Random geometric graphs

Jesper Dall* and Michael Christensen
Fysisk Institut, SDU–Odense Universitet, Campusvej 55, DK–5230 Odense M, Denmark
(Received 4 March 2002; published 24 July 2002)

We analyze graphs in which each vertex is assigned random coordinates in a geometric space of arbitrary dimensionality and only edges between adjacent points are present. The critical connectivity is found numerically by examining the size of the largest cluster. We derive an analytical expression for the cluster coefficient, which shows that the graphs are distinctly different from standard random graphs, even for infinite dimensionality. Insights relevant for graph bipartitioning are included.

DOI: 10.1103/PhysRevE.66.016121 PACS number(s): 05.10.Ln, 64.60.Ak, 89.75.Da

I. INTRODUCTION

The interest in complex networks has exploded over the last five years [1,2], where data on very large networks such as the world wide web [3–5], collaborations in the scientific community [6], transportation [7], movie actor collaborations [8] etc., have become accessible.

Random graphs are often used to model complex networks [9]. Ever since Erdös and Rényi's groundbraking work more than forty years ago [10], intense theoretical research on random graphs has been taking place [4,11–13]. In contrast to random graphs the interactions between the sites in a lattice are usually between nearest neighbors, reflecting a myopic world. Lattices are therefore often said to be at the other end of the spectrum of network models [14,15].

Properties of real networks such as robustness [16,17], growth [11,18–20], and topology have attracted much attention, primarily from physicists. It has been consistently shown that many of the networks possess small world characteristics [8,21,22]. Like random graphs, small world networks are characterized by short average distances between any two sites, and by a high degree of localness, much like in lattices. However, individually, random graphs and lattice models in their pure forms are poor models of many real world networks. One could argue that high-dimensional lattices have the necessary high clustering and low average path length, though this has not been explored much [23]. In the present paper, we provide results on high-dimensional systems.

A random geometric graph (RGG) is a random graph with a metric. It is constructed by assigning each vertex random coordinates in a d-dimensional box of unit volume, i.e., each coordinate is drawn from a uniform distribution on the unit interval. RGGs have been used sporadically in real networks modeling [24] and extensively in continuum percolation [25–29], but almost exclusively in two and three dimensions. Although RGGs are the continuum version of lattices, they deserve some attention of their own, since percolating continuum systems display behavior that lattices are incapable of [30]. In addition, the connectivity in RGGs can be increased in a more natural way than by adding new bonds randomly in lattices.

Recently, continuum percolation has been used in the study of the stretched exponential decay of the correlation function in random walks on fractals and the conjectured relation to relaxation in complex systems [31]. However, continuous systems in general and RGGs in particular are relevant whenever we need a multidimensional system with a metric, as for example when modeling the spread of diseases [32].

In this paper, we study RGGs in arbitrary dimensions. In low dimensions the systems are dominated by local interactions. For higher dimensions RGGs are usually believed to approach standard random graphs, which we show is true only in some respects. We focus on ''phase transitions'' [13,33,34] at the percolation threshold by looking at the size of the largest cluster, and we determine how the value of the critical parameter in RGGs approaches that of random graphs as the dimension increases. We also extract the distribution of cluster sizes in the critical region. Furthermore, an expression for the cluster coefficient, a quantity that has attracted much interest in network theory recently, is derived. Results relevant for graph bipartitioning are established. Finally, we discuss how to implement random geometric graphs efficiently.

The layout of this paper is as follows. In Secs. II and III, we describe random graphs and random geometric graphs, respectively. In Sec. IV we present our results and Sec. V contains the details regarding the implementation. Finally, we sum up in Sec. VI.

II. RANDOM GRAPHS

Random graphs consist of N vertices (points/sites) and K edges (lines) where each possible edge is present with probability p, i.e. $K=pN(N-1)/2$.[1] To keep the discussion independent of the system size N, graphs are often characterized by the connectivity (degree) $\alpha=2K/N=pN$, i.e. the average number of connections per vertex, instead of K or p. As the connectivity increases clusters of vertices appear, where a cluster consists of all vertices linked together by edges, directly or indirectly.

The size of the largest cluster in the macroscopic limit

*Email address: j.dall@fysik.sdu.dk

[1]From here on we consider $N\simeq N-1$ in accordance with the literature [4,35], since we are only investigating large systems.

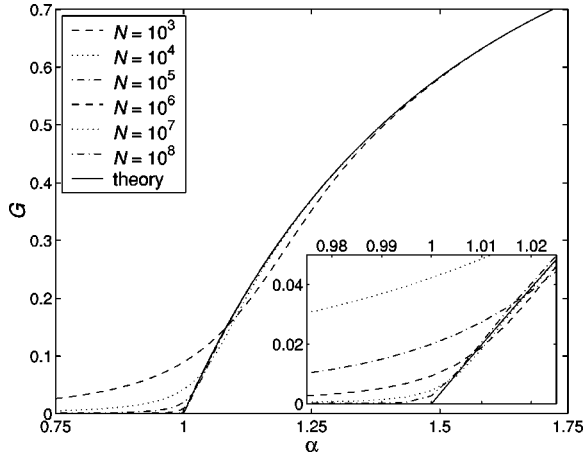

FIG. 1. The size of the largest cluster in random graphs as a function of the connectivity. Note that for $N>10^6$ the Monte Carlo data are almost indistinguishable from the theoretical result in Eq. (3). Error bars are not shown since they are in all cases less than the width of the lines. Inset: A closer look at the percolation threshold $\alpha_c=1$.

$N\to\infty$ can be calculated analytically [10,12]. It is $NG(\alpha)$, where

$$G(\alpha)=1-\frac{1}{\alpha}\sum_{n=1}^{\infty}\frac{n^{n-1}}{n!}(\alpha e^{-\alpha})^n. \tag{1}$$

By the use of [12]

$$y=\sum_{n=1}^{\infty}\frac{n^{n-1}}{n!}x^n \Leftrightarrow x=ye^{-y}, \tag{2}$$

we can invert Eq. (1), getting

$$\alpha(G)=-\frac{1}{G}\ln(1-G) \tag{3}$$

from which it is trivial to show that $\alpha_c=1$. With Eq. (3) it is an easy task to plot the fraction of vertices in the largest cluster—the giant component—as done in Fig. 1, where we see the prototype of a phase transition in combinatorial problems.

In random graphs the probability distribution of edges p_k is binomial

$$p_k=\binom{N}{k}p^k(1-p)^{N-k}\simeq\frac{\alpha^k e^{-\alpha}}{k!}, \tag{4}$$

where the approximation resulting in the Poisson distribution is valid for large systems sizes N, which is exactly the limit in which we are interested. The critical connectivity α_c for graphs with arbitrary random degree distribution p_k has recently been derived by other techniques than those originally leading to Eq. (1) [4,36,37]. Unfortunately, we cannot use these results in connection with random geometric graphs, as will become clear in the following section.

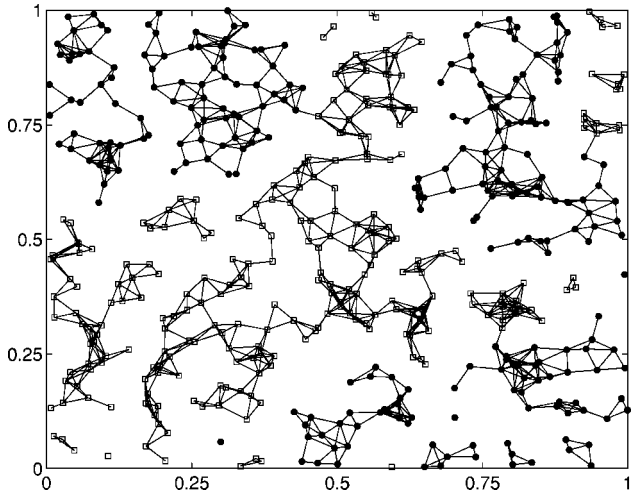

FIG. 2. A 2D random geometric graph with $N=500$ and $\alpha=5$. The graph is bipartitioned—see Sec. IV E. There are no edges across the boundaries, i.e., the boundary conditions are open, not continuous.

III. RANDOM GEOMETRIC GRAPHS

A d-dimensional RGG is a graph where each of the N vertices is assigned random coordinates in the box $[0,1]^d$, and only points ''close'' to each other are connected by an edge. The degree distribution of a RGG with average connectivity α is therefore given by Eq. (4) as well. However, a RGG is a special kind of random graph with properties not captured by the theoretical tools mentioned above. For one thing, the probability that three vertices are cyclically connected is different in random graphs and RGGs, regardless of the degree distribution of the random graph.

RGGs are sometimes named spatial graphs [8]. Figure 2 illustrates a RGG in two dimensions (2D). As in lattices, different boundary conditions can be applied. We will see that toroidal (continuous) boundary conditions make a vital difference compared to having open boundary conditions.

The volume of a d-dimensional (hyper)sphere with radius r is

$$V_{sphere}=\frac{\pi^{d/2}r^d}{\Gamma\left(\frac{d+2}{2}\right)}, \tag{5}$$

where $\Gamma(x)$ is the gamma function. This volume is needed in order to find the edges in RGGs.

To ''visualize'' a RGG in general, one can think of a box filled with small spheres with radius r and volume V given by Eq. (5), where points are connected by an edge only if the distance between their centers is $<2r$, i.e., if the spheres overlap. Since the total volume of our box is 1, the probability that two arbitrarily chosen vertices are connected is equal to the volume of a sphere with radius $R=2r$. In continuum percolation theory this volume is denoted the *excluded volume* V_{ex}, where $V_{ex}=2^dV$ in a RGG. The excluded volume is the basic quantity of interest because it is directly related to the connectivity

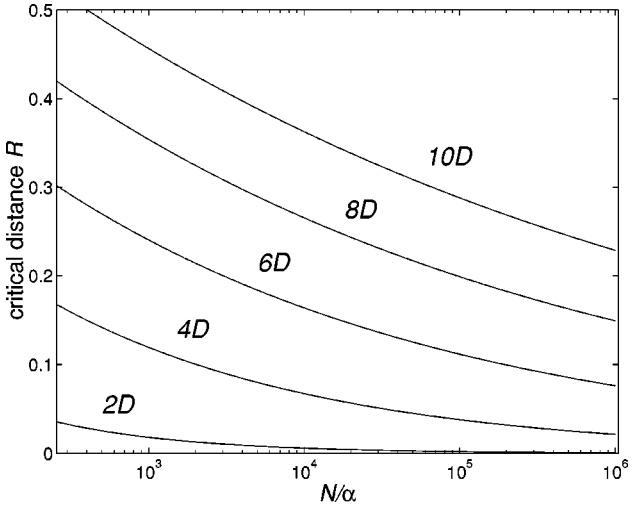

FIG. 3. The critical distance in random geometric graphs in various dimensions. Points within this distance of each other are connected by an edge. The critical distance is equivalent to the radius R of the excluded volume associated with each point.

$$\alpha = Np = NV_{ex}, \tag{6}$$

from which it is clear why the connectivity is frequently called the total excluded volume of the system. Equations (5) and (6) give us

$$R = \frac{1}{\sqrt{\pi}}\left[\frac{\alpha}{N}\Gamma\left(\frac{d+2}{2}\right)\right]^{1/d}. \tag{7}$$

Figure 3 shows the radius R of the excluded volume as a function of $N/\alpha = 1/p = 1/V_{ex}$. R decreases monotonically: for a given connectivity α the spheres have to become smaller when more vertices are added to the graph.

Equation (7) provides us with the required relation between α and R when creating a RGG. The distance between every pair of vertices must be calculated, and an edge is added if the distance is less than R. Thus, it seems unavoidable to have a runtime of $O(N^2)$ making it unfeasible to investigate as large systems as with random graphs—see Fig. 1—where the number of calculations for a given α needed to create all the edges is $O(N)$. To overcome this obstacle we have designed a data structure that is described in Sec. V, with a runtime of $O(N^\beta)$, where $\beta \simeq 1.3$. This allows us to study RGGs with up to $N = 4^{11} > 4 \times 10^6$ vertices, which is more than an order of magnitude larger than usually accomplished [38].

IV. RESULTS

In our simulations of RGGs we define α_c to be the lowest connectivity at which the fraction of vertices in the largest cluster is >0 in the macroscopic limit. We make the bold claim that the systems we are able to analyze consist of enough points to make the critical connectivity almost as sharply defined as in Fig. 1. However, our main purpose is not to derive high precision percolation thresholds. Instead, we are more interested in the critical connectivity as a function of the dimension of the RGGs.

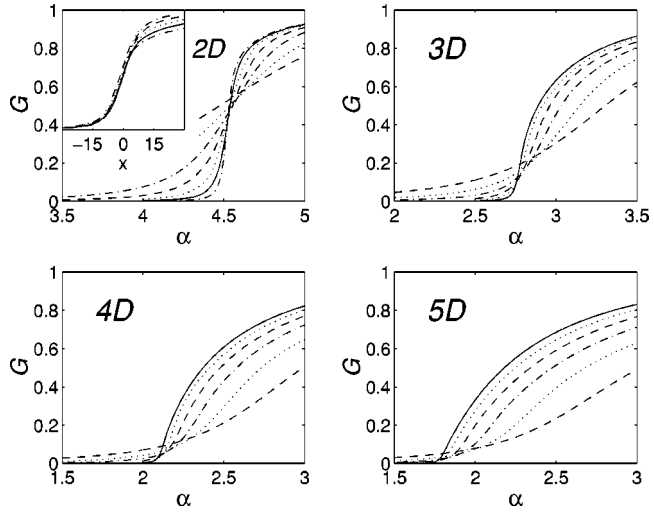

FIG. 4. The average fraction of vertices in the largest cluster for various system sizes N (see the legend in Fig. 5) in random geometric graphs with no edges across the boundaries. The inset in 2D illustrates a finite size scaling—see the text. In higher dimensions the general shape of the curves as N increases is nontrivial. Compare with Fig. 5. Error bars are $<10^{-3}$ for all curves and therefore omitted.

In this paper, we express our threshold values in terms of α. Other popular choices are the fractional volume s occupied by the spheres [30] or the density N of spheres. The relation between these parameters at the percolation threshold is

$$\alpha_c = N_c V_{ex} = -2^d \ln(1 - s_c) \tag{8}$$

(see e.g. Ref. [25] for a derivation). Usually, in continuum percolation the volume V of each sphere is fixed while N is the independent variable in a system of size $[0,L]^d$. The approach of measuring N_c or s_c for various values of L has been used in both two [39] and three [38] dimensions, i.e., for discs and spheres, where the critical values are determined by the use of finite size scaling. This procedure resembles site percolation in lattices. From the previous sections it is clear that we take a route closer to bond percolation in lattices by fixing $L=1$ while tuning α for different values of N. In Sec. V we describe how this has been carried out in practice.

A. The size of the largest cluster

Let $G_d(\alpha)$ denote the fraction of vertices in the largest cluster in d dimensions. Since a RGG in the limit of infinite dimension is often assumed equivalent to a random graph, we expect that Eq. (3) provides us with an expression for $G_\infty(\alpha)$. But what does $G_d(\alpha)$ look like for finite d? And what is the behavior of $\alpha_c(d)$? How does it approach $\alpha_c(\infty)$ as d increases? These are the questions addressed in this and the following section.

Figures 4 and 5 illustrate the average size of the largest cluster in RGGs in 2, 3, 4, and 5 dimensions with and without toroidal boundary conditions. The curves correspond to

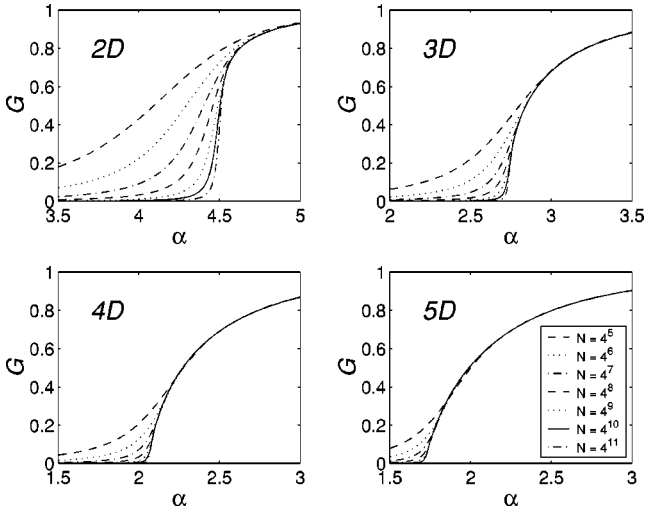

FIG. 5. Like Fig. 4 but with *continuous* boundary conditions. We see that the point at which the largest cluster becomes macroscopic is sharply defined and can immediately be determined by the eye with high precision (Table I). The overall behavior of the graphs for higher dimensions is much closer to Fig. 1 than Fig. 4 is. As d increases the α interval where there is a significant difference between curves with different N get smaller and smaller. Error bars are $<10^{-3}$ for all curves and therefore omitted.

$N=4^k$ vertices with $k=5,6,\ldots,11$, where the larger systems display the sharpest transitions. The legend in Fig. 5 applies to all diagrams in Figs. 4 and 5. In these eight diagrams each curve is based on 300 data points. In other words, $G_d(\alpha)$ is calculated in intervals of $\Delta\alpha=0.005$ resulting in the smooth lines in the figures. For every data set we have averaged over enough runs for error bars to be completely negligible.

Since continuous boundary conditions mean addition of extra edges, the size of the largest component $G(\alpha)$ obviously grows faster in Fig. 5 than in Fig. 4, especially in the smaller systems. These relatively few extra edges make a decisive difference, connecting vertices not already in the same cluster. Since toroidal systems are models of bulk systems, G is much less N dependent in that case. However, ''unphysical'' RGGs with open boundaries may seem that they are the most popular RGG versions in the literature. Consequently, we consider them alongside the continuous case.

From Figs. 4 and 5 we see that the continuous boundary conditions make the transition where $G>0$ is more abrupt, but that an estimation of α_c does not depend much on the boundary conditions if only we base our judgment on large enough systems. This is confirmed in the inset of Fig. 4, where $\alpha_c=4.53$ is obtained by finite size scaling, i.e., plotting $G(x)$, where $x=N^{1/\nu}(\alpha-\alpha_c)$. However, it is clearly easier to make precise estimates of the critical connectivity with than without continuous boundary conditions. We note in passing that the exponent $\nu=3$ is equal to the value of ν found in random graphs [13].

B. The critical connectivity

With numerically obtained knowledge of $G(\alpha)$, it is possible to extract α_c. The procedure is simple. By inspection of Fig. 5 we can estimate α_c for $d\leqslant 5$. To obtain further data points we have to run our algorithm on RGGs with $N=4^{10}$ for systems of larger dimensions as well. Though this results in increased runtime per graph, the results get more homogeneous and fewer runs are needed in order to get a decent estimate of $G_d(\alpha)$. Our findings presented in Table I and Fig. 6 strongly suggest that

$$\alpha_c(d)=\alpha_c(\infty)+Ad^{-\gamma}, \tag{9}$$

where $\alpha_c(\infty)=1$, $\gamma=1.74(2)$, and $A=11.78(5)$. As expected, Eq. (9) predicts that $\alpha_c(\infty)$ is equal to α_c in random graphs, confirming that RGGs and random graphs become more and more similar as d increases. However, when we derive the cluster coefficient, we will see that this is not true in all respects.

TABLE I. The critical connectivity α_c in random geometric graphs of dimension d with continuous boundary conditions. The data are plotted in Fig. 6. The estimated errors in α_c in the last row are rather conservative.

d	2	3	4	5	6	7	8
α_c	4.52	2.74	2.06	1.72	1.51	1.39	1.30
$\pm$	0.01	0.01	0.02	0.02	0.02	0.02	0.02

Finally, we note that our findings are in accordance with the most precise estimates that we know of $\alpha_c=4.512\,23(5)$ [29] and $\alpha_c=2.734(6)$ [38] in 2D and 3D, respectively, obtained by the use of finite size scaling. For $d>3$, we have not been able to find any estimates of α_c to compare with [40].

C. The distribution of cluster sizes

Having examined the size of the largest cluster and the critical connectivity, we now look at the distribution of cluster sizes in RGGs.

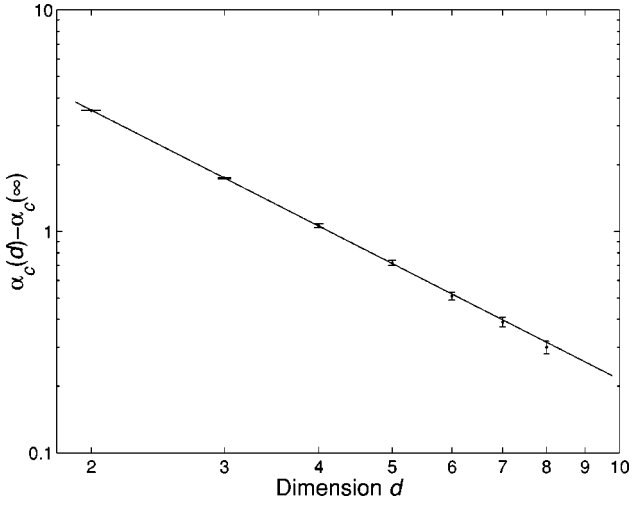

FIG. 6. Scaling of the critical connectivity as a function of the dimension of the random geometric graphs reveals a power-law relation [Eq. (9)]. For $d\leqslant 5$ the data points are estimated by close inspection of Fig. 5. For $d>5$, α_c is based on runs with $N=4^{10}$ points. Error bars are included. See Table I.

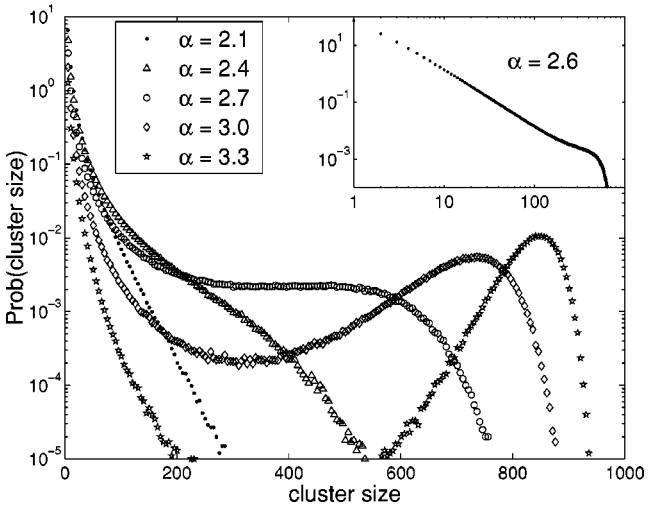

FIG. 7. The distribution of cluster sizes in 3D random geometric graphs with $N=1000$ vertices in the vicinity of the critical connectivity $\alpha_c=2.74$. The inset shows that for $\alpha \simeq \alpha_c$ the cluster sizes are given by a power law. For each value of α the data points are based on 10^6 graphs.

The inset illustrates the scale-free power-law distribution at $\alpha=2.6$. Right below α_c, clusters of all sizes can be encountered. The small hump at large cluster sizes is always present because the clusters cannot contain more than all of the vertices. The clusters pile up when their size approaches this boundary, in this case a cluster size of 1000, just below the inevitable cutoff.

Our simulations show that for α significantly below α_c the distribution is approximately exponential. As the connectivity increases the distribution becomes power-law-like. As α is further increased the distribution is separated in two parts; there are no clusters of medium size, only the largest macroscopic cluster and a few small ones around it. We have observed this overall behavior in all our tests of the distribution of cluster sizes in various dimensions.

Figure 7 shows our data in 3D. For $\alpha=2.1$ ($\cdot$) the data points lie on an almost straight line indicating an exponential distribution. Increasing the connectivity to $\alpha=2.4$ ($\triangle$) results in a broader distribution that is no longer exponential. Right at the critical connectivity ($\bigcirc$) the distribution flattens out. Clusters of all sizes are observed. Right above α_c ($\diamond$) two separate regions begin to materialize. Already at $\alpha=3.3$ ($\star$) the largest cluster makes it highly unlikely that a cluster of medium size can be present as well. The distribution is cut in two.

D. The cluster coefficient

In network theory the cluster coefficient C is an often calculated quantity [1,21,23], which is defined in the following way. Let the vertices i and j be connected directly to a common vertex k. C is then the probability that vertex i and vertex j are directly connected as well. From this we see that the cluster coefficient is a measure of the ''cliquishness'' of the graph. In this section, we derive $C=C_d$ analytically in arbitrary dimensions d, showing that C_d decreases in an exponential fashion.

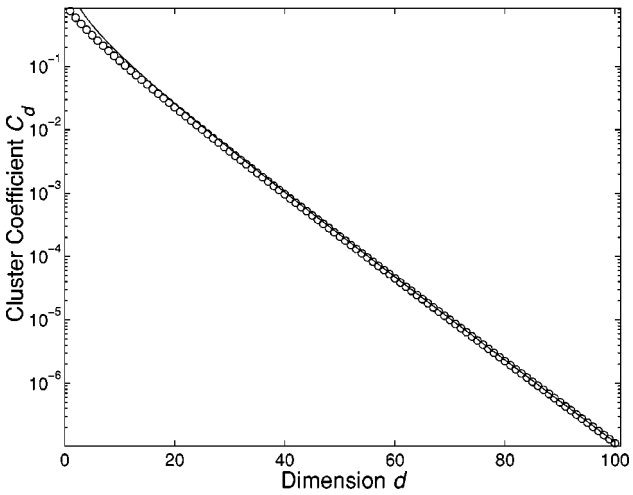

FIG. 8. The cluster coefficient C in random geometric graphs. The full line is the asymptotic solution [Eq. (13)] valid for large d only.

To determine C_d we make use of concept of the excluded volume V_{ex}. If we again use the vertices i, j, and k, then i and j must both be within the excluded volume of k. Put differently, the probability that i and j are connected is equal to the probability that two randomly chosen points in a sphere of volume V_{ex} and radius R is less than a distance R apart. In other words, given the coordinates of vertex i the probability that there is an edge between i and j is equal to the fraction of the excluded volume of vertex i that lies inside the excluded volume of k. By averaging over all points in V_{ex}, we get the cluster coefficient C_d.

The task of calculating C_d is considerably simplified by the spherical symmetry of the problem. The fractional volume ''overlap'' ρ_d of two spheres only depends on the distance r between the centers and not on any angular parts, i.e., $\rho_d=\rho_d(r)$. In general, the cluster coefficient can therefore be written as

$$C_d=\frac{1}{V_{ex}}\int_{V_{ex}}\rho_d(r)dV. \tag{10}$$

In the Appendix, we derive that

$$C_d=\begin{cases}1-H_d(1) & \text{even } d\\ \frac{3}{2}-H_d(\frac{1}{2}) & \text{odd } d,\end{cases} \tag{11}$$

where

$$H_d(x)=\frac{1}{\sqrt{\pi}}\sum_{i=x}^{d/2}\frac{\Gamma(i)}{\Gamma\left(i+\frac{1}{2}\right)}\left(\frac{3}{4}\right)^{i+1/2}. \tag{12}$$

When d is large, Eq. (11) reduces to (see the Appendix)

$$C_d\sim 3\sqrt{\frac{2}{\pi d}}\left(\frac{3}{4}\right)^{(d+1)/2}. \tag{13}$$

The cluster coefficient is plotted in Fig. 8 ($\bigcirc$) together with

the asymptotic solution in Eq. (13) (full line).

Equation (11) shows that the cluster coefficient is a purely geometric quantity depending only on the dimension d; neither the connectivity α nor the system size N are present. In random graphs $C=\alpha/N$, since there is per definition no correlation between edges. So, in contrast to what is usually believed, RGGs are *not* identical to random graphs when $d\to\infty$.

In higher dimensions, the cluster coefficient in RGGs becomes exceedingly small. This peculiar fact can be explained by noting that the distribution of distances between two connected vertices gets more and more peaked at the maximal distance R as d increases. This implies that if the vertices i and j are both connected to vertex k in a high-dimensional space, then it is highly unlikely that i and j are directly connected by an edge as well. Only in low dimensions are RGGs dominated by small loops. On the contrary, the way that a standard random graph is designed implies a cluster coefficient that can only be interpreted statistically, and not geometrically. Despite the fact that $\alpha_c=1$ in both random graphs and RGGs of infinite dimensionality, they do not have the same topology.

E. Graph bipartitioning

Random geometric graphs are useful outside network modeling and percolation theory as well. In this section we look at RGGs in relation to graph bipartitioning, a well-known problem in combinatorial optimization.

The NP-hard problem of partitioning a graph with N vertices in two subsets with $N/2$ vertices each, in such a way that the cutsize E, i.e., the number of edges between vertices in different subsets, is minimized, is called the graph bipartitioning (GBP) problem. Figure 2 illustrates a bipartitioned RGG, where $N/2$ of the points are marked by squares, the other half being dots.

The GBP problem of RGGs with open boundary conditions has been tested by various heuristics [41–43]. In this section we use our numerical findings to establish the critical connectivity in relation to GBP. Additionally, for $\alpha>\alpha_c^{GBP}$ we argue that the cutsize E depends on N and α in a simple way.

In GBP the connectivity is critical when $G=1/2$. As soon as the largest cluster contains more than half of the vertices, it becomes impossible to bipartition the graph without violating any edges. For random graphs Eq. (3) immediately gives us $\alpha_c^{GBP}=2\ln 2\simeq 1.386$.

In RGGs $\alpha_c^{GBP}(d)$ can be extracted in the same way as α_c was in Sec. IV A. Our numerical findings in RGGs with continuous boundary conditions are presented in Table II. We stress that the results are valid only for large N, as a closer look at Fig. 5 reveals. In 2D the average fraction of vertices in the largest cluster is independent of N only for $\alpha>\alpha_c^{GBP}$. This means that if one looks at GBP in 2D with $N=1000$, one cannot use the value of α_c^{GBP} in Table II. In higher dimensions the interval around α_c where $G_d(\alpha)$ is size dependent gets smaller and does not play a role in relation to GBP.

TABLE II. The critical connectivity α_c^{GBP} in random geometric graphs with toroidal boundary conditions. Only in 2D does α_c^{GBP} depend noticeably on N for $N>1000$ (see Fig. 5). Note that without continuous boundaries Fig. 4 shows that α_c^{GBP} is highly size dependent for $d>2$. The estimated errors in α_c^{GBP} in the last row are on the safe side.

d	2	3	4	5
α_c^{GBP}	4.52	2.84	2.275	1.99
$\pm$	0.02	0.01	0.005	0.005

With open boundary conditions the picture is messy, as Fig. 4 shows. In this case $G(\alpha)$ is highly N dependent, and it is not possible to speak of a critical connectivity α_c^{GBP} without specifying N. This is true despite the fact that $G(\alpha)$ is an averaged quantity, i.e., for small N will a fraction of the graphs contain a cluster with more than $N/2$ vertices even when $\alpha<\alpha_c^{GBP}$. Figure 4 clearly shows that α_c^{GBP} is a decreasing function of N for $d>2$. In 2D, however, all curves cross at almost the same (pivotal) point, and it is reasonable to speak of α_c^{GBP} without specifying N. As the inset in Fig. 4 shows this would lead to an estimate of $\alpha_c^{GBP}=4.53(1)$, close to α_c^{GBP} in RGGs with toroidal boundary conditions.

The size of the largest cluster near α_c grows so rapidly in 2D that $\alpha_c=\alpha_c^{GBP}$ cannot be ruled out on the basis of our numerical data. This is true with both open and continuous boundary conditions. However, as this would imply that the phase transition is of first order in 2D only, we believe that the two critical connectivities are close but not identical.

When bipartitioning a RGG, it is obvious that the ''area of contact'' [44] between the two subsets in the optimal configuration must be close to a minimum. In 2D this means that the best achievable partition must be close to simply cutting the graph into two at the coordinate values $x_1=1/2$ or $x_2=1/2$. This observation is especially relevant for large connectivities where the cutsize is, fluctuations neglected, proportional to the length of the dividing line. All this tentatively indicates how the cutsize E in GBP behaves as a function of N and α by looking at RGGs partitioned at $x_i=1/2$, where $1\leqslant i\leqslant d$. As we are about to argue, we expect a scaling relation like [45,46]

$$E_d\propto N^{1/\nu}\alpha^{\beta}(d), \tag{14}$$

where the exponents ν and β only depend on the dimension of the RGG.

The exponents in Eq. (14) can be determined in the following way. Given the radius R of the excluded volume of each vertex, the cutsize must be proportional to NR, since only vertices with $1/2-R<x_i<1/2$ contribute to the cutsize (to avoid counting the violated edges twice we only look at the vertices at one side of the partitioning plane at $x_i=1/2$) times the average number of violated edges per vertex in this region, which is proportional to NR^d. In other words,

$$E_d\propto N^2R^{d+1}. \tag{15}$$

If instead of R we want to express the result in terms of $\alpha(d) \propto NR^d$, we get

$$1/\nu = 1 - \frac{1}{d}, \quad \beta = 1 + \frac{1}{d}. \tag{16}$$

Since $E \propto N^2$ in Eq. (15), the relation $1/\nu + \beta = 2$ holds in arbitrary dimensions.

Now, it is obvious that the scaling ansatz is reasonable only for $\alpha > \alpha_c^{GBP}$. As Fig. 2 illustrates, the optimal partition at $\alpha \sim \alpha_c^{GBP}$ is highly complex and not at all close to a straight line. If we incorporate that $E=0$ for $\alpha < \alpha_c^{GBP}$ and replace Eq. (14) with

$$E_d \propto N^{1/\nu} [\alpha(d) - \alpha_c^{GBP}]^{\beta}, \tag{17}$$

we do not expect Eq. (16) to hold if we focus only on a region near the critical connectivity. By the use of extremal optimization, a heuristic that works particularly well near phase transitions in hard combinatorial problems, Boettcher and Percus [45,46] have found $\alpha_c^{GBP} \simeq 4.1$, $1/\nu \simeq 0.6$, and $\beta \simeq 1.4$ in 2D for $4 < \alpha < 6$, not far from our estimates in Eq. (16) valid for large connectivities. Note that the low estimate of α_c^{GBP} is expected; the algorithm does not always find the best partition, and some graphs with $\alpha < \alpha_c$ does have $E > 0$.

V. IMPLEMENTATION

The implementation is of major importance when studying random geometric graphs, since a straightforward check of all possible edges between the N points will result in unfeasible runtimes $O(N^2)$. We now outline how our program works and describe how to avoid runtimes $O(N^2)$.

The main idea is to divide and conquer. Partition the d-dimensional box in smaller sub-boxes and determine which sub-box each vertex belongs to. Given the connectivity and thereby the radius R of the excluded volume, for each vertex we then only have to look for potential edges to vertices in the sub-boxes adjacent to the sub-box where the vertex itself is located. This leads to a huge reduction in the number of comparisons. And this just gets better when N increases, resulting in a decrease in R as we saw in Fig. 3. By partitioning the box further as N increases we avoid a linear increase in the number of comparisons per vertex, which would lead to the undesirable $O(N^2)$ growth.

The algorithm used when looking at RGGs is simple. It works like this: (1) generate d coordinates for each vertex; (2) partition the space in small sub-boxes; (3) find the edges; (4) calculate the relevant quantities (G, cluster sizes etc.) as α increases. Obviously, a trade-off in step 2 is involved when choosing the number of small boxes.

Being the most time consuming part of the algorithm, step 3 is the main contributor when deciding how the runtime depends on N. The runtimes for most of our runs are shown in Fig. 9. We see that the runtime is $O(N^{\beta})$, where $\beta \approx 1.3$, resulting in ''feasible'' runtimes for graphs with $N > 4 \times 10^6$. Note that the runtime of the much simpler algorithm used on random graphs also grows like a power law with $\beta = 1.15$,

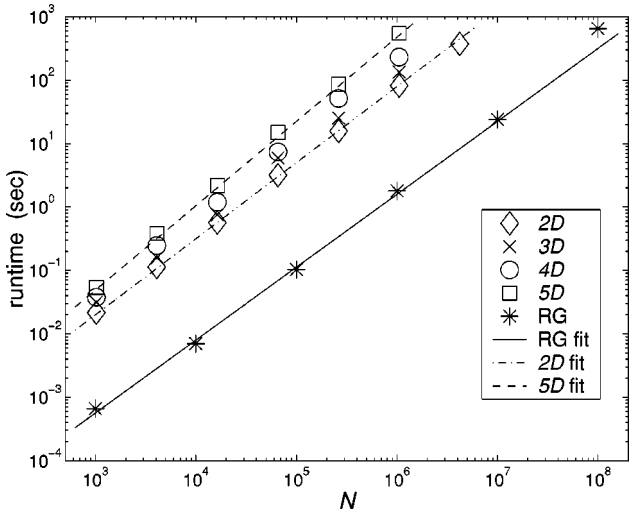

FIG. 9. The runtimes (on a 400 MHz SUN) of the algorithms used in Secs. II and IV. The straight lines indicate $t \sim N^{\beta}$, where $\beta = 1.2$ in 2D, $\beta = 1.33$ in 5D, and $\beta = 1.15$ in random graphs (RG).

even though the number of operations is clearly $O(N)$. In fact, the number of comparisons with potential neighbors per vertex is very nearly constant in our implementation, i.e., the total number of neighbor tests is $O(N)$ in RGGs as well. Of course, this is only possible if the number of sub-boxes also increases with N. Managing the partitioning part of the algorithm adds to the runtime. To sum up, the power-law increase in the runtime illustrated in Fig. 9 for both random graphs and RGGs is probably mainly due to cache misses. The slightly higher values of β in the RGGs stems from the additional time used when partitioning the d-dimensional box into smaller boxes.

Step 4 is worth a comment. When running the algorithm, we are interested in information at certain values of α. Instead of generating a new graph for every data point needed, we first set up the graph with the minimal connectivity we want to look at. This is easily accomplished with our algorithm. Given an α window $[\alpha_{min}, \alpha_{max}]$ in which we want to examine the graph, we find all the edges belonging to the graph when $\alpha = \alpha_{max}$, but we only add the edges corresponding to $\alpha = \alpha_{min}$. The rest of the edges, those who are to be added when α is gradually increased to α_{max}, are stored in a priority queue. It is then a simple task to increase α as one wishes. As mentioned earlier, in Figs. 4 and 5 each curve is based upon 300 data points, i.e., $\Delta\alpha = 0.005$.

The source code, written in C, is available upon request. For a more accurate and technical discussion of fast algorithms in relation to RGGs, see e.g. Ref. [47].

VI. SUMMARY

In this paper, we have illustrated the usefulness of random geometric graphs in network theory and how to implement them efficiently. Several properties of random geometric graphs in the vicinity of the critical connectivity α_c have been analyzed. We have determined the size of the largest cluster numerically and shown that $\alpha_c(d)$ approaches $\alpha_c(\infty) = 1$ found in random graphs in a power-law fashion.

We have verified that the distribution of cluster sizes is cut into two just when the connectivity becomes larger than α_c. Interestingly, the derivation of the cluster coefficient shows that, even in the limit of infinite dimensionality d, random geometric graphs are not identical to random graphs.

Random geometric graphs share properties with both lattice models and standard random graphs. Random geometric graphs allow us to work with random graphs with a local structure. In addition, it is straightforward to add ''long'' edges if one wishes to simulate, e.g., a small world network. With all this in mind, we hope this paper will make random geometric graphs more widely used in network theory.

ACKNOWLEDGMENTS

We thank J. Neil Bearden, Stefan Boettcher, and Paolo Sibani for helpful comments. We are especially indebted to Allon Percus for very valuable discussions and correspondence. This work was financially supported by Statens Naturvidenskabelige Forskningsråd.

APPENDIX: DERIVATION OF C_d

In order to determine the cluster coefficient for arbitrary d, one must find the fractional overlap ρ_d. Since ρ_d has no angular dependence, Eq. (10) reduces to

$$C_d = \frac{d}{R^d}\int_0^R \rho_d(r) r^{d-1} dr. \tag{A1}$$

Since $\rho_1 = 1-(r/2R)$, $C_1 = \frac{3}{4}$. From Fig. 10 we see that in 2D the overlapping area—the area circumscribed by the fat lines—is $2(A-B)$, where A is the area of the part of the circle swept out by the angle $\theta = 2\ \arccos(r/2R)$ between the two dashed lines originating from the center of the lowest circle, and B is the area of the dashed triangle. Now, $A = \frac{1}{2}\theta R^2$ and $B = R^2\ \cos(\theta/2)\sin(\theta/2) = \frac{1}{2}R^2\ \sin\theta$. The area of the overlap is then $R^2(\theta - \sin\theta)$, so $\rho_2 = 1(\theta - \sin\theta)/\pi$ and $C_2 = 1 - (3\sqrt{3}/4\pi)$.

For $d \geqslant 3$, the use of cylindrical coordinates and the relation

$$2\pi \prod_{i=2}^{n-1} \int_0^{\pi} \sin^{n-i}\theta_i d\theta_i = \frac{n\pi^{n/2}}{\Gamma\left(\frac{n+2}{2}\right)} \tag{A2}$$

results in

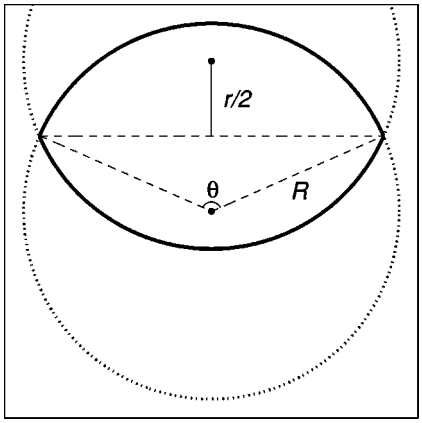

FIG. 10. Determination of the cluster coefficient C, which in 2D is equal to the average fractional area overlap of the two circles. R is the radius of the circles and r the distance between their centers. The area of the overlap is confined within the fat arcs originating from the two circles (dotted). The dashed lines are helpful in the derivation of the overlap—see the text.

$$\rho_d(r) = \frac{2}{\sqrt{\pi}} \frac{\Gamma\left(\frac{d+2}{2}\right)}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^{\arccos(r/2R)} \sin^d\theta d\theta. \tag{A3}$$

By reversing the integration in C_d, we get

$$C_d = \frac{3}{\sqrt{\pi}} \frac{\Gamma\left(\frac{d+2}{2}\right)}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^{\pi/3} \sin^d\theta d\theta, \tag{A4}$$

which can be solved by integration by parts. The use of the duplicate formula for the Γ function then finally leads to Eq. (11).

For large d, the ratio of the Γ functions in Eq. (A4) is given by Stirling's approximation. By putting $x = \cos\theta - 1/2$, the cluster coefficient can therefore be written as

$$C_d \simeq \sqrt{\frac{6d}{\pi}}\left(\frac{3}{4}\right)^{d/2} \int_0^{1/2} \exp\left[\frac{d-1}{2}\ln f(x)\right] dx, \tag{A5}$$

where $f(x) = 1 - [4x(1+x)/3]$. Since the contributions to the integral for large d are significant only when $x \simeq 0$, $\ln f$ can be expanded to first order and Eq. (13) is recovered.

[1] Albert-László Barabási, and Réka Albert, e-print cond-mat/0106096v1.
[2] Steven H. Strogatz, Nature (London) **410**, 268 (2001).
[3] Bernardo A. Huberman, Nature (London) **401**, 131 (1999).
[4] M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. E **64**, 026118 (2001).
[5] Réka Albert and Albert-László Barabási, Science **286**, 509 (1999).
[6] M.E.J. Newman, Phys. Rev. E **64**, 025102 (2001).
[7] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[8] Duncan J. Watts and Steven H. Strogatz, Nature (London) **393**, 440 (1998).
[9] Stuart Kauffman, *At Home in the Universe* (Oxford University Press, Oxford, 1995).
[10] P. Erdös and A. Rényi, Publ. Math. **5**, 17 (1960).

[11] Duncan S. Callaway, John E. Hopcroft, Jon M. Kleinberg, M.E.J. Newman, and Steven H. Strogatz, Phys. Rev. E **64**, 041902 (2001).
[12] Béla Bollobás, *Random Graphs* (Academic Press, New York, 1985).
[13] Scott Kirkpatrick and Bart Selman, Science **264**, 1297 (1994).
[14] Jon M. Kleinberg, Nature (London) **406**, 845 (2000).
[15] D. Stauffer and A. Ahorony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1992).
[16] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[17] Réka Albert, Hawoong Jeong, and Albert-László Barabási, Nature (London) **406**, 378 (2000).
[18] P.L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[19] Réka Albert and Albert-László Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[20] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **62**, 1842 (2000).
[21] D.J. Watts, *Small Worlds* (Princeton University Press, Princeton, 1999).
[22] M.E.J. Newman and D.J. Watts, Phys. Rev. E **60**, 7332 (1999).
[23] M.E.J. Newman, J. Stat. Phys. **101**, 819 (2000).
[24] Jayanth R. Banavar, Amos Maritan, and Andrea Rinaldo, Nature (London) **399**, 130 (1999).
[25] W. Xia and M.F. Thorpe, Phys. Rev. A, **38**, 2650 (1988).
[26] I. Balberg, Phys. Rev. B **31**, 4053 (1985).
[27] U. Alon, A. Drory, and I. Balberg, Phys. Rev. A **42**, 4634 (1990).
[28] U. Alon, I. Balberg, and A. Drory, Phys. Rev. Lett. **66**, 2879 (1991).
[29] J. Quantanilla, S. Torquato, and R.M. Ziff, J. Phys. A **33**, L399 (2000).
[30] I. Balberg, Philos. Mag. B **56**, 991 (1987).
[31] Philippe Jund, Rémi Jullien, and Ian Campbell, Phys. Rev. E **63**, 036131 (2001).
[32] Romualdo Pastor-Satorras and Alessandro Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[33] Peter Cheeseman, Bob Kanefsky, and William M. Taylor, Proceedings of the Twelfth International Joint Conference on Artificial Intelligence, IJCAI-91, Sidney, Australia, pp. 331–337.
[34] Philip W. Anderson, Nature (London) **400**, 115 (1999).
[35] K.Y.M. Wong and D. Sherringham, J. Phys. A **20**, L793 (1987).
[36] Michael Molloy and Bruce Reed, Random Struct. Algorithms **6**, 161 (1995).
[37] Michael Molloy and Bruce Reed, Combinatorics, Probab. Comput. **7**, 295 (1998).
[38] M.D. Rintoul and S. Torquato, J. Phys. A **30**, L585 (1997).
[39] Edward T. Gawlinski and H. Eugene Stanley, J. Phys. A **14**, L291 (1981).
[40] Salvatore Torquato, *Random Heterogenous Materials: Microstructure and Macroscopic Properties* (Springer, New York, 2002).
[41] David S. Johnson *et al.*, Oper. Res. **37**, 865 (1989).
[42] Peter Merz and Bernd Freisleben, Lect. Notes Comput. Sci. **1498**, 765 (1998).
[43] Stefan Boettcher and Allon G. Percus, Artif. Intel. **64**, 275 (2000).
[44] Stefan Boettcher and Allon G. Percus, Phys. Rev. E **64**, 026114 (2001).
[45] Stefan Boettcher and Allon G. Percus, in *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*, (Morgan Kaufmann, San Francisco, 1999), pp. 825–832.
[46] Stefan Boettcher, J. Phys. A **32**, 5201 (1999).
[47] Matthew T. Dickerson and David Eppstein, Comp. Geom. Theory Appl. **5**, 277 (1996).