  takes different forms  $f_i$  for each case  $i$ . This means that we can write

$$\mathbf{E}\bar{k} = \frac{(n-1)f_i}{(ab)^2}, \quad (2)$$

with

$$f_i = \begin{cases} f_1 & 0 \leq r \leq b, \\ f_2 & b \leq r \leq a, \\ f_3 & a \leq r \leq \sqrt{a^2 + b^2}, \end{cases} \quad (3)$$

and our task is now to find the analytical expressions for  $f_i$ .

We consider the rectangle in Fig. 2, which shows three quarter circles of different radii (each corresponding to one of the three cases) as they intersect the interior of the rectangle. We consider only quarter circles instead of circles, then quadruple the result at the end. For each of these quarter circles, we divide them into vertical rectangular strips of width  $\Delta x$ , which will approximate the areas of the intersection between the quarter circles and the full rectangle; this will become exact in the limit as  $\Delta x \rightarrow 0$ . We now consider several possibilities for these strips.

First, the strips may approximate an area that is not rectangular, which occurs when the height of the strip is smaller than the height  $b$  of the rectangle. For a strip of distance  $x$  from the left of the rectangle, this corresponds to  $0 \leq x \leq r$  for the smaller quarter circle (case 1),  $\sqrt{r^2 - b^2} \leq x \leq r$  for the medium quarter circle (case 2), and  $\sqrt{r^2 - b^2} \leq x \leq a$  for the largest quarter circle (case 3). Setting  $p = \min(a,r)$ ,  $q = \min(b,r)$ , we have  $\sqrt{r^2 - q^2} \leq x \leq p$ .

Since we need to integrate these areas over all possible quarter circles, given a fixed radius, we wish to know how we can translate the quarter circle in the figure and preserve a particular strip. That is, for a particular strip of distance  $x$  from the left of the rectangle, we may find a corresponding strip on the other quarter circles of the same radius. Since we have a rectangular strip of width  $\Delta x$ , height  $\sqrt{r^2 - x^2}$ , and distance  $x$  from the center of the (quarter) circle, we may find this strip in any of the  $(a - x - \Delta x)$  positions horizontally, and  $(b - \sqrt{r^2 - x^2})$  vertically. Thus, we can use integration to find the total area of all these strips by multiplying  $(a - x - \Delta x)(b - \sqrt{r^2 - x^2})$  by the area of the strip  $\sqrt{r^2 - x^2} \Delta x$  and taking the limit to obtain

$$\begin{aligned} I_1 &= \int_{\sqrt{r^2 - q^2}}^p (a - x)(b - \sqrt{r^2 - x^2})\sqrt{r^2 - x^2} dx \\ &= \int_{\sqrt{r^2 - q^2}}^p (a - x)[b\sqrt{r^2 - x^2} - (r^2 - x^2)] dx. \end{aligned} \quad (4)$$

Second, we note that if these strips are translated far enough in the bottom-top (B-T) direction, they become truncated by the top of the rectangle. For a particular truncated strip, we may still find a corresponding strip in any of the  $(a - x - \Delta x)$  positions horizontally, and the truncated height  $t$  of a strip may be any value between 0 and the full height of the strip. Thus, integrating gives

$$\begin{aligned} I_2 &= \int_{\sqrt{r^2 - q^2}}^p (a - x) \int_0^{\sqrt{r^2 - x^2}} t dt dx \\ &= \int_{\sqrt{r^2 - q^2}}^p \frac{1}{2}(a - x)(r^2 - x^2) dx. \end{aligned} \quad (5)$$

Alternatively, we may have  $\sqrt{r^2 - b^2} > b$ , in which case the rectangular strip is exact and of height  $b$ . In this case, the only contribution is from the truncated strips. We note that this applies for  $0 \leq x \leq \sqrt{r^2 - q^2}$  by a similar argument as before, and we integrate as follows:

$$\begin{aligned} I_3 &= \int_0^{\sqrt{r^2 - q^2}} (a - x) \int_0^b t dt dx \\ &= \int_0^{\sqrt{r^2 - q^2}} \frac{1}{2}(a - x)b^2 dx. \end{aligned} \quad (6)$$

Thus, we have the expression for  $f$  as four times the sum of the above integrals,

$$\begin{aligned} f &= 4(I_1 + I_2 + I_3) \\ &= \int_0^{\sqrt{r^2 - q^2}} 2(a - x)b^2 dx \\ &\quad + \int_{\sqrt{r^2 - q^2}}^p (a - x)[4b\sqrt{r^2 - x^2} - 2(r^2 - x^2)] dx, \end{aligned} \quad (7)$$

which can be written as

$$f = \begin{cases} 0 \leq r \leq b & \pi r^2 ab - \frac{4}{3}(a+b)r^3 + \frac{1}{2}r^4, \\ b \leq r \leq a & -\frac{4}{3}ar^3 - r^2b^2 + \frac{1}{6}b^4 + a\left(\frac{4}{3}r^2 + \frac{2}{3}b^2\right)\sqrt{r^2 - b^2} \\ & + 2r^2ab \arcsin\left(\frac{b}{r}\right), \\ a \leq r \leq \sqrt{a^2 + b^2} & -r^2(a^2 + b^2) + \frac{1}{6}(a^4 + b^4) - \frac{1}{2}r^4 \\ & + b\left(\frac{4}{3}r^2 + \frac{2}{3}a^2\right)\sqrt{r^2 - a^2} + a\left(\frac{4}{3}r^2 + \frac{2}{3}b^2\right)\sqrt{r^2 - b^2} \\ & - 2abr^2\left[\arccos\left(\frac{b}{r}\right) - \arcsin\left(\frac{a}{r}\right)\right]. \end{cases} \quad (8)$$

Now we evaluate computationally how good the expression for the average degree of the RRGs is. In Fig. 3(a) we plot the values of the average degree observed for RRGs with three different values of the rectangle side length. These observed values (represented by solid squares, circles, and triangles) are the average of 100 random realizations of RRGs with  $n = 1500$  nodes. The solid lines represent the expected values according to Eq. (8). The Pearson correlation coefficients for the linear regression between the observed and expected values are larger than 0.9999 in the three cases. We enlarge the region of small radii for the case  $a = 30$  [see Fig. 3(b)] where it can be seen that it is a perfect fit also for this region with Pearson correlation coefficient as good as for the general case.

### B. Connectivity

In this section we are interested in determining the connection radius for specific values of  $n$ , i.e.,  $r(n)$ , which guarantees that the RRG  $G[n, r(n)]$  is asymptotically connected with probability one. For the RGG, Penrose [31] proved that if  $M_n$  is the maximum length of an edge in the graph, then the probability that  $n\pi M_n^2 - \log n \leq \alpha$  for a given  $\alpha \in \mathbb{R}$  is

$$\lim_{n \rightarrow \infty} P[n\pi M_n^2 - \log n \leq \alpha] = \exp[-\exp(-\alpha)]. \quad (9)$$

In 2D  $M_n = r$ , such that we have for the RGG

$$\lim_{n \rightarrow \infty} P[n\pi r^2 - \log n \leq \alpha] = \exp[-\exp(-\alpha)]. \quad (10)$$

This means that for  $\alpha \rightarrow +\infty$  the RGG is almost surely connected when  $n \rightarrow \infty$ , and almost surely disconnected when  $\alpha \rightarrow -\infty$ .

The term  $n\pi r^2$  is just the average degree  $\bar{k}$  in RGG (when no boundary effects are considered). Thus, we can write

$$\lim_{n \rightarrow \infty} P[\bar{k} - \log n \leq \alpha] = \exp[-\exp(-\alpha)]. \quad (11)$$

In the previous section we have obtained an analytic expression for  $\bar{k}$  in the RRG. Thus, we can replace this in Eq. (11) and obtain an analogous expression for the RRG. That is, for a RRG we have that

$$\lim_{n \rightarrow \infty} P\left[\frac{(n-1)f_i}{(ab)^2} - \log n \leq \alpha\right] = \exp[-\exp(-\alpha)], \quad (12)$$

where  $f_i$  is given by Eq. (8).

Because the parameter  $\alpha$  is unknown and it depends on the specific RRG considered, we can obtain a lower bound for

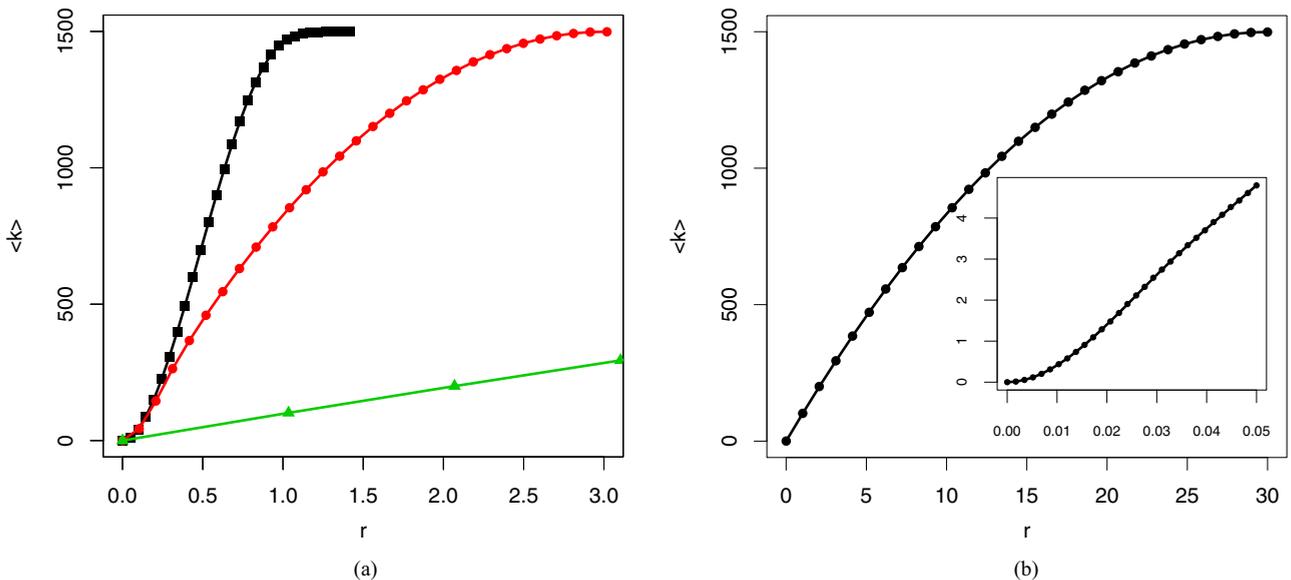


FIG. 3. (Color online) (a) Illustration of the fit between the observed [black squares ( $a = 1$ ), red circles ( $a = 3$ ), and green triangles ( $a = 30$ )] and expected (solid line) values of the average degree for RRGs with different side lengths of the rectangle. (b) Wider range of radii for  $a = 30$  (zooming for the small radii in the inset).

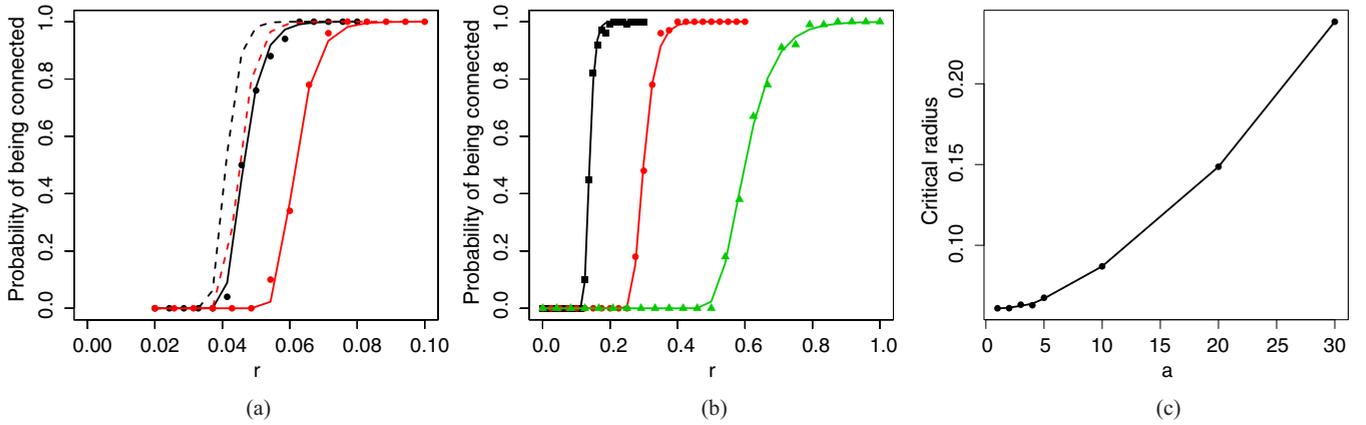


FIG. 4. (Color online) Dependence of the probability of being connected with the connection radius in RRGs. (a) Illustration of the goodness of the bound Eq. (13) for two RRGs with  $a = 1$  (black curves) and  $a = 5$  (red curves). The observed values for RRGs with 1500 nodes averaged over 100 random realizations are illustrated as solid circles and the values obtained by the theoretical bound are given by broken lines. (b) Observed values of the probability of being connected for RRGs with 1500 nodes averaged over 100 random realizations and three different rectangle side lengths  $a = 1$  (black squares),  $a = 5$  (red circles), and  $a = 10$  (green triangles). (c) Variation of the radius for which the RRG is connected (critical radius) with the rectangle side length for RRGs with 1500 nodes and 11 250 edges.

$\exp[-\exp(-\alpha)]$  using Eq. (12). That is,

$$\exp \left\{ -\exp \left[ -\left( \frac{(n-1)f_i}{(ab)^2} - \log n \right) \right] \right\} \leq \exp[-\exp(-\alpha)]. \quad (13)$$

Now we consider the computational evaluation of the connectivity of RRGs with 1500 nodes as a function of the connection radius and the rectangle side length. We start by analyzing the goodness of the upper bound that we have found for the probability of a RRG to be connected [see Eq. (13)]. In Fig. 4(a) we illustrate the plot of the probability of being connected versus the connection radius for graphs embedded into rectangles of side length  $a = 1$  (black circles) and  $a = 10$  (red circles). The solid circles represent the average values of 100 random realizations for these graphs. The values corresponding to the upper bound are plotted as broken lines. As can be seen, both the observed and the upper bound follow the same distribution and the upper bound is relatively close to the average observed values. In Fig. 4(b) we illustrate the change in the probability that the RRG is connected as a function of the connection radius, i.e., the minimum radius needed to make the network connected, for three different values of  $a$ . As the square is elongated the critical radius increases with the value of  $a$ . For instance, for  $a = 1$  the critical radius is about 0.2, and for  $a = 10$  it is about 0.9. Then, for  $a = 1$ ,  $\alpha \geq 185.3$  for the RRG to be connected. This value increases up to  $\alpha \geq 750.8$  for  $a = 5$  and to  $\alpha \geq 3813.8$  for  $a = 10$ . The main reason for this increase in the critical radius is that as we elongate the rectangle the points have to cover a longer region and consequently their separation increases. As a consequence, we need to increase the connection radius in order to guarantee the connectivity of the network. In other words, increasing the value of  $a$  necessarily implies having to increase the connection radius to make the network connected. The global increase of the critical radius with the side length of the rectangle is given in Fig. 4(c).

### C. Degree distribution

In the RRG the  $n$  nodes are distributed uniformly and independently on the unit rectangle. Then, the degree distribution can be easily estimated by considering the probability density function of having a node  $i$  of degree  $k_i$  given that there are  $n - 1$  other nodes uniformly distributed in the unit rectangle (see, for instance, Ref. [32]). This gives rise to the binomial distribution of the node degrees, which when  $(n - 1) \sim n$  is given by

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (14)$$

When  $n \rightarrow \infty$  and  $p$  is sufficiently small, this binomial distribution approaches very well a Poisson distribution of the form

$$p(k) \simeq \frac{\bar{k}^k \exp(-\bar{k})}{k!}. \quad (15)$$

As we have previously obtained an analytic expression for the average degree  $\bar{k}$  we can easily compute the degree distribution for RRGs. We select RRGs with 5000 nodes and radius of connection equal to 0.025. Then, we obtain the degree distribution for different values of the rectangle side length and take the average of 100 random realizations. In Fig. 5 we also plot the expected distribution using Eq. (15), in which we have plugged the values of the expected average degree obtained previously. As can be seen, independently of the side length of the rectangle, the RRG displays Poisson degree distributions. That is, the elongation of the rectangle does not affect the shape of the degree distribution of the nodes.

### D. Average shortest path distance

Let  $G = (V, E)$  be a simple connected graph. A *path* of length  $k$  in  $G$  is a set of nodes  $i_1, i_2, \dots, i_k, i_{k+1}$  such that for all  $1 \leq l \leq k$ ,  $(i_l, i_{l+1}) \in E$  with no repeated nodes. The *shortest-path* or *geodesic distance* between two nodes  $u, v \in V$

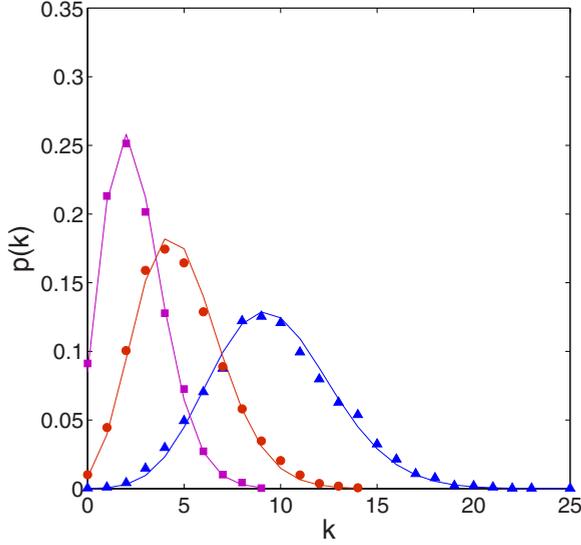


FIG. 5. (Color online) Degree distribution of RRGs with  $n = 5000$ , connection radius  $r = 0.025$ , and rectangle side lengths  $a = 1$  (triangles),  $a = 50$  (circles),  $a = 100$  (squares). The solid figures (triangles, circles, and squares) correspond to the average of 100 random realizations for the given network. The solid lines correspond to the shape of the Poisson distribution Eq. (15) with the corresponding average degree obtained from Eq. (8).

is defined as the length of the shortest path connecting these nodes. We will write  $d(u, v)$  to denote the distance between  $u$  and  $v$ . Here we will define, as is usual in network theory, the average path length to be the following quantity:

$$\langle l \rangle = \frac{2}{n(n-1)} \sum_{u < v} d(u, v). \quad (16)$$

The diameter of the graph is defined as the maximum of all the shortest path lengths in the graph, i.e.,  $D = \max_{u, v} d(u, v)$ . We consider here an upper bound for the average path length. Then, we start by considering that the  $n$  points are distributed homogeneously in the rectangle of sides  $a$  and  $b = a^{-1}$  in such a way that the points are equally spaced in the rectangle and are separated by a Euclidean distance  $r$ . The maximum number of contiguous nodes is realizable by placing them on one of the diagonals of the rectangle. In this case, the point which is the farthest from the rest of the  $n - 1$  points in the rectangle is one at any of the extremes of this diagonal. Let us designate this point as  $i$ . If the average path length of this node is  $\langle l_i \rangle$ , it is straightforward to realize that  $\langle l \rangle \leq \langle l_i \rangle$ , and we will have the desired upper bound. Let us designate the length of the main diagonal as  $c$ . Then, there are  $D = \frac{c}{r}$  connected nodes in this line. Let us now consider  $\langle l_i \rangle$  based only on those  $j$  nodes along the main diagonal of the rectangle (notice that this is an upper bound for  $\langle l_i \rangle$ ). That is

$$\sum_{j \in \text{diag}} d_{ij} = 1 + 2 + \dots + D = \frac{D(D+1)}{2}.$$

Consequently,

$$\langle l_i \rangle \leq \frac{D+1}{2}.$$

Because,  $\langle l \rangle \leq \langle l_i \rangle$ , we have

$$\langle l \rangle \leq \langle l_i \rangle \leq \frac{D+1}{2}.$$

It is easy to note that  $D < \frac{a}{r} + \frac{1}{ar}$ . That is, the length of the diagonal is smaller than the sum of the length of the two sides of the rectangle (in terms of number of nodes), which are  $\frac{a}{r}$  and  $\frac{b}{r} = \frac{1}{ar}$ . Then, we have

$$\langle l \rangle \leq \langle l_i \rangle \leq \frac{D+1}{2} \leq \frac{\frac{a}{r} + \frac{1}{ar} + 1}{2} = \frac{a^2 + ar + 1}{2ar}. \quad (17)$$

An important consequence of this bound is that it allows us to find the asymptotic behavior of the average path length as the rectangle becomes more elongated. That is, if we fix the connection radius we can see that  $\lim_{a \rightarrow \infty} \langle l \rangle = \infty$ . In other words, as the rectangle becomes more elongated the RRG becomes a large world with a very large average shortest path distance. As  $a \rightarrow \infty$ , the connection radius  $r$  has to be increased to keep the connectivity of the RRG. Thus, the infinite growth of the path length is not observed in practice due to the connectivity constraints imposed by the connection radius. Our further calculations show that the average path length is of the same order of magnitude as the right-hand side of Eq. (17), which means that we can write it as

$$\langle l \rangle \sim \frac{a^2 + ar + 1}{2ar}. \quad (18)$$

We now study computationally RRGs with 1500 nodes. For every value of  $a$  we report the average of 100 random realizations. First, we analyze the goodness of the upper bound found here for the average path length. In Fig. 6(a) we illustrate the variation of  $\langle l \rangle$  with the connection radius for two different values of  $a$ , namely  $a = 5$  and  $a = 10$ . In the same plot we illustrate the values of the upper bound obtained with Eq. (17), where it can be seen that the upper bound is very close to the average shortest path obtained for these RRGs. Particularly, for large values of  $r$  the observed values are almost identical to those of the upper bound. As observed in Fig. 6(b) the average path length not only changes with the variation of the connection radius but also with the rectangle side length. This is already expected from Eq. (18) where it can be seen that as  $a \rightarrow \infty$  the average path length also grows to infinity for fixed  $r$  (see observation at the end of the previous paragraph).

These results agree with our intuition that as we elongate the rectangle there are nodes that are farther apart from each other and as a result the average path length of the whole graph increases. For a better analysis of this relation we plotted the results of the average path length for 100 random realizations of the previously mentioned RRGs versus  $a$ , with  $m = 11250$  edges, in Fig. 6(c). We also plot here the upper bound [Eq. (17)]. It can be seen that there is an almost linear increase of  $\langle l \rangle$  for values of  $1 \leq a \lesssim 15$ , after which the dependence is very flat. In this region we have that  $a \rightarrow \infty$ , which corresponds to a good approximation of a one-dimensional RGG. For  $a = 1$  it is known that the average path length depends only on the inverse of the radius,  $\langle l_e \rangle = \Theta(1/r)$  [33]. The actual radius used for the plot in Fig. 7(c) is  $r = 0.0578$ , which gives an estimate of the average

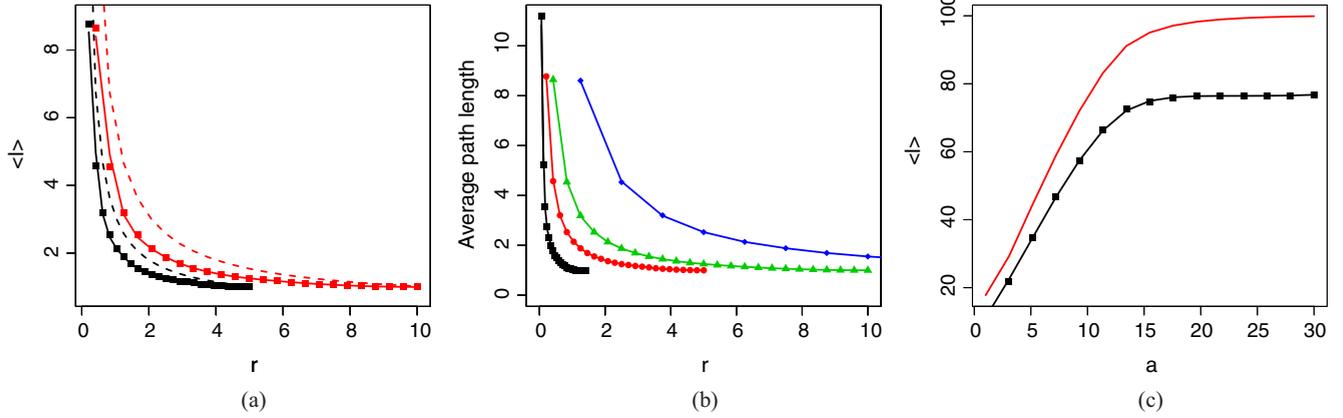


FIG. 6. (Color online) (a) Plot of the average path length for RRGs with  $n = 1500$  nodes with different connection radii for  $a = 5$  (black squares) and for  $a = 10$  (red squares). The solid squares represent the average of  $\langle l \rangle$  for 100 random realizations (lines connecting the squares are used to guide the eye). The dotted lines represent the upper bounds [Eq. (17)] for  $\langle l \rangle$  in RRGs for the corresponding values of  $a$ . (b) Change of the average path length versus radius for networks with  $n = 1500$  nodes for different values of  $a$ : black squares ( $a = 1$ ), red circles ( $a = 5$ ), green triangle ( $a = 10$ ), blue rhombus ( $a = 30$ ). (c) Illustration of the upper bound (solid line) for the average path length as a function of  $a$  for a network with the same size as before and connection radii that guarantees  $m = 11\,250$  edges for a given value of  $a$ .

path length of 17.3, which is not too far from the observed value in the plot for  $a = 1$ . In the case of  $a = 30$  we are in the presence of a very elongated rectangle, which is very similar to a one-dimensional RGG. A crude estimate of the average path length in this case would be  $\langle l_e \rangle = n / \langle k \rangle$ , which in the current case will give  $\langle l_e \rangle \approx 100$ , which is relatively close to the observed value of  $\langle l \rangle \approx 75$  for  $a = 30$ .

### E. Clustering coefficient

One of the most important network parameters is the *local clustering coefficient*, usually known as the *Watts-Strogatz clustering coefficient*, of a node  $i$ . This parameter is defined as [10]

$$C_i = \frac{2t_i}{k_i(k_i - 1)}, \quad (19)$$

where  $t_i$  is the number of triangles involving the node  $i$  and  $k_i$  is the degree of the node  $i$ . Taking the mean of these values as  $i$  varies among the nodes in  $G$ , one gets the *average clustering coefficient* of the network:  $\langle C \rangle = \frac{1}{n} \sum_{i=1}^n C_i$ .

The average clustering coefficient of a RGG has been obtained by Dall and Christensen [17] for  $r^2 = \frac{\log n + \alpha}{n\pi}$  when  $n \rightarrow \infty$  and  $\alpha \rightarrow \infty$ , where  $\alpha \in \mathbb{R}$  is a constant for a given number of nodes (see Sec. III B):

$$\langle C_d \rangle = \begin{cases} 1 - H_d(1) & d \text{ even} \\ \frac{3}{2} H_d(1/2) & d \text{ odd,} \end{cases} \quad (20)$$

where  $d$  is the dimension of the hypercube in which the nodes are embedded and

$$H_d(x) = \frac{1}{\sqrt{\pi}} \sum_{i=x}^{d/2} \frac{\Gamma(i)}{\Gamma(i + \frac{1}{2})} \left(\frac{3}{4}\right)^{i + \frac{1}{2}}, \quad (21)$$

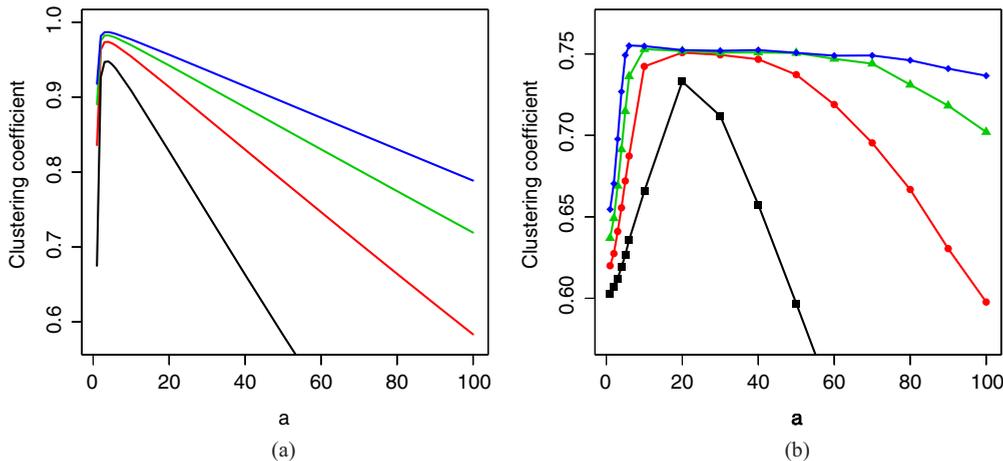


FIG. 7. (Color online) Illustration of the dependence of the clustering coefficient with the rectangle side length for different connection radii:  $r = 0.05$  (black),  $r = 0.1$  (red),  $r = 0.15$  (green),  $r = 0.20$  (blue). In plot (a) we show the analytical results and in plot (b) the observed ones. Both results are obtained for RRG with 1500 nodes and for the observed ones we averaged the results of 100 random realizations.

where  $\Gamma(i)$  is the  $\Gamma$  function. Thus, for  $d = 2$ ,  $\langle C_2 \rangle = 1 - \frac{3\sqrt{3}}{4\pi} \approx 0.5865$  and for  $d = 1$ ,  $\langle C_1 \rangle = 3/4 = 0.75$ .

Here, however, we are interested in an expression that accounts for the variations of the clustering coefficient with both the connection radius and the rectangle side length. Our strategy is similar to the one used in Ref. [17]. That is, let  $i$  and  $j$  be two connected nodes in a RRG, which are separated at a Euclidean distance  $\delta$  from each other. Let us draw two circles of radius  $r$  centered, respectively, at  $i$  and  $j$ . Let  $\delta \leq r$  such that the two nodes are connected. Then, because  $\delta < 2r$  the two circles overlap. Because  $i$  and  $j$  are connected, any point in the area formed by the overlap of the two circles will form a triangle with the nodes  $i$  and  $j$ . In addition, any node inside the two circles that is not in the overlapping area forms a path of length two with the nodes  $i$  and  $j$ . Thus, if we quantify the ratio of the overlapping area to the total area of the circle we account for the ratio of the number of triangles to open triads in which the nodes  $i$  and  $j$  take place, i.e., the clustering coefficient. This ratio is given by

$$\langle C \rangle = \frac{2r^2 \arccos\left(\frac{\delta}{2r}\right) - \frac{1}{2}\delta\sqrt{4r^2 - \delta^2}}{\pi r^2}. \quad (22)$$

At this point we only need an estimation of the length  $\delta$  between two connected nodes in a RRG. We use here a simple approach based on the following intuition. Let us start by considering  $n$  nodes in a square in such a way that they form a regular square lattice. Then,  $\delta$  is proportional to the length of the side of the rectangle  $a$  divided by the number of circles along this side. As we have a square, the number of points along the side of length  $a$  is the same as that for the other side. Consequently,  $\delta \sim n^{-1/2}$ . If we elongate the rectangle to  $a \rightarrow \infty$ , which resembles a straight line, we will have that the separation between the two points is just the length of the straight line divided by the number of nodes,  $\delta \sim an^{-1}$ . For a general rectangle the separation between two points in a line along the edge side of length  $a$  is given by  $\delta \sim an^{-\gamma}$ , where  $\gamma \sim a(a+b)^{-1}$ . Notice that when  $a = 1$  ( $b = 1$ ) we have  $\delta \sim n^{-1/2}$  and when  $a \rightarrow \infty$  ( $b \rightarrow 0$ ) we have  $\delta \sim an^{-1}$ . We remind the reader that here we consider only the case  $b = a^{-1}$ .

Plugging  $\delta \sim an^{-\gamma}$ , with  $\gamma \sim a(a+b)^{-1}$  into Eq. (22) we obtain some surprising results. In Fig. 7(a) we illustrate the dependence of  $\langle C \rangle$  with  $a$  for different radii based on Eq. (22) with the estimated value of  $\delta$  given before. Notably, the clustering coefficient is predicted to change nonmonotonically with the rectangle side length. Instead, for small values of  $a$  the clustering coefficient is predicted to increase to a maximum value and after it the clustering decreases linearly. In addition, according to this model, as the connection radius increases the clustering coefficient is expected to increase for the same value of  $a$ . In closing, the largest value of the clustering is expected for certain specific value of  $a$  and a relatively large connection radius.

In order to verify these findings we compute the average clustering coefficient of RRGs with  $n = 1500$  and the same connection radii used in the simulations with the analytical formulas. The results of the variation of the clustering with the rectangle side length are illustrated in Fig. 7(b). As can be seen the clustering increases up to a maximum, whose

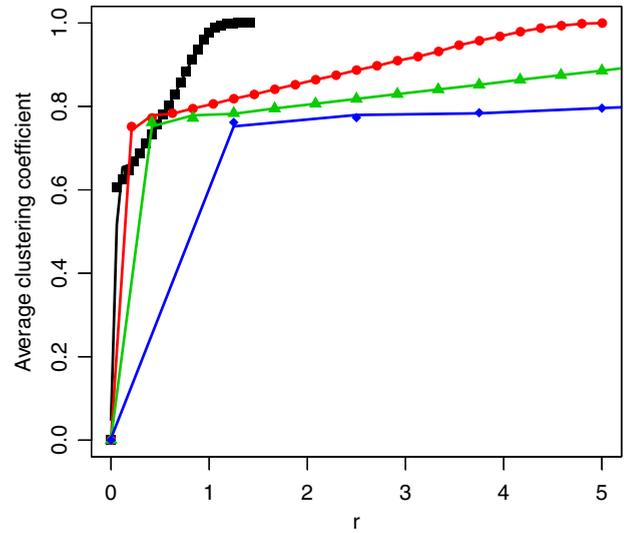


FIG. 8. (Color online) Variation of the average path length with the radius for RRGs with 1500 nodes and  $a = 1, 5, 10, 30$ . Every point is the average of 100 random realizations.

location depends on the connection radius, and then decays with the increase of the elongation of the rectangle. We have not been able to capture the dependence of  $\delta$  with the radius in our previous reasoning, but we very well captured the behavior of the clustering of having a nonmonotonic change with  $a$ . Also, these experiments show that the increase of the connection radius increases the average clustering coefficient as predicted by our analytical results. As can be seen in Fig. 7(b) for  $a = 1$  and small radius the average clustering coefficient is  $\langle C \rangle \approx 0.61$ , which is very close to the expected value for the two-dimensional RGG according to Ref. [17]. When  $a = 30$  and the radius is relatively large, the average clustering coefficient is  $\langle C \rangle \approx 0.75$ , which coincides with the exact value expected for the one-dimensional RGG according to Ref. [17]. Consequently, the RRG generalizes the values of the clustering coefficient of both, the one- and two-dimensional RGG, for  $a = 1$  and  $a \rightarrow \infty$ , respectively. In addition, it provides a series of intermediate values of the clustering coefficient for intermediate values of the side length of the rectangle.

We now further explore the relation between the radius  $r$  and the clustering for RRGs with different side lengths. We consider graphs with  $n = 1500$  nodes and  $a = 1, 5, 10, 30$ . As the radius increases the graph is becoming more and more dense, which means that the clustering coefficient is characterized by an abrupt increase at the beginning of the plot and then a linear increase until the value of  $\langle C \rangle = 1$  is reached for the complete graph (see Fig. 8).

#### IV. CONCLUSIONS AND FUTURE OUTLOOK

We have introduced here a generalization of the RGG in which we embed the points into a unit rectangle instead of on a unit square. We consider a rectangle with sides of lengths  $a$  and  $b = 1/a$ , such that as when  $a = 1$  we have the particular case of the *classical* random geometric graph embedded in a unit square. Also, when  $a \rightarrow \infty$  we have a very

elongated rectangle which resembles a one-dimensional RGG. We have provided computational and analytical evidence that reaffirm the fact that the topological properties of the RRG differ significantly from those of the RGG. In this respect we have obtained analytical expressions or bounds for the average degree, degree distribution, connectivity, average path length, and the clustering coefficient of RRGs. In general, these properties depend on the connection radius as well as on the rectangle side length. Most of the dependencies found here for these properties in terms of the rectangle side length are monotonic. The only exception is the clustering coefficient. This index first increases up to a critical value of  $a$ , which depends on the connection radius, and then decays linearly with the increased elongation of the rectangle.

The introduction of the RRGs opens new possibilities for studying spatially embedded random graphs. For instance, the study of dynamical processes taking place on the nodes and edges of these graphs is of great interest to explore how

the shape constraints influence the dynamics on the RRGs. On the other hand, the analysis of rectangular proximity graphs, such as the rectangular Gabriel graphs and random rectangular neighborhood graphs, is also interesting for many of the practical applications of these graphs as mentioned in the Introduction. The generalization of the RRG model to higher dimensions is also of both theoretical and practical interest. In closing, the current work is expected to open new horizons for the study of random spatial graphs and its applications in physics and beyond.

#### ACKNOWLEDGMENTS

E.E. thanks the Royal Society for a Wolfson Research Merit Award. M.S. thanks Weird Group PLC and EPSRC for partial financial support of his work. We thank the two anonymous referees for valuable suggestions that helped to improve this paper.

- 
- [1] E. Estrada, in *Mathematical Tools for Physicists*, edited by M. Grinfeld (John Wiley & Sons, New York, 2014).
  - [2] G. Berkolaiko, R. Carlson, S. A. Fulling, and P. Kuchment, *Quantum Graphs and Their Applications* (AMS, Providence, RI, 2006).
  - [3] E. Estrada, *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, Oxford, 2011).
  - [4] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
  - [5] L. d. F. Costa, O. Oliveira, G. Travieso, F. A. Rodrigues, P. Villas Boas, L. Antigueira, M. Viana, and L. Correa Rocha, *Adv. Phys.* **60**, 329 (2011).
  - [6] M. J. E. Newman, [arXiv:cond-mat/0202208](https://arxiv.org/abs/cond-mat/0202208).
  - [7] B. Bollobás, *Random Graphs* (Academic Press, New York, 1985).
  - [8] P. Erdős and A. Rényi, *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 17 (1960).
  - [9] A. L. Barabási and R. Albert, *Science* **286**, 509 (1999).
  - [10] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
  - [11] M. Barthélemy, *Phys. Rep.* **499**, 1 (2011).
  - [12] D. Urban and T. Keitt, *Ecology* **82**, 1205 (2001).
  - [13] A. Perna, S. Valverde, J. Gautrais, C. Jost, R. Solé, P. Kuntz, and G. Theraulaz, *Physica A* **387**, 6235 (2008).
  - [14] J. Buhl, J. Gautrais, R. V. Solé, P. Kuntz, S. Valverde, J. L. Deneubourg, and G. Theraulaz, *Eur. Phys. J. B* **42**, 123 (2004).
  - [15] E. Santiago, J. X. Velasco-Hernández, and M. Romero-Salcedo, *Exp. Syst. Appl.* **41**, 811 (2014).
  - [16] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, 2003).
  - [17] J. Dall and M. Christensen, *Phys. Rev. E* **66**, 016121 (2002).
  - [18] P. Gupta and P. R. Kumar, in *Stochastic Analysis, Control, Optimization and Applications* (Birkhäuser, Boston, 1999).
  - [19] G. J. Pottie and W. J. Kaiser, *Comm. ACM* **43**, 51 (2000).
  - [20] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, in *Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking* (ACM, New York, 1999), p. 263.
  - [21] E. N. Gilbert, *Ann. Math. Stat.* **30**, 1141 (1959).
  - [22] P. Wang and M. C. González, *Phil. Trans. Royal Soc. A: Math., Phys. Eng. Sci.* **367**, 3321 (2009).
  - [23] A. Díaz-Guilera, J. Gómez-Gardeñes, Y. Moreno, and M. Nekovee, *Int. J. Bif. Chaos* **19**, 687 (2009).
  - [24] M. Nekovee, *New J. Phys.* **9**, 189 (2007).
  - [25] V. Isham, J. Kaczmarek, and M. Nekovee, *Phys. Rev. E* **83**, 046128 (2011).
  - [26] Z. Toroczkai and H. Guclu, *Physica A* **378**, 68 (2007).
  - [27] D. Watanabe, *The Ninth International Symposium on Operations Research and Its Applications* (ORSC & APORC, Chengdu-Jiuzhaigou, China, 2010).
  - [28] M. Desai and D. Manjunath, *Comm. Lett., IEEE* **6**, 437 (2002).
  - [29] C. H. Foh, G. Liu, B. S. Lee, B. C. Seet, K. J. Wong, and C. P. Fu, *Comm. Lett., IEEE* **9**, 31 (2005).
  - [30] E. Godehardt and J. Jaworski, *Rand. Struct. Algorithm.* **9**, 137 (1996).
  - [31] M. D. Penrose, *Ann. Appl. Prob.* **7**, 340 (1997).
  - [32] A. Antonioni and M. Tomassini, *Phys. Rev. E* **86**, 037101 (2012).
  - [33] G. Ercal, in *NETWORKING* (Springer, Berlin/Heidelberg, 2011), p. 281.