Distance Graphs: From Random Geometric Graphs to Bernoulli Graphs and Between

UCLA CSD TR 060011

Chen Avin

Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095-1596, USA.
avin@cs.ucla.edu

Abstract. A random geometric graph $\mathcal{G}(n,r)$ is a graph resulting from placing n points uniformly at random on the unit area disk, and connecting two points iff their Euclidean distance is at most the *radius* $r(n)$. Recently, this class of graphs have received much attention as a model for wireless networks. The *Bernoulli* graph $\mathcal{B}(n,p)$ is a random graph in which each edge is chosen independently with *edge probability* $p(n)$. The critical parameter for connectivity played a major role in the study of both $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$, and in what may seem surprising, it has been shown that both graphs have closely related critical connectivity thresholds for the radius and the edge probability. In particular, if $\pi r^2 = p = \frac{\log n + \gamma_n}{n}$ then both graphs are connected *w.h.p.* iff $\gamma_n \to +\infty$ and disconnected *w.h.p.* iff $\gamma_n \to -\infty$. To explain the similarities in the connectivity thresholds, we introduce an extension of the random geometric graphs: the *random distance graph*, $\mathcal{D}(n,g)$: A graph resulting from placing n points uniformly at random on the unit disk and connecting every two points independently with probability $g(d)$, where d is the distance between the nodes and g is the *connection function*. We give a connection function $g_r^\alpha(d)$ with parameters r and α such that: When $\alpha = 1$, $\mathcal{D}(n, g_r^\alpha)$ is identical to $\mathcal{G}(n,r)$ and when $\alpha = \pi r^2$, $\mathcal{D}(n, g_r^\alpha)$ is identical to $\mathcal{B}(n,p)$. Using results from continuum percolation we are able to bound the connectivity threshold of $\mathcal{D}(n, g_r^\alpha)$ for $\pi r^2 \leq \alpha \leq 1$. We show that the connectivity is governed by the probability that a node is isolated which was previously known for $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$.
Note, however, that in the evolution stage of connectivity these two graphs differ significantly, $\mathcal{G}(n,r)$ has a large diameter and many small cliques while $\mathcal{B}(n,p)$ has a small diameter, and no small cliques. So, interestingly, neither $\mathcal{G}(n,r)$ nor $\mathcal{B}(n,p)$ are suitable to model social networks. On the contrary, we show that a typical case of random distance graph is adequate to model social networks and, in particular, is a "Small World" graph, capturing both high clustering and small average path length. As opposed to previous Small World models that rely on deterministic sub-structures to grantee connectivity, random distance graphs offer a completely randomized model with a proven connectivity threshold.

1 Introduction

A random geometric graph $\mathcal{G}(n,r)$ [1–4] is a graph resulting from placing n points uniformly at random on the unit disk[1] and connecting two points iff their Euclidean

[1] This work is concerned with the 2-dimenstional case

distance is at most the *radius* $r(n)$. These graphs have traditionally been studied in relation to subjects such as statistical physics and hypothesis testing, and have been used in the last few years as models for ad-hoc and sensor networks [4]. The *Bernoulli* graph $\mathcal{B}(n,p)$ (a.k.a. Erdős-Rényi graph) is a random graph with n nodes in which each edge (out of the $\binom{n}{2}$ possible edges) is chosen independently at random with an *edge probability* $p(n)$. These graphs were first offered by Gilbert [5] and have been thoroughly investigated since the seminal work of Erdős and Rényi [6, 7].

One of the main goals of studying random graphs is to elucidate the properties of what is called a *typical graph*: the graph we likely to see after setting the parameter of the model to a specific function of n (in our case setting $r(n)$ or $p(n)$ as functions of n) and leting n go to infinity. If a property Q exists with probability going to 1 as n goes to infinity (i.e. with high probability), we say that a typical graph has property Q. In this context we can also speak of the *evolution* of the typical graph: the way the properties of a typical graph change as the typical graph evolves when we increase the order of the functions we set the parameter to. The parameter threshold is called *critical* if Q exhibits a sharp threshold: the difference between the parameter's settings for which the property Q holds with high probability (*w.h.p.*) and the parameter's settings for which the property Q holds with low probability goes to zero as $n \to \infty$.

The critical parameter for connectivity, both for $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$, has been of special interest. In what may seem surprising, it has been shown that both graphs have closely related critical thresholds for the radius and the edge probability. In particular if $\pi r^2 = \frac{\log n + \gamma_n}{n}$ then $\mathcal{G}(n,r)$ is connected *w.h.p.* iff $\gamma_n \to +\infty$ and disconnected *w.h.p.* iff $\gamma_n \to -\infty$ [1, 2] and likewise, if $p = \frac{\log n + \gamma_n}{n}$ then $\mathcal{B}(n,p)$ is connected *w.h.p.* iff $\gamma_n \to +\infty$ and disconnected *w.h.p.* iff $\gamma_n \to -\infty$ [6, 7]. This gives rise to the natural question of how to explain the similarities in the connectivity threshold between these two graphs? Recently, more results on random geometric graphs have been coupled to known results on Bernoulli graphs, revealing some interesting similarities around the critical connectivity thresholds, for example both graphs have sharp threshold for all monotone properties [8] and both graphs have optimal cover time near the connectivity threshold [9, 10].

Note, however, that these two graphs have quite different characteristics: in $\mathcal{B}(n,p)$ nodes appear to be only "place holders" for random edges selection, while in $\mathcal{G}(n,r)$ nodes are assigned some properties (i.e. coordinates x and y), on which the existence of edges depends. This leads to major differences in structure. For example, in the connectivity regime, $\mathcal{B}(n,p)$ has small diameter, and no small cliques while $\mathcal{G}(n,r)$ has large diameter and many small cliques. Another important distinction between these graphs concerns the conditional probability that given the existence of edges $k \sim i$ and $k \sim j$ there is also an edge $i \sim j$. (i.e. $P(i \sim j \mid k \sim i, k \sim i)$.) In $\mathcal{B}(n,p)$ these events are independent so $P(i \sim j \mid i \sim k, j \sim k) = P(i \sim j)$, while in $\mathcal{G}(n,r)$ the two events are not independent and $P(i \sim j \mid i \sim k, j \sim k) \gg P(i \sim j)$.

This lack of independence, also called "locality", is one of two important properties of social networks, and in particular, *Small World* graphs [11]. The locality property is measured by the *clustering coefficient* [11]: a number between 0 and 1 that reflects the fraction of a vertex's neighbors which are neighbors themselves. This reflects the observation that in social networks people that have a common friend are more likely to be friends with each others than two people chosen at random from the population; this intuition is confirmed by a wide range of real life data [11, 12]. The other major property of Small World graphs is the existence of a small average path length between nodes. (and from an algorithmic perspective, the ability to find such a path locally [13].) Interestingly, in the evolution stage of connectivity $\mathcal{G}(n, r)$ displays locality but not small path length, and $\mathcal{B}(n, p)$ displays small path length but not locality, so neither graphs are adequate to model social networks.

This motivates our investigation of whether a more general class of random graph exist, one that contains both $\mathcal{G}(n, r)$ and $\mathcal{B}(n, p)$ as special pathological cases, but whose typical case, in the evolution stage of connectivity, is adequate to model social networks. That is a class of graphs that present both locality and small average path length.

2 Statement of Results

Let $\mathcal{U}$ be the unit area disk centered at the origin of the $\mathbb{R}^2$ plane.

Definition 1 (Random Distance Graph) *For n nodes and a function $g : [0, \frac{2}{\sqrt{\pi}}] \to [0, 1]$ let $\mathcal{D}(n, g)$ be a random distance graph resulting from the following process. First place n nodes uniformly at random in the unit disk $\mathcal{U}$. Second, for each pair of nodes i, j with distance $d(i, j)$ place an edge between i and j independently from all other edges with probability $p_{ij} = g(d(i, j))$.*

Intuitively, we add edges between nodes as a function of their distance and therefore the underlying structure of the graph depends on the connection function g. Here we primarily consider a specific type of function, a step function g_r^α, with parameters r and α. This function creates edges with probability $\alpha = \alpha(n)$ for nodes at distance less or equal to a radius $r = r(n)$ (short edges), and with probability $\beta = \beta(n)$ for nodes at distance larger than r (long edges). Moreover we choose $\beta(\alpha)$ as a function of α, and in order to keep the average degree of the graph invariant with respect to α, we require:

$$(1 - \alpha)\pi r^2 = \beta(1 - \pi r^2) \tag{1}$$

Formally, for $\pi r^2 \leq \alpha \leq 1$ the function g_r^α is defined as follow:

$$g_r^\alpha(d) = \begin{cases} \alpha & \text{if } d \leq r, \\ \beta = \frac{(1-\alpha)\pi r^2}{1-\pi r^2} & \text{if } d > r. \end{cases} \tag{2}$$

For $\pi r^2 \leq \alpha \leq 1$, $\mathcal{D}(n, g_r^\alpha)$ is an extension of random geometric graphs that can capture both $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$. On one hand when $\alpha = 1$, $\beta = 0$ and we have $\mathcal{D}(n, g_r^1) \equiv \mathcal{G}(n,r)$, a random geometric graph[2]. On the other hand when $\alpha = \pi r^2$, from (1), we get that $\beta = \pi r^2$. Since $\alpha = \beta$ edges do not depend any more on the distance and every edge is chosen independently with probability $p = \alpha$ so $\mathcal{D}(n, g_r^{\pi r^2}) \equiv \mathcal{B}(n, p = \pi r^2)$ which is the original random Bernoulli graph. Note that both these graphs, as well as $\mathcal{D}(n, g_r^\alpha)$ for any $\pi r^2 \leq \alpha \leq 1$, have the same average degree, $\delta_{\text{avg}} = (n-1)\pi r^2$ and the same expected number of edges $m_{\text{avg}} = \binom{n}{2}\pi r^2$. When $r = \Theta(\frac{\log n}{n})$ then $\delta_{\text{avg}} = \Theta(\log n)$ and $m_{\text{avg}} = \Theta(n \log n)$. Our first result extends the sharp threshold for connectivity from $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$ to $\mathcal{D}(n, g_r^\alpha)$:

Theorem 1 (Connectivity) *Let* $\pi r^2 = \frac{\log n + \gamma_n}{n}$. *Then for* $\pi r^2 \leq \alpha \leq 1$, $\mathcal{D}(n, g_r^\alpha)$ *is connected* w.h.p. *iff* $\gamma_n \to \infty$ *and is disconnected* w.h.p. *iff* $\gamma_n \to -\infty$.

Next we can prove the following about clustering and the diameter of $\mathcal{D}(n, g_r^\alpha)$:

Theorem 2 (Clustering) *Let* $\pi r^2 = \frac{\log n + \gamma_n}{n}$ *and* $\gamma_n \to \infty$. *For* $\pi r^2 \leq \alpha \leq 1$ *the clustering coefficient of* $\mathcal{D}(n, g_r^\alpha)$ *represented as* $C = P(i \sim j \mid k \sim i, k \sim j)$ *is* w.h.p. $C = \alpha * c' + o(1)$ *where* $c' \approx 0.5865$.

Theorem 3 (Diameter) *Let* $\pi r^2 = \frac{\log n + \gamma_n}{n}$ *and* $\gamma_n \to \infty$. *For a constant* $\epsilon > 0$ *and* $\pi r^2 \leq \alpha \leq 1 - \epsilon$ *the diameter of* $\mathcal{D}(n, g_r^\alpha)$ *is* w.h.p. $\Theta(\frac{\log n}{\log \log n})$.

Another important perspective of Small World graphs mentioned earlier is the algorithmic one [13]. From this perspective, motivated by the original Milgram experiment [14], social network not only have a short average path length, but enable for such a path to be found locally in a distributed manner, namely by *local routing*. Local routing is a mechanism for which a message is sent from a source to a destination using only local information available at each node. The destination's location is known (for example in the message header), and each node can only forward a message to one of its immediate neighbors based on their locations, thus we can route the message in a distributed way without global knowledge. We are interested in a graph that permits *short* local routing. Namely, the expected route length (number of steps) from source to destination is on the same order as the graph diameter. Finally we can show the following:

Theorem 4 (Local Routing) *There exist a random distance graph* $\mathcal{D}(n, g)$ *that has the following properties with high probabilty: It is connected, has average degree* $\Theta(\log n)$, *high clustering, small diameter and allows short local routing.*

[2] where $\equiv$ stand for the same generating process.

3 Related Work

This work was influenced by a large variety of recent work on random geometric graphs, and in particular work that exposed the similarities to the Bernoulli model, in connectivity [1, 2], monotone properties [8] and the cover time of random walk [10]. The origin of distance graphs goes back to the work of Gilbert [15] and later to the random connection model used in *continuum percolation* by Penrose and others [4, 16]. Those models where concern with a Poisson process on the entire plane while the random distance graph is define on the unit disk and with a uniform point distibution. We are, however, not aware of previous work that explores the connection between $\mathcal{B}(n,p)$ and $\mathcal{G}(n,r)$ by showing that the integral over the connection function $g(x)$ is the same for both graphs. In [2] Gupta and Kumar conjectured that if $\pi r^2 p(n) = \frac{\log n + \gamma_n}{n}$ the same results on connectivity holds, but this seems to be harder case since it increases the border effect. (their proof for connectivity of $\mathcal{G}(n,r)$, as well as ours, depends on the negligibility of the effect of nodes that are less than r away from the border.)

In [13] Kleinberg first propose the algorithmic perspective of the Milligram experiment and offered a grid-based distance graph that support local routing. In his model, as well as in the original Small World model of Watts and Strogatz [11], the graph starts from a deterministic connected graphs (i.e ring or grid) and random edges are only later introduced or rewired. In Watts and Strogatz model the initial graph is a connected ring of clusters, when edges are rewired the graph become a Small World, at the extreme enough edges are rewired to result in a Bernoulli graph. Random Intersection Graphs (RIG)[17] are another model of random graphs motivated by social structure, in particular by a collaboration networks such as authors-papers or actors-movies, but it does not have any geometric flavor. Under certain parameters values this model is also identical to $\mathcal{B}(n,p)$ [18] and under others differs significantly. Unfortunately, across its parameters range it has only two asymptotic clustering coefficient values, either $o(1)$ or 1-o(1).

4 Proofs

Recall that $\mathcal{U}$ is the unit area disk at the origin. Let $\bar{\mathcal{U}}$ be the disk of radius $\frac{1}{\pi} - r$ and let $\Delta\mathcal{U}$ be $\mathcal{U} \setminus \bar{\mathcal{U}}$. Let $disk_r(i)$ be the disk of radius r centered at X_i, and $lune_r(i,j)$ be the intersection of $disk_r(i)$ and $disk_r(j)$. (we omit r where there is no confusion.) If, for nodes uniformly distributed in $\mathcal{U}$, it is given that a node $i \in disk(j)$ then i is uniformly distributed in $disk(j)$ and if it is given that $i \notin disk(j)$ then i is uniformly distributed in $\mathcal{U} \setminus disk(j)$.

4.1 Proof of Theorem 1 (Connectivity)

Proof. We make use of results from *continuum percolation*, in particular a model that was first introduced by Gilbert [15] and later analyzed rigorously by Penrose [19]. Here we

concern ourselves only with the two dimensional case; Let $g(x),\ x \in \mathbb{R}^2$, be a measurable function taking values in $[0,1]$ such that

$$g(x) = g(|x|), \quad x \in \mathbb{R}^2 \tag{3}$$

$$0 < \int_{\mathbb{R}^2} g(x)dx < \infty \tag{4}$$

Let $\mathscr{P}$ be a homogeneous Poisson process on $\mathbb{R}^2$ with rate ρ: the expected number of point in any region is equal to the area of the region multiplied by ρ. Let $\{X_1, X_2, X_3 \dots\}$ be the set of points placed by $\mathscr{P}$ in $\mathbb{R}^2$. In addition, a point $X_0 = 0$ is added and when considering this point at the origin it is assumed to be an "arbitrary point of the Poisson process" [19].

Let $\mathcal{P}(\rho, g)$ denote the following random graph: Given a generalization of $\mathscr{P} \cup X_0$ we connect every two points $\{X_i, X_j\}$ $i \neq j$ with probability $g(d(X_i, X_j))$, independently of any other pair of points. The connected components of $\mathcal{P}(\rho, g)$ are called *clusters*, and let $C(0)$ be the "cluster at the origin", the set of points that have a path to X_0 in $\mathcal{P}(\rho, g)$. Let $card(C(0))$ be the cardinality of $C(0)$. For a given $g(x)$, $\mathcal{P}(\rho, g)$ and an integer k, let $q_k(\rho)$ denote the probability that $C(0)$ has k points, i.e. the probability that $card(C(0)) = k$

We make use of the following results:

I. **Probability of Isolation** (Theorem 3 [19]): If g enclose zero (essentially g is symmetric, g has bounded support, and g is bounded away from zero in some open neighborhood of the surface, see
[19]; all the functions considered in this paper encloses zero) then for large ρ the probability of the origin to be in a size one cluster and the probability of the origin to be in any finite size cluster are the same:

$$\frac{P_\rho(card(c(0)) < \infty)}{P_\rho(card(C(0)) = 1)} = \frac{1}{q_1(\rho)} \sum_{k=1}^{\infty} q_k(\rho) \to 1 \quad \text{as} \quad \rho \to \infty \tag{5}$$

In other words, for large ρ, *w.h.p.* the point at the origin is either isolated or connected to an infinite cluster.

II. **Uniqueness Theorem** (Theorem 6.3 [20]): *w.h.p.* there is at most one infinite cluster in $\mathcal{P}(\rho, g)$.

Putting these results together we get that in order to bound the probability of $\mathcal{P}(\rho, g)$ to be connected (for an appropriate g), it is sufficient to bound $q_1(\rho)$, the probability that the point at the origin is isolated, i.e. $card(C(0)) = 1$. In [19] it has been shown that

$$q_1(\rho) = P_\rho[card(C(0)) = 1] = \exp\left(-\rho \int_{\mathbb{R}^d} g(x)dx)\right) \tag{6}$$

If we let $g(x) = g_r^\alpha$ and $\rho = n$, then we can bound the probability of isolation at the origin:

$$\begin{aligned}\int_{\mathcal{U}} g_r^\alpha(x)dx &= \alpha(\pi r^2) + \beta(1-\pi r^2) \\ &= \alpha(\pi r^2) + \frac{(1-\alpha)\pi r^2}{1-\pi r^2}(1-\pi r^2) \\ &= \pi r^2 \\ &= \frac{\log n + \gamma_n}{n}\end{aligned}$$

So from (6) the probability of the origin to be isolated is:

$$\begin{aligned}q_1(n) &= \exp\left(-n\int_{\mathbb{R}^2} g(x)dx\right) \\ &= \exp\left(-n\frac{\log n + \gamma_n}{n}\right) \\ &= \exp(-\log n - \gamma_n) \\ &= \frac{1}{n}e^{-\gamma_n}\end{aligned}$$

Let $E^1(G)$ be the expected number of order 1 components in $\mathcal{U}$ for random graph G and $P^1(G)$ the probability that there is at least one order 1 component in $\mathcal{U}$ for G. Since we are still (not for long) considering $\mathbb{R}^2$ any other point in $\mathcal{U}$ has the same probably to be isolated and we have

$$P^1(\mathcal{P}(n,g)) \leq E^1(\mathcal{P}(n,g)) = nq_1(n) = e^{-\gamma_n} \tag{7}$$

When $\lim_{n\to\infty}\gamma_n = \infty$ then $P^1(\mathcal{P}(n,g)) \to 0$ and the graph is connected with high probability since every point is in the unique infinite cluster. The problem is that this result relays on the fact that $\mathcal{P}(n,g)$ is defined over $\mathbb{R}^2$ and each point in $\mathcal{U}$ is isomorphic to the origin. In our case we are concerned with a process that throws points only inside the unit disk, and the problem of the border's effect arise.

For a connection function $g(x)$ let $\mathcal{P}^{\mathcal{U}}(n,g)$ be the graph resulting from the Poisson process with density n on the unit disk $\mathcal{U}$ and $\mathcal{D}(n,g)$, as before, the graph resulting from uniformly distributing n points in $\mathcal{U}$. Let I be the indicator function and for a random graph $\mathcal{D}(j,g)$ with j nodes let $\mathcal{D}(j,g)_i$ stand for the event "X_i is isolated in $\mathcal{D}(j,g)$" (and similarly we have $\mathcal{G}(j,r)_j$ and $\mathcal{B}(j,p)_j$). By definition of the Poisson process and

expectation we have:

$$\begin{aligned} P^1(\mathcal{P}^{\mathcal{U}}(n,g)) &= \sum_{j=1}^{\infty} P^1(\mathcal{D}(j,g))e^{-n}\frac{n^j}{j!} \\ &\leq \sum_{j=1}^{\infty} E^1\big(\mathcal{D}(j,g)\big)e^{-n}\frac{n^j}{j!} \\ &= \sum_{j=1}^{\infty} E\left[\sum_{i=1}^{j} I_{\mathcal{D}(j,g)_i}\right]e^{-n}\frac{n^j}{j!} \\ &= \sum_{j=1}^{\infty} jP(\mathcal{D}(j,g)_j)e^{-n}\frac{n^j}{j!} \end{aligned}$$

In [2] Gupta and Kumar proved that for $\pi r^2 = \frac{\log n+\gamma_n}{n}$

$$P^1(\mathcal{P}^{\mathcal{U}}(n,g_r^1)) \leq \sum_{j=1}^{\infty} jP(\mathcal{D}(j,g_r^1)_j)e^{-n}\frac{n^j}{j!} \leq e^{-\gamma_n} \tag{8}$$

More over they bound the probability P_{diss} that $\mathcal{D}(n,g)$ is disconnected[3]. For $\epsilon > 0$

$$P_{diss}(\mathcal{D}(n,g)) \leq 2(1+6\epsilon)\left[P^1(\mathcal{P}^{\mathcal{U}}(n,g)) + \sum_{j=1}^{\infty} jP(\mathcal{D}(j,g)_j)e^{-n}\frac{n^j}{j!}\right] \tag{9}$$

In particular for $\pi r^2 = \frac{\log n+\gamma_n}{n}$ and $\lim_{n\to\infty}\gamma_n = c$, using (8) they proved:

$$\limsup_{n\to\infty} P_{diss}(\mathcal{G}(n,r)) = \limsup_{n\to\infty} P_{diss}(\mathcal{D}(n,g_r^1)) \leq 4e^{-c}$$

To prove our result we want to show that for $\pi r^2 \leq \alpha \leq 1$ and $1 \leq j$

$$\limsup_{n\to\infty} P_{diss}(\mathcal{D}(n,g_r^\alpha)) \leq 4e^{-c}$$

This will follow from (8) and (9) and by showing that for $\pi r^2 \leq \alpha \leq 1$

$$P(\mathcal{D}(j,g_r^1)_j) \geq P(\mathcal{D}(j,g_r^\alpha)_j) \geq P(\mathcal{D}(j,g_r^{\pi r^2})_j) \tag{10}$$

meaning that, for any α in the range, the probability that a node is isolated in $\mathcal{D}(n,g_r^\alpha)$ is smaller than it is in random geometric graph $\mathcal{G}(n,r)$ and larger than in random Bernoulli graph $\mathcal{B}(n,p=\pi r^2)$.

[3] They prove it for $\mathcal{G}(n,r)$, but the result follows for the general case.

There are two cases: (i) $X_j \in \bar{\mathcal{U}}$ and (ii) $X_j \in \Delta\mathcal{U}$. For the first case we have that for $\mathcal{G}(j,r)$ and $\mathcal{B}(j,\pi r^2)$ the probability that X_j is isolated is $(1-\pi r^2)^{j-1}$. For $\mathcal{D}(j,g_r^\alpha)$ we have:

$$\begin{aligned}
P(\mathcal{D}(j,g_r^\alpha)_j) &= ((1-\alpha)\pi r^2 + (1-\beta)(1-\pi r^2))^{j-1} \\
&= ((1-\alpha)\pi r^2 + (1-\frac{(1-\alpha)\pi r^2}{1-\pi r^2})(1-\pi r^2))^{j-1} \\
&= ((1-\alpha)\pi r^2 + (1-\pi r^2) - (1-\alpha)\pi r^2)^{j-1} \\
&= (1-\pi r^2)^{j-1}
\end{aligned}$$

For (ii) X_j is less than r away from the border of $\mathcal{U}$ and let $A = disk_r(X_j) \cap \mathcal{U} < \pi r^2$. Then for $\pi r^2 < \alpha < 1$ we have

$$\begin{aligned}
P(\mathcal{D}(j,g_r^\alpha)_j) &= ((1-\alpha)A + (1-\beta)(1-A))^{j-1} \\
&= ((1-\alpha)A + (1-A) - \frac{(1-\alpha)\pi r^2}{1-\pi r^2}(1-A))^{j-1}
\end{aligned}$$

When $\alpha = 1$ we get the case for $\mathcal{G}(j,r)$, $P(\mathcal{G}(j,r)_j) = (1-A)^{j-1}$ and when $\alpha = \pi r^2$ we get the case for $\mathcal{B}(j,\pi r^2)$, $P(\mathcal{B}(j,\pi r^2)_j) = (1-\pi r^2)^{j-1}$.

To confirm that (10) hold also for case (ii) it is sufficient to show that $\frac{\partial}{\partial\alpha}P(\mathcal{D}(j,g_r^\alpha)_j) \geq 0$, and in particular it is enough to show that

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\left((1-\alpha)A + (1-A) - \frac{(1-\alpha)\pi r^2}{1-\pi r^2}(1-A)\right) &= \frac{\pi r^2}{1-\pi r^2}(1-A) - A \\
&\geq \pi r^2 - A \geq 0
\end{aligned}$$

The second part of the theorem also follow from (10). Since it is known that for $\mathcal{B}(n,\pi r^2)$ and $\gamma_n \to -\infty$, $\mathcal{B}(n,\pi r^2)$ is disconnected *w.h.p.* , and in particular that there is an isolated node, from (10) the same will hold for $\mathcal{D}(n,g_r^\alpha)$ and $\alpha \geq \pi r^2$. □

4.2 Proof of Theorem 2 (Clustering)

Proof. When computing the conditional probability $P(i \sim j \mid k \sim i, k \sim j)$ we ignore the border effect (nodes in $\Delta\mathcal{U}$) since the fraction of these nodes is vanishing and their contribution to the final result goes to zero. We can consider three disjoint cases: (i) $i,j \in disk(k)$. (ii) one of i,j is in $disk(k)$ and the other is not. (iii) $i,j \notin disk(k)$. First let's define the following helping probabilities:

$$\begin{aligned}
p^* &= P(j \in disk(i) \mid j,i \in disk(k)) = P(j \in lune(k,i) \mid j,i \in disk(k)) \\
p^{**} &= P(j \in disk(i) \setminus lune(k,i) \mid j \notin disk(k) \wedge i \in disk(k))
\end{aligned}$$

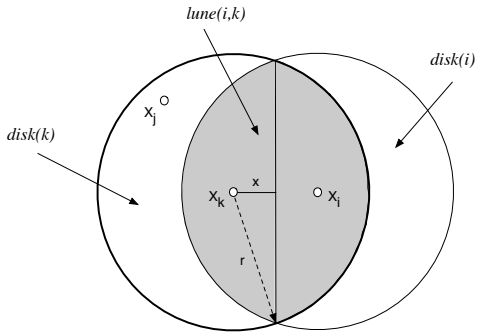

Fig. 1. Computing the conditional probability $P(i \sim j \mid k \sim i, k \sim j)$

Let $i \in disk(k)$ and let $y = 2x$ be the distance between k and i. Then $lune(k, i)$ is equal to twice the half lune (see Fig 1):

$$lune(k,i) = r^2\left(2\arccos\left(\tfrac{x}{r}\right) - \sin\left(2\arccos\left(\tfrac{x}{r}\right)\right)\right) \tag{11}$$

taking the integral over $ydy = 2x2dx$ we get:

$$\begin{aligned}
p^* &= \int_0^{r/2} \frac{2\pi 2x2}{\pi r^2} r^2 \frac{2\arccos(\frac{x}{r}) - \sin(2\arccos(\frac{x}{r}))}{\pi r^2} dx \\
&= \frac{8}{\pi r^2}\int_0^{r/2} x2\arccos(\frac{x}{r}) - \sin(2\arccos(\frac{x}{r}))dx \\
&= \frac{8}{\pi r^2}\left(\frac{-x(r^2+2x^2)\sqrt{1-\frac{x^2}{r^2}} + 4rx^2\arccos(\frac{x}{r}) + r^3\arcsin(\frac{x}{r})}{4r}\right)\Bigg|_0^{r/2} \\
&= \frac{2}{\pi r^3}\left(-\frac{r}{2}\left(r^2+\frac{r^2}{2}\right)\frac{\sqrt{3}}{2} + r^3\frac{\pi}{3} + r^3\frac{\pi}{6} - 0\right) \\
&= \frac{2}{\pi}\left(-\frac{\sqrt{3}3}{8} + \frac{\pi}{3} + \frac{\pi}{6}\right) \\
&\approx 0.5865
\end{aligned}$$

and

$$\begin{aligned}
p^{**} &= \int_0^{r/2} \frac{2\pi 2x2}{\pi r^2}\left(\frac{\pi r^2 - r^2(2\arccos(\frac{x}{r}) - \sin(2\arccos(\frac{x}{r}))}{1-\pi r^2}\right)dx \\
&= \frac{\pi x^2 2}{1-\pi r^2} - \frac{8}{1-\pi r^2}\left(\frac{-x(r^2+2x^2)\sqrt{1-\frac{x^2}{r^2}} + 4rx^2\arccos(\frac{x}{r}) + r^3\arcsin(\frac{x}{r})}{4r}\right)\Bigg|_0^{r/2} \\
&= \frac{2\pi r^2}{4(1-\pi r^2)} - \frac{2r^2}{1-\pi r^2}\left(\frac{\sqrt{3}3}{8} + \frac{\pi}{3} + \frac{\pi}{6}\right) \\
&= o(1)
\end{aligned}$$

where the last step is a result of $\pi r^2 = o(1)$. Now we can calculate the clustering coefficient as a function of α (note that $\beta = o(1)$ and $\pi r^2 = o(1)$):

$$\begin{aligned}
P(i\sim j \mid k\sim i, k\sim j) &= P(i\sim j \mid k\sim i, k\sim j, (\text{i})) + P(i\sim j \mid k\sim i, k\sim j, (\text{ii})) + P(i\sim j \mid k\sim i, k\sim j, (\text{iii})) \\
&= (\alpha p^* + \beta(1-p^*)) + (\alpha p^{**} + \beta(1-p^{**})) + (\alpha\pi r^2 + \beta(1-\pi r^2)) \\
&= \alpha p^* + o(1) \\
&\approx \alpha * 0.5865 + o(1)
\end{aligned}$$

□

4.3 Proof of Theorem 3 (Diameter)

Proof. It is well known that the diameter of a connected $\mathcal{B}(n,p)$ is $\frac{\log n}{\log np}$ and that the unique giant component of order n emerge when $p > \frac{1}{n}$ [21]. Since in our case $\alpha < 1-\epsilon$ is bounded away from 1 we get that $\beta \geq \frac{\epsilon\pi r^2}{1-\pi r^2} \geq \frac{\epsilon' \log n}{n}$ for a constant $\frac{\epsilon}{1-\pi r^2} \geq \epsilon' > 0$. The graph $\mathcal{D}(n, g_r^\alpha)$ can be thought of as being built in two phases: in the first phase we give each node its location and create $\mathcal{B}(n,p)$ with $p = \beta$, and in the second phase we add the rest of the short edges with the appropriate probability. (i.e. $\frac{\alpha-\beta}{1-\beta}$.) For the above $p = \beta$ we can say the following about $\mathcal{B}(n,p)$ and its gaint component[21]: Let $I(n)$ be the set of nodes *not* in the giant component, then the expected size of $I(n)$ is $E(|I(n)|) = \frac{x(c)}{2c}n$ where $\frac{x(c)}{2c}$ is the fraction of nodes not in the giant component, and:

$$x(c) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(2ce^{-2c})^k$$

where $c > 1/2$ is defined by $p = 2c/n$. In our case $c \geq \epsilon' \log n \to \infty$ so $x(c)/2c \to 0$. In particular

$$E(|I(n)|) = \frac{x(c)}{2c}n = O(n^{1-2\epsilon'})$$

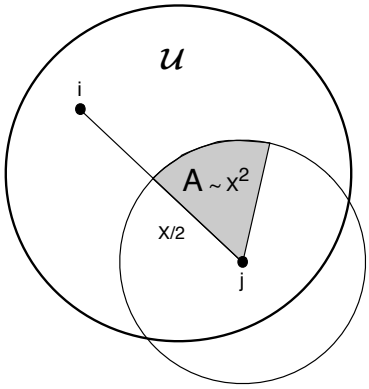

Fig. 2. an area that is proportional to x^2 when local routing from i to j with $x = d(i, j)$.

From [22] we have that for $\log n > np \to \infty$ the diameter of the giant component of $\mathcal{B}(n, p)$ is $(1 + o(1))\frac{\log n}{\log np}$. To prove the theorem we need to show that there is no path longer than $\frac{\log n}{\log np}$ in $\mathcal{D}(n, g_r^\alpha)$ from nodes not in the giant component of $\mathcal{B}(n, \beta)$. Since we already proved that $\mathcal{D}(n, g_r^\alpha)$ is connected the theorem follows. A necessary condition to have a path longer than $\frac{\log n}{\log np}$ is to have $\frac{\log n}{\log np}$ nodes from $I(n)$ in an area of size less than or equal to $(\frac{\log n}{\log np})\pi r^2 = \frac{\log^2 n}{n \log np}$. For $i \in I(n)$ let S_i be the number of nodes from $I(n)$ that are in an area of $\frac{\log^2 n}{n \log np}$ which i belongs to. Let $S = \max\{S_i \mid i \in I(n)\}$. Now since:

$$E(S_i) = E(|I(n)|)\frac{\log^2 n}{n \log np} = O\left(\frac{\log^2 n}{n^{-2\epsilon'} \log np}\right) \to 0 \tag{12}$$

it follows that $P(S > \frac{\log n}{\log np}) \to 0$ and so the diameter of $\mathcal{D}(n, g_r^\alpha)$ is of the same order as the diameter of the giant component of $\mathcal{B}(n, p = \beta)$. □

4.4 Proof of Theorem 4 (Local Routing)

Proof. The proof follows the same idea of Kleinberg [13], we give a connection function with the following guarantees (in expectation): after a constant number of steps the distance to the destination is reduced by a factor of two. We then modify the function to make sure the graph is also connected with average degree of $\Theta(\log n)$. Let $g'(x) = (x\sqrt{n\pi}+1)^{-2}$ for $x \in [0, \frac{2}{\pi}]$ and 0 otherwise, so it is defined up to the maximum distance on $\mathcal{U}$. Observe that for any two points $i, j \in \mathcal{U}$ where $x = d(i, j)$, there is an area A of at least $\pi(\frac{x}{2})^2/6$ s.t $A \subset \mathcal{U}$ and all the points in A are at most distance $x/2$ from j and at most distance x from i. The expected number of nodes (for large enough A)[4] is $\Theta(nA)$.

[4] A is large since for $x < r \log n$ we can route to destination, without using long edges, in $\Theta(\log n)$ steps.

Consider a message at i on its way to j, the probability that i will have a long range neighbor in A is at least

$$n\frac{\pi(\frac{x}{2})^2/6}{(x\sqrt{n\pi}+1)^2} = \frac{\pi n x^2}{24(\pi n x^2 + 2x\sqrt{\pi n}+1)}$$
$$= \frac{1}{24} - o(1)$$

To guarantee that the graph is connected and that if a step at i fails (there is no long edge to a node in A) there is a close neighbor k with the same order of distance to j, we can just compose $g'(x)$ with g_r^1 s.t $g(x) = max\{g'(x), g_r^1(x)\}$. Now the expected number of steps to reduce the distance to j by a factor of two is constant, and the total number of steps to reach the destination is $\Theta(\log n)$. In this case the average degree of $\mathcal{D}(n,g)$ is bounded by:

$$n\int_{\mathbb{R}} g(x) \leq \int_{\mathbb{R}} g'(x) + \int_{\mathbb{R}} g_r^1(x)$$
$$= n(\frac{1}{n+2n^{1.5}} + \frac{\log(1+2\sqrt{n})}{n} - \frac{1}{n}) + n\frac{\log n + \gamma_n}{n}$$
$$= \Theta(\log n)$$

The high clustering is guaranteed by $\alpha = 1$ and $\mathcal{D}(n, g_r^1)$, it can be decreased by taking $\alpha = 1 - \epsilon$ for a small enough ϵ. □

5 Conclusions

We offer the perspective that the similarity of the connectivity threshold results for $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$ is a consequence of the integral over the connection function when the nodes are uniformly distributed on the unit disk. In both cases the integral is $\frac{\log n}{n}$ and it determines the probability of isolated nodes which, in turn, govern the connectivity threshold by a result from continuum percolation. Following this view, we introduce a class of random distance graphs, $\mathcal{D}(n, g_r^\alpha)$, with a connection function that has the same integral as $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$. This guarantees that the connectivity threshold is identical for $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$. In addition, we show that for a wide range of parameters this class behaves as a Small World graph, contrary to $\mathcal{G}(n,r)$ and $\mathcal{B}(n,p)$. As opposed to previous Small World models, we propose a completely random model which seems to be more suitable for real-life situations.

We conjecture that a similar connectivity result can be obtained for more general functions than presented here and we state the following:

Conjecture 1 *Let $c > 1$ be a constant and let $g(x)$ be a function $[0, \frac{2}{\sqrt{\pi}}] \to [0,1]$ s.t $g(x)$ encloses zero and $\int_{\mathcal{U}} g(x) = \frac{c\log n}{n}$. Then $\mathcal{D}(n,g)$ is connected with high probability.*

In future work we would like to prove more properties of random distance graphs like the sharp threshold of monotone properties, the cover time and the mixing time and to offer a random scale-free graph that is based on distance graphs.

References

1. Penrose, M.D.: The longest edge of the random minimal spanning tree. The Annals of Applied Probability **7** (1997) 340–361
2. Gupta, P., Kumar, P.R.: Critical power for asymptotic connectivity in wireless networks. In Stochastic Analysis, Control, Optimization and Applications: A Volume in Honor of W. H. Fleming (1998) 547–566
3. Díaz, J., Petit, J., Serna, M.: Random geometric problems on $[0, 1]^2$. Lecture Notes in Computer Science **1518** (1998)
4. Penrose, M.D.: Random Geometric Graphs. Volume 5 of Oxford Studies in Probability. Oxford University Press (2003)
5. Gilbert, E.N.: Random graphs. The Annals of Mathematical Statistics **30** (1959) 1141–1144
6. Erdős, P., Rényi, A.: On random graphs. Publicationes Mathemticae (Debrecen) **6** (1959) 290–297
7. Bollobás, B.: Random Graphs. Academic Press, Orlando, FL (1985)
8. Goel, A., Rai, S., Krishnamachari, B.: Sharp thresholds for monotone properties in random geometric graphs. In: Proc. of the thirty-sixth annual ACM symposium on Theory of computing. (2004) 580–586
9. Cooper, C., Frieze, A.: The cover time of sparse random graphs. In: Proc. of the fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-03). (2003) 140–147
10. Avin, C., Ercal, G.: On the cover time of random geometric graphs. In: Proc. Automata, Languages and Programming, 32nd International Colloquium, ICALP05. (2005) 677–689
11. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393** (1998) 440–442
12. Newman, M., Watts, D., Strogatz, S.: Random graph models of social networks. In: Proc. Natl. Acad. Sci. Volume 99. (2002) 2566–2572
13. Kleinberg, J.: The small-world phenomenon: an algorithm perspective. In: STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing, New York, NY, USA, ACM Press (2000) 163–170
14. Milgram, S.: The small world problem. Psychology today **2** (1967) 60–67
15. Gilbert, E.N.: Random plane networks. Journal of the Society for Industrial and Applied Mathematics **9** (1961) 533–543
16. Meester, R., Penrose, M.D., Sarkar, A.: The random connection model in high dimensions. Statistics & Probability Letters **35** (1997) 145–153
17. Singer, K.B.: Random intersection graphs. PhD thesis, Johns Hopkins University (1995)
18. Fill, J.A., Scheinerman, E.R., Singer-Cohen, K.B.: Random intersection graphs when m= w(n): an equivalence theorem relating the evolution of the g(n, m, p) and g(n,p) models. Random Struct. Algorithms **16** (2000) 156–176
19. Penrose, M.D.: On a continuum percolation model. Advances in Applied Probability **23** (1991) 536–556
20. Meester, R., Roy, R.: Continuum percolation. Cambridge Univesity Press, Cambridge, UK. (1996)
21. Palmer, E.M.: Graphical Evolution: An Introduction to the Theory of Random Graphs. Jhon Wiley & Sons, Inc. (1985)
22. Chung, F., Lu, L.: The diameter of sparse random graphs. Adv. Appl. Math. **26** (2001) 257–279