# Percolation in the $k$-nearest neighbor graph

P. Balister and B. Bollobás

**Abstract.** Let $\mathcal{P}$ be a Poisson process of intensity one in $\mathbb{R}^2$. For a fixed integer $k$, join every point of $\mathcal{P}$ to its $k$ nearest neighbors, creating a directed random geometric graph $\vec{G}_k(\mathbb{R}^2)$. We prove bounds on the values of $k$ that, almost surely, result in an infinite connected component in $\vec{G}_k(\mathbb{R}^2)$ for various definitions of "component". We also give high confidence results for the exact values of $k$ needed. In particular, for percolation on the underlying (undirected) graph of $\vec{G}_k(\mathbb{R}^2)$, we prove that $k = 11$ is sufficient, and show with high confidence that $k = 3$ is the actual threshold for percolation.

## 1. Introduction

Let $\mathcal{P}$ be a Poisson process of intensity one in $\mathbb{R}^d$, $d \geq 2$. For a fixed integer $k$, we join every point of $\mathcal{P}$ to its $k$ nearest neighbors, creating a directed random geometric graph $\vec{G}_k(\mathbb{R}^d)$ in which every vertex has out-degree exactly $k$. In this paper we shall mainly consider the case $d = 2$. The connectivity of these graphs restricted to a finite region in $\mathbb{R}^2$ was studied in [14, 3, 4]. Here we shall study percolation in the infinite region $\mathbb{R}^2$, i.e., the existence or otherwise of infinite connected graph components. Since we are dealing with directed graphs, there are several possible definitions we can use for percolation.

- **U**: The underlying undirected graph has an infinite component.
- **O**: The directed graph has an infinite directed out-component.
- **I**: The directed graph has an infinite directed in-component.
- **S**: The directed graph has an infinite strongly connected subgraph.
- **B**: The directed graph has an infinite subgraph consisting of bidirectional edges.

Here an *out-component* is a subgraph with a spanning subtree whose edges are all directed away from a root vertex, while an *in-component* is a subgraph with a spanning subtree whose edges are all directed towards a root vertex. As all

The work of both authors was partially supported by the NSF grant CCF-0728928. The work of the second author was also partially supported by NSF grant CNS-0721983 and ARO grant W911NF-06-1-0076.

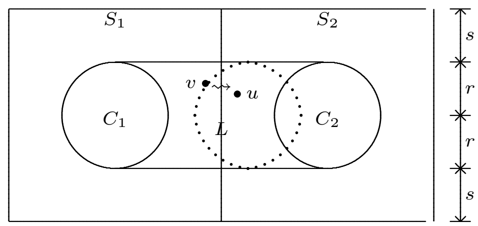


FIGURE 1. The regions defining $\mathcal{E}_{\mathbf{X},S_1,S_2}$, $\mathcal{E}'_{\mathbf{X},S_1,S_2}$, and the rolling ball $D_v$ (dotted circle).

degrees are almost surely finite in this model, conditions **O** and **I** are equivalent to the existence of infinite paths directed away, respectively towards, a root vertex. A *strongly connected* subgraph is a subgraph where there are directed paths from $u$ to $v$ for any choice of vertices $u$ and $v$ in the component. An edge $uv$ is *bidirectional* if both $\vec{uv}$ and $\vec{vu}$ lie in $\vec{G}_k(\mathbb{R}^d)$. Clearly we have the following implications

$$\mathbf{B} \Rightarrow \mathbf{S} \Rightarrow (\mathbf{I} \text{ and } \mathbf{O}), \qquad (\mathbf{I} \text{ or } \mathbf{O}) \Rightarrow \mathbf{U}.$$

From now on, let $\mathbf{X}$ denote any of $\mathbf{U}$, $\mathbf{O}$, $\mathbf{I}$, $\mathbf{S}$, or $\mathbf{B}$. Let $\theta_{\mathbf{X}}(k,d)$ denote the probability that $G_k(\mathbb{R}^d)$ contains an infinite connected component according to definition $\mathbf{X}$.

**Lemma 1.** *For all values of $k$, $d$, and $X$, $\theta_{\mathbf{X}}(k,d) \in \{0,1\}$.*

*Proof.* The existence of an infinite $\mathbf{X}$-component is translation invariant, so by ergodicity of the Poisson process under translation, the probability that such a component exists is either 0 or 1. □

It is clear that $\theta_{\mathbf{X}}(k,d)$ is non-decreasing in $k$. Define $k_{\mathbf{X},d}$ to be the *critical out-degree*, i.e., the smallest $k$ such that $\theta_{\mathbf{X}}(k,d) > 0$ (equivalently $\theta_{\mathbf{X}}(k,d) = 1$). Our aim in this paper is to present rigorous bounds on the critical out-degrees $k_{\mathbf{X},2}$ for each choice of $\mathbf{X}$ described above (Section 2, Theorem 2), as well as providing high confidence results for their exact values (Section 3).

## 2. Bounds

It has been shown by Häggström and Meester [8] that $k_{\mathbf{U},d} > 1$ for all $d$ (see also [13]) and that $k_{\mathbf{U},d} = 2$ for sufficiently large $d$. (Actually, the proof in [8] shows that $k_{\mathbf{O},d} = 2$ for sufficiently large $d$.) Also, Teng and Yao [13] have shown that $k_{\mathbf{U},2} \le 213$, while Bagchi and Bansal [1] have improved this to $k_{\mathbf{U},2} \le 188$. We improve and generalize these last bounds as follows.

**Theorem 2.** $k_{\mathbf{U},2} \le 11$, *and* $k_{\mathbf{O},2}, k_{\mathbf{I},2}, k_{\mathbf{S},2} \le 13$, *and* $k_{\mathbf{B},2} \le 15$.

To prove this result we shall compare the process to various bond percolation models on $\mathbb{Z}^2$. In these models, the states of the edges will not be independent, however they will satisfy the following property.

**Definition 1.** A bond percolation model is 1-*indepenedent* if whenever $E_1$ and $E_2$ are sets of edges at graph distance at least 1 from each another (i.e., if no edge of $E_1$ is incident to an edge of $E_2$) then the state of the edges in $E_1$ is independent of the state of the edges in $E_2$.

We shall use the following result, which is Theorem 2 of [5] (together with the remarks following its proof).

**Theorem 3.** *If every edge in a* 1*-independent bond percolation model on* $\mathbb{Z}^2$ *is open with probability at least* 0.8639*, then almost surely there is an infinite open component. Moreover, for any bounded region, there is almost surely a cycle of open edges surrounding this region.* □

*Proof of Theorem* 2. Let us first consider the case of **U**-percolation. Write $u \rightsquigarrow_{\mathbf{U}} v$ (or just $u \rightsquigarrow v$) if either $\vec{uv} \in \vec{G}_k(\mathbb{R}^2)$ or $\vec{vu} \in \vec{G}_k(\mathbb{R}^2)$, i.e., if $uv$ is an edge of the underlying undirected graph. (Of course, this definition is symmetric, so that $u \rightsquigarrow v$ iff $v \rightsquigarrow u$. However, when we generalize this argument to the other types of percolation this may no longer hold.)

For percolation we need to find an infinite $\rightsquigarrow$-path, i.e., a sequence $u_1, u_2, \dots$ with $u_i \rightsquigarrow u_{i+1}$ for all $i$. Consider the rectangular region consisting of two adjacent squares $S_1$, $S_2$ shown in Figure 1. Both $S_1$ and $S_2$ have side length $2r + 2s$, where $r$ and $s$ are to be chosen later. Also, $S_2$ may be to the right, left, above or below $S_1$, in which case Figure 1 should be rotated accordingly. We define the *basic good event* $\mathcal{E}_{\mathbf{U},S_1,S_2}$ to be the event that every vertex $u_1$ in the central disk $C_1$ of $S_1$ is joined to at least one vertex $v$ in the central disk $C_2$ of $S_2$ by a $\rightsquigarrow$-path, regardless of the state of the Poisson process outside of $S_1 \cup S_2$, and moreover that $C_1$ contains at least one vertex.

Now consider the following percolation model on $\mathbb{Z}^2$. Each vertex $(i, j) \in \mathbb{Z}^2$ corresponds to a square $[Ri, R(i+1)] \times [Rj, R(j+1)]$ in $\mathbb{R}^2$, where $R = 2r+2s$, and an edge is open between adjacent vertices (corresponding to squares $S_1$ and $S_2$) if *both* the corresponding basic good events $\mathcal{E}_{\mathbf{U},S_1,S_2}$ and $\mathcal{E}_{\mathbf{U},S_2,S_1}$ hold. Note that this is indeed a 1-independent model on $\mathbb{Z}^2$ since the event $\mathcal{E}_{\mathbf{U},S_1,S_2}$ depends only on the Poisson process within the region $S_1 \cup S_2$, and thus sets of edges at distance at least one apart in $\mathbb{Z}^2$ depend on the Poisson process in disjoint regions of $\mathbb{R}^2$. Any open path $p_1, p_2, p_3, \dots$ in $\mathbb{Z}^2$, corresponds to a sequence of basic good events $\mathcal{E}_{S_1,S_2}, \mathcal{E}_{S_2,S_3}, \dots$ that occur, where $S_i$ is the square associated to $p_i$. Every vertex $u_1$ of the original Poisson process that lies in the central disk $C_1$ of $S_1$ now has an infinite $\rightsquigarrow$-path leading away from it, since one can find points $u_i$ in the central disk of $S_i$ and $\rightsquigarrow$-paths from $u_{i-1}$ to $u_i$ inductively for every $i > 1$. In particular, each such $u_1$ lies in an infinite **U**-component. Moreover, such vertices exist in $C_1$, so there is an infinite **U**-component. From the bounds in Table 1 (which will be proved in Lemma 4 below), we see that for $k = 11$ one can choose $r$ and $s$ so that

$$\mathbb{P}(\mathcal{E}_{\mathbf{U},S_1,S_2} \text{ fails}) \leq 0.0653 < 0.06805 = (1 - 0.8639)/2,$$

so in particular

$$\mathbb{P}(\mathcal{E}_{\mathbf{U},S_1,S_2} \text{ and } \mathcal{E}_{\mathbf{U},S_2,S_1} \text{ hold}) \geq 0.8639.$$

The result now follows from Theorem 3.

For $k_{\mathbf{B},2}$ we follow the same proof as above, except that we define $u \rightsquigarrow_{\mathbf{B}} v$ to hold if *both* $\vec{uv} \in \vec{G}_k(\mathbb{R}^2)$ and $\vec{vu} \in \vec{G}_k(\mathbb{R}^2)$. The event $\mathcal{E}_{\mathbf{B},r,s}$ is now defined as for $\mathcal{E}_{\mathbf{U},r,s}$ except that we use $\rightsquigarrow_{\mathbf{B}}$ in place of $\rightsquigarrow_{\mathbf{U}}$, and the result follows from the bound (for $k = 15$) given in Table 1, since $0.0676 < 0.06805$.

Similarly, the bounds in Table 1 give $k_{\mathbf{O},2} \leq 13$, where we follow the same proof as above using $u \rightsquigarrow_{\mathbf{O}} v$, which is defined to hold if $\vec{uv} \in \vec{G}_k(\mathbb{R}^2)$. In this case $\rightsquigarrow$ is not symmetric, however the above proof still gives an infinite outwards directed path from some vertex.

At first sight it seems from Table 1 that the bound for $k_{\mathbf{I},2}$ (using $u \rightsquigarrow_{\mathbf{I}} v$, which holds iff $\vec{vu} \in \vec{G}_k(\mathbb{R}^2)$) will not be as good. Moreover, we do not have an analogous proof for $k_{\mathbf{S},2}$. However, it turns out that our bound on $k_{\mathbf{O},2}$ applies to $k_{\mathbf{S},2}$ as well. To see this, note that the above argument shows that (for $k = 13$, suitable $r$, $s$, and $\rightsquigarrow_{\mathbf{O}}$) we have an infinite path $p_1p_2\ldots$ in $\mathbb{Z}^2$ corresponding to a sequence of squares $S_1, S_2, \ldots$, with each edge $p_ip_{i+1}$ corresponding to basic good events $\mathcal{E}_{\mathbf{O},S_i,S_{i+1}}$, $\mathcal{E}_{\mathbf{O},S_{i+1},S_i}$ in *both* directions. Let $Z_i$ be the set of vertices in $\vec{G}_k(\mathbb{R}^2)$ that are in the central disk of $S_i$. Then $Z_1$ is almost surely finite. Fix $N > 0$. Each $v_{i_1} \in Z_1$ is joined to some $u_{i_1} \in Z_N$ by a directed $\rightsquigarrow_{\mathbf{O}}$-path in $\vec{G}_k(\mathbb{R}^2)$. But similarly $u_{i_1}$ is joined by a directed $\rightsquigarrow_{\mathbf{O}}$-path to some element $v_{i_2} \in Z_1$. Iterating this process starting at $v_{i_2}, v_{i_3}, \ldots$ in turn we must eventually repeat some vertex of $Z_1$. Hence some vertex of $Z_1$ lies on a directed $\rightsquigarrow_{\mathbf{O}}$-cycle[1] meeting at least $N$ vertices, (at least one from each of $Z_1, Z_2, \ldots, Z_N$, which are disjoint sets). Since this holds for every $N$, and $Z_1$ is finite, there exists a single $v \in Z_1$ which lies on arbitrarily large directed cycle. Thus in particular $v$ lies in an infinite strong component. Thus $k_{\mathbf{S},2} \leq 13$. Finally **S**-percolation implies **I**-percolation, so $k_{\mathbf{I},2} \leq 13$ also holds. $\square$

We note that in the above proof we declared an edge in $\mathbb{Z}^2$ to be open if *both* $\mathcal{E}_{\mathbf{U},S_1,S_2}$ and $\mathcal{E}_{\mathbf{U},S_2,S_1}$ held. It would seem that (at least in the **U** and **B** cases) that we would need only one, say $\mathcal{E}_{\mathbf{U},S_1,S_2}$ with $S_2$ either to the right or above $S_1$. However, in this case one needs an *oriented* open path in $\mathbb{Z}^2$, which at each step goes either to the right or up, to obtain an infinite $\rightsquigarrow_{\mathbf{U}}$-path. This is because $\mathcal{E}_{\mathbf{U},S_1,S_2}$ and $\mathcal{E}_{\mathbf{U},S_3,S_2}$ do not force a path from $S_1$ to $S_3$. Unfortunately no good bounds appear to have been proved for 1-independent oriented bond percolation in $\mathbb{Z}^2$, and in any case such bounds are unlikely to improve much on the method used above.

To complete the proof of Theorem 2, we need to show the following.

[1]By *cycle* we mean a closed trail. That is, we allow repeated vertices.

TABLE 1. Upper bounds on $\min_{r,s} \mathbb{P}(\mathcal{E}_{\mathbf{X},S,S'}$ fails). (All numbers rounded up.)

| $\mathbf{X}\backslash k$ | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|
| **U** | .1786 | .1090 | .0653 | .0386 | .0225 | .0130 | .0075 |
| **O** | .3424 | .2215 | .1402 | .0871 | .0533 | .0322 | .0192 |
| **I** | .4906 | .3511 | .2476 | .1725 | .1189 | .0812 | .0551 |
| **B** | .6217 | .4472 | .3151 | .2183 | .1492 | .1009 | .0676 |

**Lemma 4.** *The probabilities that the $\mathcal{E}_{\mathbf{X},S,S'}$ fail can be bounded (for suitable choices of $r$ and $s$) by the values given in Table 1.*

*Proof.* To bound the probability that a basic good event fails, we shall use the following "rolling ball" method. Let $C_1$, $C_2$, and $L$ be as in Figure 1. ($L$ is the region between the two disks $C_1$ and $C_2$.) For $\mathbf{X} \in \{\mathbf{U}, \mathbf{O}, \mathbf{I}, \mathbf{B}\}$, define $\mathcal{E}'_{\mathbf{X},S_1,S_2}$ to be the event that for every point $v \in C_1 \cup L$, there is a $u$ such that

(a) $v \rightsquigarrow_{\mathbf{X}} u$;
(b) $\|u - v\| \le s$; and
(c) $u \in D_v$, where $D_v$ is the disk of radius $r$ inside $C_1 \cup L \cup C_2$ with $v$ on its $C_1$-side boundary (the dotted disk in Figure 1).

Note in particular that (b) implies that the condition $u \rightsquigarrow v$ in (a) is independent of the Poisson process outside of $S_1 \cup S_2$. This is because both $u$ and $v$ are at distance at least $s$ from the exterior of $S_1 \cup S_2$, so the event that $u$ is among the $k$ nearest neighbors of $v$, or that $v$ is among the $k$ nearest neighbors of $u$, only depends on the points within $S_1 \cup S_2$. If $\mathcal{E}'_{\mathbf{X},S_1,S_2}$ holds, then every vertex $v$ in $C_1$ must be joined by a $\rightsquigarrow$-path to a vertex in $C_2$, since each vertex in $C_1 \cup L$ is joined to a vertex whose disk $D_v$ is further along in $C_1 \cup L \cup C_2$. Thus if we let $\mathcal{F}_{S_1}$ be the event that there is at least one vertex in $C_1$, we have $\mathcal{E}'_{\mathbf{X},S_1,S_2} \cap \mathcal{F}_{S_1} \subseteq \mathcal{E}_{\mathbf{X},S_1,S_2}$. The probability that $\mathcal{F}_{S_1}$ fails is simply the probability that there is no vertex in $C_1$, which is $e^{-\pi r^2}$. The probability that $\mathcal{E}'_{\mathbf{X},S_1,S_2}$ fails is bounded by the expected number of points $u$ for which the above conditions (a)–(c) fail. The expected number of points in $C_1 \cup L$ is $|C_1 \cup L| = 2r(2r + 2s)$. Thus

$$\mathbb{P}(\mathcal{E}'_{\mathbf{X},S_1,S_2} \text{ fails}) \le 2r(2r + 2s)p_{\mathbf{X},r,s} \tag{1}$$

where $p_{\mathbf{X},r,s}$ is the probability that (a)–(c) fail for some fixed $v$. Note that this probability is independent of the location of $v$ in $C_1 \cup L$.

To bound $p_{\mathbf{X},r,s}$ we consider the probability that the vertex $u$ *closest* to $v$ inside $D_v$ fails (a)–(c) (or does not exist). Let us consider the $\mathbf{X} = \mathbf{U}$ case first. Condition on the existence of a vertex $u \in \mathcal{P} \cap D_v$, and define the regions $A$, $B$, and $C$ as in Figure 2. Let $p_{\mathbf{U}}(u)$ be the probability that $u$ is the closest point to $v$ inside $D_v$, but that $v \rightsquigarrow_{\mathbf{U}} u$ fails. Let $\alpha = \|u-v\|$ be the distance between $u$ and $v$, and define for any $x \in \mathbb{R}^2$, $B(x, \alpha) = \{y : \|x - y\| \le \alpha\}$ to be the disk of radius $\alpha$

about $x$. Then there are at least $k$ points of the Poisson process in $A = B(v,\alpha)\setminus D_v$, at least $k$ points in $C = B(u,\alpha)$, but no points in $B = B(v,\alpha)\cap D_v$. Thus

$$\begin{aligned} p_{\mathbf{U}}(u) &= \sum_{i=0}^{\infty} \sum_{j,l\geq \max\{0,k-i\}} \mathbb{P}(\mathrm{Po}(|A\cap C|)=i)\,\mathbb{P}(\mathrm{Po}(|A\setminus C|)=j) \\ &\qquad\qquad \times \mathbb{P}(\mathrm{Po}(|C\setminus (A\cup B)|)=l)\,\mathbb{P}(\mathrm{Po}(|B|)=0) \\ &= \sum_{i=0}^{\infty} \sum_{j,l\geq \max\{0,k-i\}} \frac{|A\cap C|^i |A\setminus C|^j |C\setminus (A\cup B)|^l}{i!j!l!} e^{-|A\cup B\cup C|}, \end{aligned} \tag{2}$$

since if there are $i$ points in $A\cap C$ then there are at least $k-i$ in $A\setminus C$ and $C\setminus(A\cup B)$, and none in $B$. Now the probability that there is no $u$ satisfying conditions (a)–(c) above is bounded by

$$p_{\mathbf{U},r,s} \leq e^{-|D_v\cap B(v,s)|} + \int_{u\in D_v\cap B(v,s)} p_{\mathbf{U}}(u)\,du. \tag{3}$$

The first term being the probability that there is no $u$ satisfying (b) and (c), and the integral gives the probability that such a $u$ exists, but that the closest one to $v$ fails (a). Explicit calculation of this upper bound to $p_{\mathbf{U},r,s}$ is rather unpleasant due to the calculation of the areas above. However, numerical bounds can be computed (see Appendix A and [2]). Finally,

$$\mathbb{P}(\mathcal{E}_{\mathbf{U},S_1,S_2} \text{ fails}) \leq \mathbb{P}(\mathcal{F}_{S_1} \text{ fails}) + \mathbb{P}(\mathcal{E}'_{\mathbf{U},S_1,S_2} \text{ fails}) \leq e^{-\pi r^2} + 2r(2r+2s)p_{\mathbf{U},r,s}, \tag{4}$$

and this bound can be minimized over various values of $r$ and $s$. The minimum values obtained are listed in Table 1 (row $\mathbf{U}$) for various values of $k$.

The calculation for the other cases is exactly analogous. For $\mathbf{B}$ we replace (2) by

$$p_{\mathbf{B}}(u) = \sum_{i=0}^{\infty} \sum_{\max\{j,l\}\geq \max\{0,k-i\}} \frac{|A\cap C|^i |A\setminus C|^j |C\setminus (A\cup B)|^l}{i!j!l!} e^{-|A\cup B\cup C|}$$

since now we require *either* at least $k$ points in $A$ *or* at least $k$ points in $C$ for $v \rightsquigarrow u$ to fail. For $\mathbf{O}$, failure occurs when there are at least $k$ points in $A$, so (2) becomes

$$p_{\mathbf{O}}(u) = \sum_{j=k}^{\infty} \frac{|A|^j}{j!} e^{-|A\cup B|}.$$

For completeness we also consider the case $\mathbf{I}$, where at least $k$ points in $C$ is required for $v \rightsquigarrow u$ to fail. In this case (2) becomes

$$p_{\mathbf{I}}(u) = \sum_{l=k}^{\infty} \frac{|C\setminus B|^l}{l!} e^{-|B\cup C|}.$$

In each case the bounds (3) and (4) generalize to

$$\mathbb{P}(\mathcal{E}_{\mathbf{X},S_1,S_2} \text{ fails}) \leq e^{-\pi r^2} + 2r(2r+2s)\Big(e^{-|D_v\cap B(v,s)|} + \int_{u\in D_v\cap B(v,s)} p_{\mathbf{X}}(u)\,du\Big), \tag{5}$$

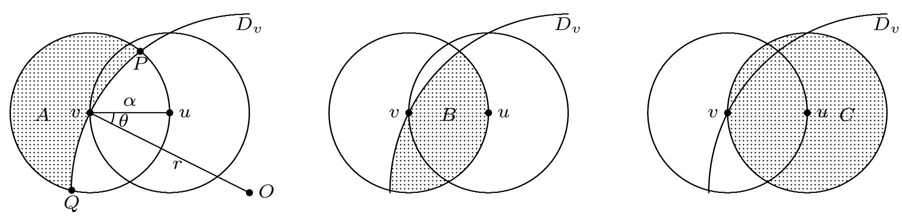


FIGURE 2. Areas $A = B(v, \alpha) \setminus D_v$, $B = B(v, \alpha) \cap D_v$, and $C = B(u, \alpha)$.

and the minimum value (over $r$ and $s$) found is listed in Table 1. □

## 3. High confidence results

In this section, we evaluate the critical out-degrees $k_{\mathbf{X},2}$ with *high confidence*. Here, high confidence means that we reduce to showing hat if a certain high dimensional integral exceeds a given value, then percolation occurs (respectively does not occur). Unfortunately, the integral is impractical to evaluate exactly, so it is estimated using a Monte-Carlo approach. The value obtained then gives a proof of the result, subject to the proviso that the random numbers used in the Monte-Carlo calculation did not lie in the very small region of the sample space that gives a misleading value for this integral. (See [7, 5] for examples of this approach being applied to other percolation questions.)

**Results.** *With high confidence,* $k_{\mathbf{U},2} = 3$, $k_{\mathbf{O},2} = k_{\mathbf{I},2} = k_{\mathbf{S},2} = 4$, $k_{\mathbf{B},2} = 5$.

The value $k_{\mathbf{U},2} = 3$ has been conjectured on the basis of simulation results by a number of authors, however we note that high confidence results are substantially more that just simulation. Slow convergence to a limit can, and has, mislead people into believing limiting results that are not true. High confidence results on the other hand must quantify the difference between the finite and infinite models.

To show percolation in the cases $\mathbf{X} \in \{\mathbf{U}, \mathbf{O}, \mathbf{B}\}$ (with $k = 3, 4, 5$ respectively), we choose $r$ and $s$ as above. We then generate a random instance of the process inside $S_1 \cup S_2$ and test for the following conditions:

- $\mathrm{UB}_1$ For more than half of the vertices $v \in C_1$ there are $\rightsquigarrow$-paths from $v$ to more than half the vertices of $C_2$, regardless of the state of the process outside of $S_1 \cup S_2$.
- $\mathrm{UB}_2$ Similarly, more than half of the vertices of $C_2$ have paths to more than half the vertices of $C_1$, regardless of the state of the process outside of $S_1 \cup S_2$.

As before, it is clear that if we have a sequence of distinct squares $S_1, S_2, \dots$ with the above holding in $S_i \cup S_{i+1}$ for all $i$, then there will be an infinite $\rightsquigarrow$-path from some vertex in $C_1$. The conditions $\mathrm{UB}_1$ and $\mathrm{UB}_2$ were chosen in place of $\mathcal{E}_{\mathbf{X},S_1,S_2}$ and $\mathcal{E}_{\mathbf{X},S_2,S_1}$ since in general they have a higher probability of success. Note that requiring strictly more than half the vertices of $C_i$ to have a property implies in

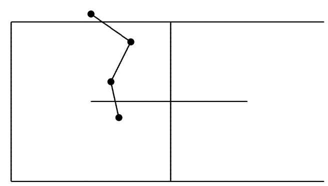

FIGURE 3. Condition $\mathrm{LB}_1$ requires that there is no path from outside $S_1 \cup S_2$ crossing the line segment joining the centers of $S_1$ and $S_2$.

particular that at least one vertex must exist in $C_i$. Since $\mathrm{UB}_1$ and $\mathrm{UB}_2$ depend only on the Poisson process within $S_1 \cup S_2$, by Theorem 3 we only need to show that these conditions hold with probability at least 0.8639. The condition that the path should be independent of the process outside of $S_1 \cup S_2$ is simply obtained by ignoring any edges $\vec{uv}$ of $\vec{G}_k(S_1 \cup S_2)$ where $d(u,v) > d(u, \partial(S_1 \cup S_2))$, since only edges $\vec{uv}$ with $d(u,v) \leq d(u, \partial(S_1 \cup S_2))$ are guaranteed to exist in $\vec{G}_k(\mathbb{R}^2)$.

Using a computer program (see [2]) we generated many instances, and counted the proportion of times these conditions held. The results are listed in the top half of Table 2. Using a similar argument as in the proof of Theorem 2, the infinite path in $\mathbb{Z}^2$ in the $\mathbf{X} = \mathbf{O}$ case actually gives us an infinite strong component, so in fact gives a bound for $k_{\mathbf{S}}$. From these we calculate the confidence level, i.e., the probability $p$ that these results (or better) could be obtained if the true probability of success was $< 0.8639$. In all cases considered $p$ is ludicrously small.

To show *lack* of percolation in the cases $\mathbf{X} \in \{\mathbf{U}, \mathbf{I}, \mathbf{O}, \mathbf{B}\}$ (with $k = 2, 3, 3, 4$ respectively), we generate, for suitable $r$, $s$, instances of the process in $S_1 \cup S_2$ and check the following condition holds:

$\mathrm{LB}_1$ Regardless of the state outside $S_1 \cup S_2$, there is no $\rightsquigarrow$-path from outside of $S_1 \cup S_2$ that crosses the line segment that joins the center point of $S_1$ to the center point of $S_2$ (see Figure 3).

Once again we define a percolation model of $\mathbb{Z}^2$ by declaring an edge open if $\mathrm{LB}_1$ holds in the corresponding rectangle $S_1 \cup S_2$. This model is also clearly 1-independent. Suppose the probability that $\mathrm{LB}_1$ occurs is at least 0.8639. Then by Theorem 3 there are open cycles in the corresponding $\mathbb{Z}^2$ process surrounding any bounded region. If an infinite $\rightsquigarrow$-path existed starting in some such region, then it would have to cross this cycle, and in particular cross the central line segment in one of rectangles $S_1 \cup S_2$ corresponding to an open edge in this cycle. However, this would contradict condition $\mathrm{LB}_1$ for this edge.

Note that we could have demanded in $\mathrm{LB}_1$ only that there is no path from one boundary point to another boundary point that crosses the center line. However, the condition given is easier to test for, and is sufficient for our purposes.

To test whether an edge of a $\rightsquigarrow$-path could come from outside of $S_1 \cup S_2$ to a vertex $v \in S_1 \cup S_2$ is somewhat harder. In the $\mathbf{X} = \mathbf{I}$ case, one can just test

TABLE 2. Results of Monte-Carlo simulation

| Test | $(r,s)$ | Successes | Trials | Confidence |
|---|---|---|---|---|
| $k_{\mathbf{U}} \leq 3$ | $(18,2)$ | 9984 | 10000 | $p < 10^{-597}$ |
| $k_{\mathbf{S}}(k_{\mathbf{O}}) \leq 4$ | $(18,2)$ | 9564 | 10000 | $p < 10^{-208}$ |
| $k_{\mathbf{B}} \leq 5$ | $(18,2)$ | 9960 | 10000 | $p < 10^{-555}$ |
| $k_{\mathbf{U}} > 2$ | $(0,60)$ | 9861 | 10000 | $p < 10^{-430}$ |
| $k_{\mathbf{O}} > 3$ | $(0,60)$ | 9667 | 10000 | $p < 10^{-269}$ |
| $k_{\mathbf{I}} > 3$ | $(0,60)$ | 9710 | 10000 | $p < 10^{-299}$ |
| $k_{\mathbf{B}} > 4$ | $(0,600)$ | 9460 | 10000 | $p < 10^{-157}$ |

whether or not the $k$ nearest neighbors of $v$ are all closer to $v$ than the boundary. If not, we assume an edge (of $\vec{G}_k(\mathbb{R}^2)$) could leave $S_1 \cup S_2$, and so $u \rightsquigarrow v$ for some $u$ outside of $S_1 \cup S_2$.

For the **O** and **U** cases, however, one must find the $k$ nearest neighbors in $S_1 \cup S_2$ of *every* possible point outside of $S_1 \cup S_2$. It is easy to see that it is enough to check points that lie on the boundary of $S_1 \cup S_2$, however there are still an infinite number of these. Instead we use the following algorithm. Pick a point $w$ on the boundary of $S_1 \cup S_2$ and find its $k+2$ nearest neighbors in $S_1 \cup S_2$. Mark the $k+1$ nearest neighbors of $w$ as possibly having an edge from outside $S_1 \cup S_2$. Let $d_i$ be the distance from $w$ to its $i$th nearest neighbor in $S_1 \cup S_2$. Now advance by a distance $(d_{k+2} - d_k)/2$ along the boundary from $w$ and then check this new point. Repeat this process until the entire boundary has been traversed. To see that this is sufficient, note that if $d(w', w) < (d_{k+2} - d_k)/2$, then the points that are not among the $k+1$ nearest neighbors of $w$ will all be further away from $w'$ than the $k$ nearest neighbors of $w$. Thus the $k$ nearest neighbors (in $S_1 \cup S_2$) of $w'$ will be a subset of the $k+1$ nearest neighbors of $w$.

Finally, for the case **B** we could use the above algorithm and also check edges leaving $S_1 \cup S_2$ as in the **I** case. However, the above boundary searching algorithm is rather slow, and the size of the rectangle $S_1 \cup S_2$ needed was rather large in this case. Thus we just checked edges leaving $S_1 \cup S_2$ as in the **I** case and assumed any $u \in S_1 \cup S_2$ with an edge leaving $S_1 \cup S_2$ also had an edge in from the exterior of $S_1 \cup S_2$. This makes the test $\mathrm{LB}_1$ slightly more pessimistic, but in practice the difference in success rate was minimal, while the program ran much faster.

The results of these computer simulations are listed in the bottom half of Table 2. Once again, in all cases considered the result is shown with extremely high confidence. All our simulations used the ARC4 algorithm [12] for pseudo-random number generation. More details, including the `C` source code, can be found in [2].

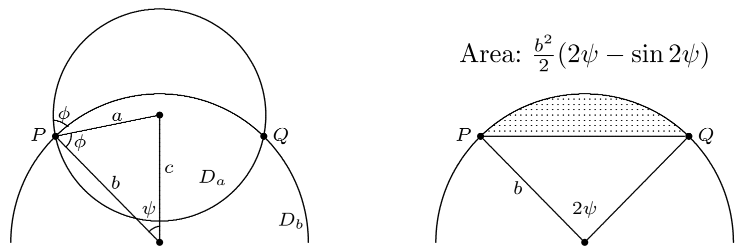


FIGURE 4. Lune used to define $L(a,b,\phi)$.

## Appendix A. Calculation of the integral

To calculate the areas in Figure 2, write $\alpha = d(u,v)$ for the distance between $u$ and $v$, and $\theta$ for the angle $Ovu$, where $O$ is the center of $D_v$ (see Figure 2). The following is a useful formula for the area $L(a,b,\phi)$ of a lune $D_a \setminus D_b$ consisting of the area inside a disk $D_a$ of radius $a$ and outside a disk $D_b$ of radius $b$, and which makes an angle of $\phi \in (0,\pi)$ at each end (see Figure 4).

$$L(a,b,\phi) = \tfrac{a^2}{2}\big(2(\psi+\phi) - \sin 2(\psi+\phi)\big) - \tfrac{b^2}{2}\big(2\psi - \sin 2\psi\big)$$

where $\psi = \cos^{-1}\big(\frac{b^2+c^2-a^2}{2bc}\big) \in [0,\pi]$ and $c^2 = a^2+b^2-2ab\cos\phi$. (The first term is the area above the line $PQ$ inside of $D_a$, and the second is the area above $PQ$ inside $D_b$.) It is also useful to define

$$L(a,b,\phi) = -L(b,a,-\phi) \qquad \text{when } \phi < 0. \tag{6}$$

The angle of the lune $A$ in Figure 2 is given by

$$\phi = \cos^{-1}(\alpha/2r) \in [0, \tfrac{\pi}{2}.]$$

(The angle $\phi$ is also one of the angles at the base of the isosceles triangle $OvP$.) Clearly $\alpha \le 2r$, and indeed, $\theta \in (-\phi,\phi)$, otherwise $u$ would not lie in $D_v$. By symmetry we may assume that $\theta \ge 0$, so that $\theta \in [0,\phi)$. Now

$$|A \cup B| = |C| = \pi\alpha^2. \tag{7}$$

and a simple calculation shows that

$$|C \setminus (A \cup B)| = |(A \cup B) \setminus C| = \big(\tfrac{\pi}{3} + \tfrac{\sqrt{3}}{2}\big)\alpha^2. \tag{8}$$

Also, by the definition of $L(a,b,\phi)$,

$$|A| = L(\alpha, r, \phi). \tag{9}$$

We now calculate $|A \cap C|$. This calculation splits into three cases depending on whether the two intersection points of the boundary of $D_v$ and $B(v,\alpha)$ ($P$ and $Q$

in Figure 2) lie inside of $B(u,\alpha)$. The result is

$$|A\cap C| = \begin{cases} L(\alpha,r,\theta) & \theta\le\phi-\frac{\pi}{3},\ \theta>\frac{\pi}{3}-\phi; \\ \frac{\alpha^2}{2}(\theta-\phi+\frac{\pi}{3})+L(\alpha,r,\phi-\frac{\pi}{3}) & \theta>\phi-\frac{\pi}{3},\ \theta>\frac{\pi}{3}-\phi; \\ |A|-|(A\cup B)\setminus C|+L(r,\alpha,\theta) & \theta\le\frac{\pi}{3}-\phi. \end{cases} \tag{10}$$

(First case when $P,Q\notin B(u,\alpha)$, third case $P,Q\in B(u,\alpha)$, second case when $P\in B(u,\alpha)$, $Q\notin B(u,\alpha)$. Also, in the second case $\phi-\frac{\pi}{3}$ will be negative if $r<\alpha$, so we also use (6).) Combining (7)–(10) allows us to evaluate the areas necessary for the calculation of $p_{\mathbf{X}}(u)$.

To prove a bound on the integral in (5), we note that (10) is monotonic in $\theta$, while (7)–(9) are independent of $\theta$. One can check that the formulae giving $p_{\mathbf{X}}(u)e^{|B|}$ (as a function of $\alpha$ and $\theta$) are monotonically increasing with both $\alpha$ and $\theta$, for each choice of $\mathbf{X}$. Also $e^{-|B|}$ is decreasing in $\alpha$. One can effectively bound $p_{\mathbf{X}}(u)$ over a small region $R$ in the $(\alpha,\theta)$-plane by multiplying the maximum value of $p_{\mathbf{X}}(u)e^{|B|}$ in $R$ by the maximum value of $e^{-|B|}$ over $R$. A bound on the total integral is then obtained by summing the bounds on the integrals over a suitable partition of $D_v\cap B(v,s)$ (see [2] for more details). Table 1 gives the results we obtained using this approach.

## References


[1] A. Bagchi and S. Bansal. On the metric distortion of nearest-neighbour graphs on random point sets, `http://arxiv.org/abs/0804.3784` (2008).

[2] P.N. Balister, website `http://www.memphis.edu/msci/people/pbalistr/knn`.

[3] P.N. Balister, B. Bollobás, A. Sarkar, M.J. Walters. Connectivity of random $k$-nearest neighbour graphs, *Advances in Applied Probability* **37** (2005) 1–24.

[4] P.N. Balister, B. Bollobás, A. Sarkar, M.J. Walters. A critical constant for the $k$-nearest-neighbour model. *Advances in Applied Probability* **41** (2009) 1–12.

[5] P.N. Balister, B. Bollobás, M.J. Walters. Continuum Percolation in the square and the disk, *Random Structures and Algorithms* **26** (2005) 392–403.

[6] J.M. González-Barrios, A.J. Quiroz, A clustering procedure based on the comparison between the $k$ nearest neighbors graph and the minimal spanning tree. *Statist. Probab. Lett.* **62** (2003) 23–34.

[7] B. Bollobás, A.M. Stacey. Approximate upper bounds for the critical probability of oriented percolation in two dimensions based on rapidly mixing Markov chains. *J. Appl. Probab.* **34** (1997) 859–867.

[8] O. Häggström and R. Meester, Nearest neighbour and hard sphere models in continuum percolation, *Random Structures and Algorithms* **9** (1996) 295–315.

[9] Y. Kim, S. Lee, Z. Lin, W. Wang, The invariance principle for the total length of the nearest-neighbor graph, *J. Theoret. Probab.* **18** (2005) 649–664.

[10] I. Kozakova, R. Meester, S. Nanda, The size of components in continuum nearest-neighbor graphs. *Ann. Probab.* **34** (2006) 528–538.

[11] S. Nanda and C.M. Newman, Random nearest neighbor and influence graphs on $\mathbb{Z}^d$ *Random Structures and Algorithms* **15** (1999) 262–278.

[12] R. Rivest, The RC4 encryption algorithm, *RSA Data Security Inc.* (1992).

[13] Shang-Hua Teng and Frances F. Yao, $k$-nearest-neighbor clustering and percolation theory, *Algorithmica* **49** (2007) 192–211.

[14] F. Xue and P.R. Kumar, The number of neighbors needed for connectivity of wireless networks. *Wireless Networks* **10** (2004) 169–181.

P. Balister
Department of Mathematical Sciences, University of Memphis, TN 38152, USA.
e-mail: `pbalistr@memphis.edu`

B. Bollobás
DPMMS, Centre for Mathematical Sciences, Cambridge University, Wilberforce Road, Cambridge CB3 0WA; and Department of Mathematical Sciences, University of Memphis, TN 38152, USA.
e-mail: `B.Bollobas@dpmms.cam.ac.uk`