Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

RARE EVENT PROBABILITY ESTIMATION FOR CONNECTIVITY OF LARGE RANDOM GRAPHS

Rohan Shah

School of Mathematics and Physics
The University of Queensland
Brisbane QLD 4702, AUSTRALIA

Christian Hirsch

Institute of Stochastics
Ulm University
89069 Ulm, GERMANY

Dirk P. Kroese

School of Mathematics and Physics
The University of Queensland
Brisbane QLD 4702, AUSTRALIA

Volker Schmidt

Institute of Stochastics
Ulm University
89069 Ulm, GERMANY

ABSTRACT

Spatial statistical models are of considerable practical and theoretical interest. However, there has been little work on rare-event probability estimation for such models. In this paper we present a conditional Monte Carlo algorithm for the estimation of the probability that random graphs related to Bernoulli and continuum percolation are connected. Numerical results are presented showing that the conditional Monte Carlo estimators significantly outperform the crude simulation estimators.

1 INTRODUCTION

Random graph models are of significant practical importance; see, e.g., Sahini and Sahimi 1994. The connectivity properties of such models are of considerable interest, for example in network reliability (Gertsbakh and Shpungin 2010, Colbourn 1987), percolation theory (Bollobás and Riordan 2006) and material design (Stenzel, Koster, Thiedmann, Oosterhout, Janssen, and Schmidt 2012). In percolation theory the focus is on infinite random graph models, which are theoretically more tractable. However physical systems of interest are necessarily finite and this suggests the use of finite random graph models in applications.

This paper studies the following problem: consider a connected 'base' graph $\mathcal{G}$, and retain vertices independently with probability p. If we use this random vertex subset to construct the induced subgraph, what is the probability that the induced subgraph is connected? Calculating such a probability exactly for a finite but large random graph model constitutes a difficult counting problem. In the network reliability setting this problem has been proved to be #P-complete (Colbourn 1987). Given the difficulties with exact computation we naturally turn to Monte Carlo methods.

However, crude Monte Carlo techniques can be cumbersome because connectivity is often a rare event and the problem becomes one of rare-event simulation. This is similar to the situation in network reliability, which also involves rare-event simulation; however in that case disconnection is the rare event, rather than connection. Typical methods for efficient rare event simulation include splitting (Kahn and Harris 1951, Glasserman, Heidelberger, Shahabuddin, and Zajic 1999, Garvels, van Ommeren, and Kroese 2002, L'Ecuyer, Demers, and Tuffin 2006, Botev and Kroese 2012), importance sampling (Glynn and Iglehart

1989, Asmussen and Rubinstein 1995) and conditional Monte Carlo (Asmussen and Glynn 2007). See Rubinstein and Kroese 2008 or Kroese, Taimre, and Botev 2011 for an overview of these techniques.

The connectivity criterion we use has previously been considered in network reliability, where it is known as *residual network connectivity* (Sutner, Satyanarayana, and Suffel 1991, Elmallah 1992, Colbourn, Satyanarayana, Suffel, and Sutner 1993, Stivaros and Sutner 1997, Chernyak 2004). It is important to note that residual network connectivity is a *non-monotone* criterion. That is, the addition of more vertices to the observed random graph may add additional connected components to the graph, making an initially connected graph disconnected. Similarly, the removal of vertices from an observed random graph may remove all but one connected component, causing an initially disconnected graph to become connected. In these non-monotone cases techniques based on permutation Monte Carlo (Elperin, Gertsbakh, and Lomonosov 1991, Lomonosov 1994, Hui, Bean, Kraetzl, and Kroese 2005) are difficult to apply.

Random *geometric* graphs are the continuous analog of random graph models. The defining property of these models is that the vertices of the graph are the points of a point process on a bounded sampling window. Although these models can be viewed as strictly combinatoric, the spatial structure of the model remains important. Even in the infinite domain case very little has been proved about the connectivity properties of such models and related critical exponents (Brereton, Hirsch, Kroese, and Schmidt 2014). This has lead to the widespread use of Monte Carlo methods to estimate unknown percolation thresholds (Quintanilla and Ziff 2007, Li and Östling 2013, Torquato and Jiao 2012). We show that the Monte Carlo estimate we propose can be applied to the Gilbert disk model with minimal change.

The rest of this paper is organized as follows. Section 2 describes the Bernoulli site percolation model on a finite lattice and the Gilbert disk model. Section 3 outlines the conditional Monte Carlo estimator for the Bernoulli site percolation model. Section 4 describes the adaptation of the conditional Monte Carlo estimator in Section 3 to the Gilbert disk model. Section 5 gives numerical results showing that the conditional Monte Carlo estimators perform significantly better than the crude simulation estimators. Appendix A describes the use of the power diagram to compute the area of the union of closed balls in $\mathbb{R}^2$. This material is used in Section 4.

2 PRELIMINARIES

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a finite connected graph. For any vertex v the *degree* of v is the number of edges incident to v, and we write $\deg(v)$. The maximum degree of any vertex in $\mathcal{G}$ is denoted by $\Delta(\mathcal{G})$. The cardinality of a finite set S is denoted by $|S|$. If S is an uncountable subset of $\mathbb{R}^d$ then $|S|$ denotes instead the Lebesgue measure of the set.

Take some $p \in (0,1)$ and let $q = 1-p$. Let $X = \{X_v\}_{v\in\mathcal{V}}$ be a collection of independent and identically distributed (iid) random variables with $X_v \sim \mathsf{Ber}(p)$. The random variable X_v is the *activation state* of vertex v. If $X_v = 1$ then v is said to be *activated*, otherwise it is said to be *deactivated*. Let $V(X)$ denote the set of activated vertices, that is

$$V(X) = \{v \in \mathcal{V} \mid X_v = 1\}.$$

The random subset $V(X)$ induces a random subgraph $G = G(X) = (V(X), E(X))$, where

$$E = E(X) = \{(v_1, v_2) \in \mathcal{E} \mid v_1, v_2 \in V(X)\}.$$

We typically omit the dependence of these random variables on X. We denote the collection of possible induced subgraphs of $\mathcal{G}$ by $\mathscr{P}(\mathcal{G})$. We will write the density of G with respect to counting measure on $\mathscr{P}(\mathcal{G})$ as $f_G(g;p)$. For $g_1, g_2 \in \mathscr{P}(\mathcal{G})$ induced by vertex subsets $V_1, V_2 \subseteq \mathcal{V}$ we will write $g_1 \cap g_2$ for the subgraph induced by the vertex set $V_1 \cap V_2$.

Models of this form for G are commonly known as *discrete site percolation models*, although $\mathcal{G}$ is often implicitly assumed to be infinite. We will also refer to the case where $\mathcal{G}$ is an arbitrary finite graph as being a discrete site percolation model. As vertices are retained independently with some probability p these models are said to be *Bernoulli site percolation models*.

Let $\mathcal{C} \subseteq \mathcal{V}$ be a subset of vertices such that the subgraph induced by the subset is *connected*. Define $\partial\mathcal{C}$ to be the *boundary vertices* of $\mathcal{C}$ in $\mathcal{G}$. That is,

$$\partial\mathcal{C} = \{v_1 \in \mathcal{V} \mid v_1 \notin \mathcal{C}, \{v_1, v_2\} \in \mathcal{E} \text{ for some } v_2 \in \mathcal{C}\}.$$

Define the connectivity probability $\ell(\mathcal{G}, p) = \mathbb{P}(G \text{ is connected})$. If we add vertices to G while maintaining a bound on the maximum vertex degree $\Delta(\mathcal{G})$, then $\ell(\mathcal{G}, p)$ will decay exponentially fast in the number of vertices. See Weichenberg, Chan, and Medard 2004 for results bounding the connectivity probability in the related network reliability setting. The idea of *prime failure events* used in Weichenberg, Chan, and Medard 2004 applies equally to our site-percolation model. Exponential decay means in particular that $\ell(\mathcal{G}, p)$ will be small for large base graphs $\mathcal{G}$ when $\Delta(\mathcal{G})$ is small. Note that the whole of $\mathcal{G}$, any single-vertex subgraph and the empty graph are all connected, so situations with p close to 0 or 1 are trivial.

The defining property of *random geometric graphs* is that their vertices are the points of a spatial point process on some bounded Borel set $R \subseteq \mathbb{R}^d$. Edges are added between vertices according to some probabilistic or deterministic rule. One possibility is to connect each vertex to the k closest other vertices; another is to connect a pair of vertices with some probability that depends on the Euclidean distance between them.

We focus on the *standard Gilbert disk model*, a special case of the *Boolean model* (Chiu, Stoyan, Kendall, and Mecke 2013). In this model the point process ξ that generates the vertices of the graph is a homogeneous Poisson point process on R with some intensity $\lambda > 0$, and any pair of vertices that are closer than some fixed distance r are connected by an edge. We will denote this model by $G_{\text{geo}}(R, \lambda, r)$, generally abbreviated to G_{geo}. The open ball of radius r around a point $x \in \mathbb{R}^d$ will be denoted by $B(x, r)$. We will consider this model in the specific case of $d = 2$.

3 CONDITIONAL MONTE CARLO FOR DISCRETE PERCOLATION

If $\left\{X^{(i)}\right\}_{i=1}^{\infty}$ are iid copies of X then the *crude simulation estimator* is

$$\widehat{\ell}_{\text{crude}}(\mathcal{G}, p) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left\{G\left(X^{(i)}\right) \text{ is connected}\right\}, \tag{1}$$

where $n \geq 1$ is an arbitrary fixed integer and $\mathbb{I}\{A\}$ denotes the indicator function of an event A. Our aim is to find an estimator that has better asymptotic properties than the crude simulation estimator, as the number of vertices in $\mathcal{G}$ is allowed to increase.

We can construct a simple conditional Monte Carlo estimator based on knowledge of a single connected component. After this connected component has been generated it is no longer necessary to simulate the states of the remaining vertices, as the connectivity probability can be computed exactly; it is the probability that the vertices not already simulated are all deactivated. See Equation (2) below for details. By the total variance formula this gives an estimate with smaller variance than the crude estimator given in Equation (1). See Billingsley 1995 for further details.

The idea of the algorithm is as follows. Select vertices randomly without replacement and generate X_v according to the $\mathsf{Ber}(p)$ distribution. Continue this process until $X_v = 1$, meaning that the selected vertex is activated, or until every vertex has been considered. The set of deactivated vertices is denoted by V_{deact}, with both $V_{\text{deact}} = \emptyset$ and $V_{\text{deact}} = \mathcal{V}$ being possible.

If an activated vertex is generated denote it by ω. We can then simulate the entire connected component C_{act} for ω by performing a depth-first search of $\mathcal{V} \setminus V_{\text{deact}}$. For every visited vertex v the random variable $X_v \sim \mathsf{Ber}(p)$ is simulated. If $X_v = 1$ then v is activated and the search continues to the neighbors of v. If no activated vertex was originally found, set $C_{\text{act}} = \emptyset$. The random object we will condition on is $Z_{\text{Ber}} = (V_{\text{deact}}, C_{\text{act}})$. It will be convenient to define $N_{\text{deact}} = |V_{\text{deact}}|$. Note that the random variables defined in this section are not just functions of the binary vector X.

The process of generating Z_{Ber} is illustrated in Figure 1. In this case V_{deact} contains the vertices v_1, v_2 and v_3, all of which were generated to be deactivated. The fourth vertex picked was simulated as being activated, so we have identified the vertex ω. The connected component for ω was then generated, and contains three vertices. Note that v_1 was already determined to be deactivated when we started to generate C_{act}. The activation state has only been generated for the marked vertices; the activation states of the unmarked vertices is unknown.

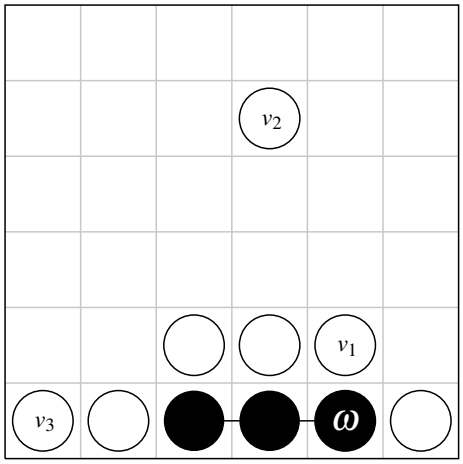


Figure 1: An example showing the process of generating Z_{Ber} where $\mathcal{G}$ is the 6 by 6 grid graph which does not include diagonal edges. C_{act} contains three vertices. V_{deact} also contains three vertices, labeled v_1, v_2 and v_3. Activated vertices are marked by filled circles and deactivated vertices by empty circles.

Let $f_N(x;n,p)$ denote the density of a $\mathsf{Binomial}(n,p)$-distributed random variable, and let $k = v_{\text{deact}} \cup \partial c_{\text{act}} \cup c_{\text{act}}$. Then the density of $G|Z_{\text{Ber}}$ is

$$f_{G|Z_{\text{Ber}}}(g \mid (v_{\text{deact}}, c_{\text{act}}); p) = f_N(|g| - |c_{\text{act}}|; |\mathcal{V} \setminus k|, p).$$

This occurs because the activation states of the vertices in $\mathcal{V} \setminus k$ are an iid Bernoulli family with success probability p. Conditional on Z_{Ber}, the only way for G to be connected is if all vertices outside the set $V_{\text{deact}} \cup \partial C_{\text{act}} \cup C_{\text{act}}$ are deactivated. This has probability

$$\begin{aligned}\mathbb{P}(G \text{ is connected } \mid Z_{\text{Ber}} = z_{\text{Ber}}) &= \mathbb{P}(\mathcal{V} \setminus (c_{\text{act}} \cup \partial c_{\text{act}} \cup v_{\text{deact}}) \text{ are deactivated }) \\ &= q^{|\mathcal{V}| - |c_{\text{act}} \cup \partial c_{\text{act}} \cup v_{\text{deact}}|}. \end{aligned} \tag{2}$$

Note that if $V_{\text{deact}} = \mathcal{V}$ then G is the empty graph which is considered connected. The conditional probability is simple to calculate, and Z_{Ber} is simple to simulate. We can use the expression

$$\ell(\mathcal{G}, p) = \mathbb{E}[\mathbb{P}(G \text{ is connected } \mid Z_{\text{Ber}})] \tag{3}$$

to construct the following conditional Monte Carlo estimator for the Bernoulli site percolation model.

Proposition 1 (Conditional Monte Carlo estimator for the Bernoulli site percolation model)

Let $\left\{Z_{\text{Ber}}^{(i)}\right\}_{i=1}^{\infty}$ be iid copies of Z_{Ber} and $\left\{X^{(i)}\right\}_{i=1}^{\infty}$ be iid copies of X, where each of the $Z_{\text{Ber}}^{(i)}$ depends only on $G\left(X^{(i)}\right)$. Define $P^{(i)} = \mathbb{P}\left(G\left(X^{(i)}\right) \text{ is connected } \middle| Z_{\text{Ber}}^{(i)}\right)$.

Then for any fixed $n \geq 1$, the Rao–Blackwell estimator $\widehat{\ell}_{\text{rao}}(\mathcal{G}, p) = \frac{1}{n}\sum_{i=1}^{n} P^{(i)}$ is unbiased for $\ell(\mathcal{G}, p)$ and has smaller variance than the crude simulation estimator introduced in (1).

Proof. This proposition follows from standard properties of conditional expectation and the total variance formula. See Billingsley 1995 for further details. □

While the variance of $\widehat{\ell}_{\text{crude}}$ can be calculated analytically, computing the variance of $\widehat{\ell}_{\text{rao}}$ appears intractable. This occurs because although we can simulate from Z_{Ber}, its distribution is unknown. However its variance can be estimated via simulation, by the sample variance of the $P^{(i)}$. Proposition 1 leads to the following algorithm for estimating $\ell(\mathcal{G}, p)$.

Algorithm 1 (Conditional Monte Carlo algorithm for the Bernoulli site percolation model)

1. Set $i = 1$.
2. Generate $N^{(i)}_{\text{deact}} = \min\left(|\mathcal{V}|, N^{(i)}_{\text{geom}}\right)$, where $N^{(i)}_{\text{geom}}$ has a $\mathsf{Geometric}(q)$ distribution on the non-negative integers.
3. If $N^{(i)}_{\text{deact}} = |\mathcal{V}|$, set $P^{(i)} = 1$, set $i = i+1$ and go to Step 2.
4. Select $N^{(i)}_{\text{deact}}$ vertices uniformly at random from $\mathcal{V}$ without replacement, and denote the chosen vertices by $V^{(i)}_{\text{deact}}$. These vertices will be deactivated.
5. Select a vertex $\omega^{(i)}$ uniformly at random from $\mathcal{V} \setminus V^{(i)}_{\text{deact}}$. This vertex will be activated.
6. Generate the connected component $C^{(i)}_{\text{act}}$ of $G^{(i)}$ containing $\omega^{(i)}$, conditional on the vertices in $V^{(i)}_{\text{deact}}$ being deactivated and $\omega^{(i)}$ being activated.
7. Calculate $P^{(i)} = \mathbb{P}\left(G^{(i)} \text{ is connected } \middle| \, Z^{(i)}_{\text{Ber}} = \left(V^{(i)}_{\text{deact}}, C^{(i)}_{\text{act}}\right)\right)$ according to (2).
8. If $i < n$ set $i = i+1$ and repeat Step 2. Otherwise return $\frac{1}{n}\sum_{i=1}^{n} P^{(i)}$.

Note that our construction of Z_{Ber} does not depend on an ordering of the vertices. Another possibility is to take some total ordering of $\mathcal{V}$ and let Z_{Ber} be the connected component of the *first* activated vertex of G. Here 'first' is with respect to the ordering of $\mathcal{V}$. Although we do not pursue this idea further in the discrete case, it leads to a very similar conditional Monte Carlo algorithm to the one described here. We continue this ordering-based approach with reference to random *geometric* graphs in Section 4.

4 CONDITIONAL MONTE CARLO FOR THE GILBERT DISK MODEL

Recall from Section 2 that for the Gilbert disk model on $\mathbb{R}^2$, the point process ξ generating the vertices of the graph is a homogeneous Poisson process with intensity λ on a bounded Borel set R of $\mathbb{R}^2$. The random graph $G_{\text{geo}} = G_{\text{geo}}(R, \lambda, r)$ is then generated by connecting any pair of vertices closer than r in the Euclidean distance by an edge. The probability to be estimated is

$$\ell(R, \lambda, r) = \mathbb{P}\left(G_{\text{geo}}(R, \lambda, r) \text{ is connected }\right).$$

Similar to Section 3, we can define the *crude simulation estimator* as

$$\widehat{\ell}_{\text{crude}}(R, \lambda, r) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left\{G^{(i)}_{\text{geo}} \text{ is connected }\right\},$$

where $n \geq 1$ is an arbitrary fixed integer and $\left\{G^{(i)}_{\text{geo}}\right\}_{i=1}^{\infty}$ are iid copies of $G_{\text{geo}}(R, \lambda, r)$.

For simplicity we will assume that R is a rectangular region with width w and height h, with bottom left corner at the origin. Similar to Section 3, the conditional Monte Carlo estimator proposed in this section is based around observing a single connected component, and then conditioning on there being no other connected components. The connected component will be chosen by assuming some total ordering of R, and then observing the connected component of ξ which contains the first point of ξ with respect to the ordering.

One natural ordering is the *lexicographic ordering*. For $x=(x_1,x_2), y=(y_1,y_2)\in\mathbb{R}^2$, the lexicographic ordering is defined by

$$(x_1,x_2) <_l (y_1,y_2) \qquad \text{if and only if} \qquad x_1<y_1 \text{ or } (x_1=y_1 \text{ and } x_2<y_2).$$

Another choice is the *distance ordering*, where for some fixed point $z\in R$ the ordering is

$$x <_d y \qquad \text{if and only if} \qquad \|x-z\|<\|y-z\|.$$

Note that we do not define the ordering among points which are equally distant from z. This is acceptable because we will only apply the ordering to the points of a Poisson process, and with probability 1 there will be no pair of points equally distant from the nonrandom point z. In the numerical examples in Section 5 we take z to be the center of R.

Let $\eta=(\eta_1,\eta_2)$ be the first point of ξ with respect to the chosen ordering of R. Let Z_{geo} be the vertices of the connected component of G_{geo} that contains η. Then conditional on Z_{geo} there must be no vertices in the region

$$R_{\text{empty}}=\{r\in R \mid r<\eta\}\subseteq R.$$

The set Z_{geo} is equal to $\xi\cap R_{\text{known}}$, where

$$R_{\text{known}}=\left(R\setminus R_{\text{empty}}\right)\cap\left(\cup_{v\in Z_{\text{geo}}}B(v,r)\right).$$

On the remainder of R the points of ξ are unknown. That is, conditional on Z_{geo} the distribution of ξ on the region

$$R_{\text{unknown}}=R\setminus\left(R_{\text{empty}}\cup R_{\text{known}}\right)$$

is that of a homogeneous Poisson point process with intensity λ. The random graph G_{geo} can be connected only if there are no points of ξ in R_{unknown}. Therefore we have

$$\mathbb{P}\left(G_{\text{geo}} \text{ is connected} \,\middle|\, Z_{\text{geo}}\right)=\exp\left(-\lambda\,|R_{\text{unknown}}|\right).$$

This leads to the following conditional Monte Carlo estimator for the Gilbert disk model.

Proposition 2 (Conditional Monte Carlo estimator for the Gilbert disk model)
Let $\left\{Z_{\text{geo}}^{(i)}\right\}_{i=1}^{\infty}$ be iid copies of Z_{geo} and $\left\{G_{\text{geo}}^{(i)}\right\}_{i=1}^{\infty}$ be iid copies of $G_{\text{geo}}(R,\lambda,r)$, where each of the $Z_{\text{geo}}^{(i)}$ depend only on $G_{\text{geo}}^{(i)}$. Define $P_{\text{geom}}^{(i)}=\mathbb{P}\left(G_{\text{geo}}^{(i)} \text{ is connected} \,\middle|\, Z_{\text{geo}}^{(i)}\right)$. Then for any fixed $n\geq 1$, the Rao–Blackwell estimator $\widehat{\ell}_{\text{rao}}(R,\lambda,r)=\frac{1}{n}\sum_{i=1}^{n}P_{\text{geo}}^{(i)}$ is unbiased and has smaller variance than the crude simulation estimator $\widehat{\ell}_{\text{crude}}(R,\lambda,r)$.

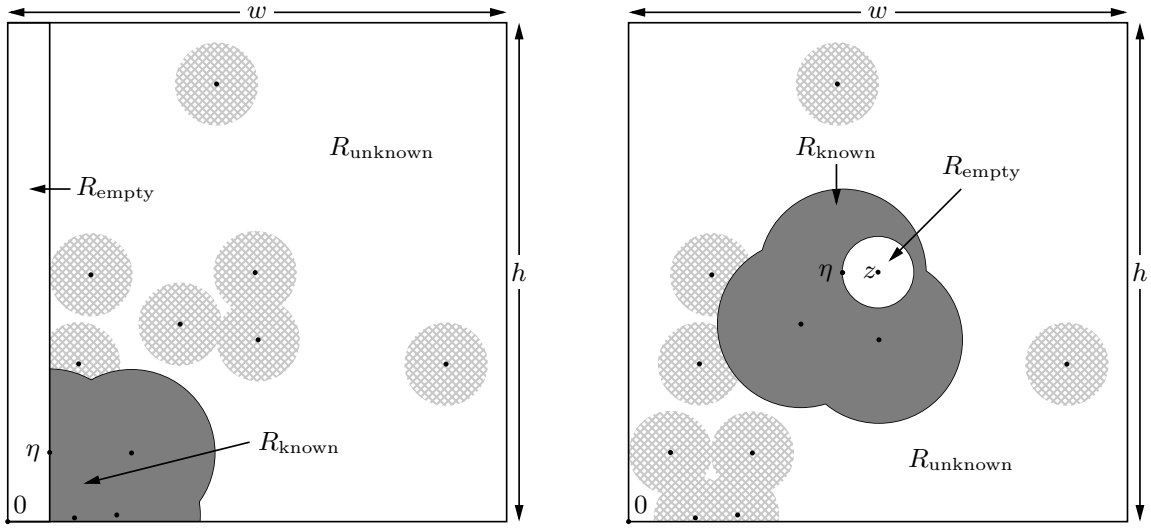
The difficulty with applying this estimator is determining the area of R_{unknown}, or equivalently $R_{\text{empty}}\cup R_{\text{known}}$. However in some cases this can be relatively straightforward. The following two propositions calculate these areas for the lexicographic ordering and distance ordering.

Proposition 3 (Lexicographic ordering) Consider the *lexicographic ordering* of R. Then $R_{\text{empty}}=[0,\eta_1)\times[0,h]$, and therefore

$$|R_{\text{unknown}}|=|R|-\left|R_{\text{empty}}\cup R_{\text{known}}\right|=h(w-\eta_1)-\left|\left(\cup_{v\in Z_{\text{geo}}}B(v,r)\right)\cap([\eta_1,w]\times[0,h])\right|.$$

Proposition 4 (Distance ordering) Consider the *distance ordering* of R with respect to a fixed point $z\in R$. Then $R_{\text{empty}}=B(z,\|z-\eta\|)$, and therefore

$$|R_{\text{unknown}}|=|R|-\left|R_{\text{empty}}\cup R_{\text{known}}\right|=hw-\left|\left(\cup_{v\in Z_{\text{geo}}}B(v,r)\cup B(z,\|z-\eta\|)\right)\cap R\right|.$$



(a) Lexicographic ordering

(b) Distance ordering

Figure 2: Illustration of the regions $R_{\text{known}}, R_{\text{unknown}}$ and R_{empty} for the lexicographic and distance orderings. Crosshatched regions represent balls of radius $\frac{r}{2}$ around other points of ξ that are not in the connected component of η. These regions are included in R_{unknown}.

Propositions 3 and 4 are easy to prove. The key observation is that we know the first point of ξ with respect to the ordering, and this excludes the possibility of observing any other points occurring in a region whose shape depends on the ordering chosen. See Figure 2 for illustrations of the regions R_{known}, R_{unknown} and R_{empty} for both orderings. In Figure 2a, region R_{empty} is the rectangular region on the left, R_{known} is the shaded region at the bottom left and R_{unknown} is the remaining region of R.

In the lexicographic ordering case, we can write $|R_{\text{unknown}}|$ as $h\eta_1 - |R_{\text{known}}|$. As R_{known} is the intersection of a union of disks of equal radius with a rectangular region, we can use the power diagram approach outlined in Appendix A to efficiently compute $|R_{\text{unknown}}|$. In the distance ordering case the set $R_{\text{empty}} \cup R_{\text{known}}$ is a union of disks of unequal radius, one of which is centered around z. We can again apply the power diagram to efficiently calculate this area.

For the lexicographic ordering, this leads to the following algorithm. The algorithm for the distance ordering is similar.

Algorithm 2 (Conditional Monte Carlo algorithm for the Gilbert disk model using lexicographic ordering)

1. Set $i = 1$.
2. Simulate $G_{\text{geo}}^{(i)}$. Determine the first vertex $\eta^{(i)} = \left(\eta_1^{(i)}, \eta_2^{(i)}\right)$ of $G_{\text{geo}}^{(i)}$ with respect to the lexicographic ordering, and let $Z_{\text{geo}}^{(i)}$ be the connected component containing $\eta^{(i)}$.
3. Construct the power diagram $V^{(i)}$ of the points in $Z_{\text{geo}}^{(i)}$, with all points taken to be the centers of disks of radius r.
4. Use $V^{(i)}$ to calculate $\left|R_{\text{known}}^{(i)}\right|$.
5. Set $P_{\text{geo}}^{(i)} = \exp\left(-\lambda\left(h\left(w - \eta_1^{(i)}\right) - \left|R_{\text{known}}^{(i)}\right|\right)\right)$.
6. If $i < n$ set $i = i+1$ and repeat Step 2. Otherwise return $\frac{1}{n}\sum_{i=1}^{n} P_{\text{geo}}^{(i)}$.

5 NUMERICAL RESULTS

The efficacy of a rare event probability estimator $\widehat{\ell}$ is generally assessed using its *relative error*, defined by

$$\mathrm{RE}\left(\widehat{\ell}\right) = \sqrt{\mathrm{Var}\left(\widehat{\ell}\right)\Big/\ell^2}.$$

We give the relative error (expressed as a percentage) for all our simulations, denoted by RE%.

If the estimators to be compared require different levels of computation, the *work normalized relative error* may be more useful. This is defined by

$$\mathrm{WNRE}\left(\widehat{\ell}\right) = \sqrt{T\left(\widehat{\ell}\right)\mathrm{Var}\left(\widehat{\ell}\right)\Big/\ell^2},$$

where $T\left(\widehat{\ell}\right)$ is the expected time required to compute the estimator $\widehat{\ell}$. We also give the work normalized relative error for all our simulations.

Example 1 Let $\mathcal{G}$ be the 6 by 6 grid graph without diagonal edges. This graph is small enough to allow complete enumeration of the 2^{36} subgraphs. We can therefore compute the probability of observing a connected subgraph exactly, for any parameter value p. A grid search for the parameter value which minimized the probability of connectivity gave a value of $p = 0.285$. The probability of connectivity for $p = 0.285$ was calculated to be 0.00125143.

Both conditional Monte Carlo and crude Monte Carlo were applied with sample size $n = 100,000$, and these simulations were repeated 1000 times. The estimated relative error was 8.93% for crude Monte Carlo and 2.48% for conditional Monte Carlo. The estimated work normalized relative error was 0.0456 for crude Monte Carlo and 0.0173 for conditional Monte Carlo.

Example 2 We started with a 20 by 20 grid graph which included diagonal edges and generated a random subgraph by retaining at random 340 of the 400 vertices. The base graph that was generated is shown in Figure 3. In this case exact computation is infeasible. Both crude Monte Carlo and conditional Monte Carlo were applied for 21 equally spaced different parameter values between $p = 0.05$ and $p = 0.99$ inclusive. For values of p between 0.097 and 0.332 inclusive the crude method did not identify any connected subgraphs and therefore estimated a probability of 0. A sample of the results where both methods estimated non-zero probabilities is shown in Table 1. The average values estimated by both methods were similar and are not shown. An up to five-fold improvement in relative error is observed when using the Rao–Blackwell estimator as compared to the crude estimator. The work normalized relative error is improved by up to a factor of 3.

Table 1: Simulation results for a randomly generated subgraph of the 20 by 20 grid graph.

p	Crude RE%	Crude WNRE	Conditional RE%	Conditional WNRE
0.05	129.64	2.88	25.08	1.10
0.43	251.41	11.65	67.54	3.85
0.47	29.54	1.11	11.19	0.62
0.52	5.76	0.22	3.08	0.19
0.57	1.73	0.09	1.03	0.06

Example 3 We considered the Gilbert disk model on a 6 by 6 square region of $\mathbb{R}^2$. The homogeneous Poisson point process generating the vertices of the graph had intensity 10, and the distance r at which points are connected was allowed to be $0.38, 0.40$ or 0.42. Both the distance and lexicographic orderings were considered. In the case of the distance ordering the fixed point z was taken to be the center of R. We used $n = 1,000,000$ samples for the crude estimator, and $n = 100,000$ samples for the conditional

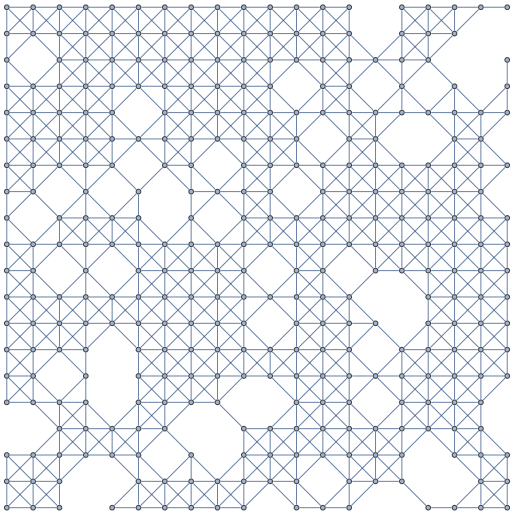


Figure 3: Subgraph of the 20 by 20 grid graph used as the base graph in Example 2.

Monte Carlo estimator. Different numbers of samples were used due to the different running times of both approaches. These simulations were repeated $1,000$ times to estimate the relative error. The simulation results are shown in Table 2. The distance ordering appears to outperform the lexicographic ordering in terms of relative error. However when the orderings are compared using the WNRE, which accounts for simulation time, the lexicographic ordering is found to perform slightly better. The WNRE values can be used to compare the performance of the crude estimator with that of the conditional Monte Carlo estimator. They show equal performance for the two approaches when $r = 0.4$, and that if $r = 0.42$ the conditional Monte Carlo has half the WRNE of the crude estimator. For $r = 0.38$ the target event is not sufficiently rare to justify the extra computation of the conditional Monte Carlo approach, and the crude estimator performs better.

Although this example suggests that the distance ordering performs better than the lexicographic ordering in terms of relative error, it is not conclusive. Further work, possibly involving a more comprehensive simulation study, could be done to compare these orderings.

Table 2: Simulation results for the Gilbert disk model on a 6×6 region with intensity 10.

Method	r	Estimate	Relative Error %	WNRE
Crude	0.38	3.55×10^{-06}	166.01	24.10
Crude	0.40	4.94×10^{-04}	14.32	1.99
Crude	0.42	1.34×10^{-02}	2.71	0.38
Lexicographic	0.38	3.55×10^{-07}	107.59	11.52
Lexicographic	0.40	4.96×10^{-05}	15.02	2.00
Lexicographic	0.42	1.34×10^{-03}	3.90	0.60
Distance	0.38	3.49×10^{-07}	96.99	13.61
Distance	0.40	4.93×10^{-05}	14.15	2.31
Distance	0.42	1.34×10^{-03}	3.64	0.64

ACKNOWLEDGMENTS

We would like to thank the referees for all their comments, which helped to improve the quality of this paper. This work was supported by the ARC Center of Excellence in Mathematical and Statistical Frontiers

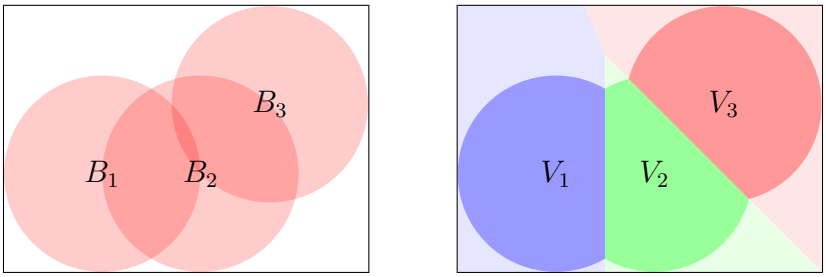


Figure 4: Disks B_1, B_2 and B_3, and the corresponding power diagram partition into regions V_1, V_2 and V_3.

for Big Data, Big Models and New Insights (ACEMS, CE140100049), and by the DAAD / Go8 Australia – Germany Joint Research Cooperation Scheme.

A THE POWER DIAGRAM

Assume that $R \subseteq \mathbb{R}^2$ is a rectangular region. Let $x_1, \ldots, x_n$ be points in R, let $r_1, \ldots, r_n$ be positive real numbers and let $B_i = B(x_i, r_i)$ be the open ball with radius r_i centered around the point x_i. In Section 4 it will be important to efficiently compute the area of the union of these open balls which is contained in R. That is, to compute the area of $A = R \cap (\bigcup_{i=1}^n B_i)$.

Computing this area is non-trivial. Let D^j be the set containing all j-element subsets of $\{1, \ldots, n\}$, where the j elements must also be distinct. Then the naive approach is to use the inclusion-exclusion principle. This involves computing

$$|A| = \sum_{j=1}^{n} (-1)^{j-1} \sum_{(i_1,\ldots,i_j)\in D^j} \left|R \cap B_{i_1} \cap \cdots \cap B_{i_j}\right| = \sum_{i_1=1}^{n} |R \cap B_{i_1}| - \sum_{\substack{i_1=1 \\ i_1 \neq i_2}}^{n} \sum_{i_2=1}^{n} |R \cap B_{i_1} \cap B_{i_2}| + \cdots \tag{4}$$

Each of the terms in (4) is an intersection of open balls with R, and the areas of such regions can be calculated easily. However if many of the B_i intersect then the number of terms that must be calculated grows extremely fast. In our application the inclusion-exclusion computation is prohibitively slow.

Another approach is to partition R into sub-regions $V_1, \ldots, V_n$, so that the part of A contained in V_j can be attributed uniquely to the open ball B_j. This eliminates the problem of multiple counting that the inclusion-exclusion principle tries to deal with, allowing $|A|$ to be decomposed as

$$|A| = \sum_{j=1}^{n} \left|V_j \cap A\right| = \sum_{j=1}^{n} \left|V_j \cap B_j\right|.$$

Such a partition of R can be constructed using the *power diagram*, also known as the *Laguerre tessellation*. Let C_i denote the circle of radius r_i around point x_i. Then for any point $y \in R$ at distance d from x_i, the *power* of y with respect to C_i is $d^2 - r^2$. Note that the power can be negative. As an example, let C be the circle of radius 2 located at the origin. Then the power of the point $(1,0)$ with respect to C is -3, while the power of $(3,0)$ with respect to C is 5.

The set V_i is constructed as being all those points whose power with respect to C_i is smaller than their power with respect to any other C_j. An example using three equally sized disks is given in Figure 4. In this case all three disks intersect V_2, while two disks intersect regions V_1 and V_3. However, the area of the union can be written as $|V_1 \cap B_1| + |V_2 \cap B_2| + |V_3 \cap B_3|$.

Note that if B_i is completely contained within B_j then the partition region V_i corresponding to B_i will be the empty set, as there will be no points with smaller power with respect to C_i than C_j. The use of the power diagram to compute the area of a union of open balls was originally proposed in Avis, Bhattacharya, and Imai 1988 and Edelsbrunner 1993. In two dimensions the power diagram can be constructed for n points

in $O(n \log n)$ time (Imai, Iri, and Murota 1985). For $d > 2$ the complexity is $O\left(n^{\lfloor \frac{d+1}{2} \rfloor}\right)$ (Aurenhammer 1987).

REFERENCES

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation : Algorithms and Analysis*. New York: Springer.

Asmussen, S., and R. Y. Rubinstein. 1995. "Steady state rare events simulation in queueing models and its complexity properties". In *Advances in Queueing*, 429–461. Boca Raton: CRC Press.

Aurenhammer, F. 1987. "Power diagrams: Properties, algorithms and applications". *SIAM Journal on Computing* 16 (1): 78–96.

Avis, D., B. Bhattacharya, and H. Imai. 1988. "Computing the volume of the union of spheres". *The Visual Computer* 3 (6): 323–328.

Billingsley, P. 1995.. *Probability and Measure*. 3rd ed. New York: J. Wiley & Sons.

Bollobás, B., and O. Riordan. 2006. *Percolation*. Cambridge: Cambridge University Press.

Botev, Z. I., and D. P. Kroese. 2012. "Efficient Monte Carlo simulation via the generalized splitting method". *Statistics and Computing* 22 (1): 1–16.

Brereton, T., C. Hirsch, D. P. Kroese, and V. Schmidt. 2014. "Pair connectedness and shortest-path scaling in critical continuum percolation". Submitted.

Chernyak, A. 2004. "Residual reliability of P-threshold graphs". *Discrete Applied Mathematics* 135 (1-3): 83 – 95.

Chiu, S. N., D. Stoyan, W. S. Kendall, and J. Mecke. 2013. *Stochastic Geometry and Its Applications*. Chichester: J. Wiley & Sons.

Colbourn, C. J. 1987. *The Combinatorics of Network Reliability*. New York: Oxford University Press, Inc.

Colbourn, C. J., A. Satyanarayana, C. Suffel, and K. Sutner. 1993. "Computing residual connectedness reliability for restricted networks". *Discrete Applied Mathematics* 44 (13): 221 – 232.

Edelsbrunner, H. 1993. "The union of balls and its dual shape". In *Proceedings of the Ninth Annual Symposium on Computational Geometry*, SCG '93, 218–231: ACM.

Elmallah, E. S. 1992. "Algorithms for K-terminal reliability problems with node failures". *Networks* 22 (4): 369–384.

Elperin, T., I. Gertsbakh, and M. Lomonosov. 1991, Dec. "Estimation of network reliability using graph evolution models". *IEEE Transactions on Reliability* 40 (5): 572–581.

Garvels, M. J. J., J.-K. C. W. van Ommeren, and D. P. Kroese. 2002. "On the importance function in splitting simulation". *European Transactions on Telecommunications* 13 (4): 363–371.

Gertsbakh, I. B., and Y. Shpungin. 2010. *Models of Network Reliability*. Boca Raton: CRC Press.

Glasserman, P., P. Heidelberger, P. Shahabuddin, and T. Zajic. 1999. "Multilevel splitting for estimating rare event probabilities". *Operations Research* 47 (4): 585–600.

Glynn, P. W., and D. L. Iglehart. 1989. "Importance sampling for stochastic simulations". *Management Science* 35 (11): 1367–1392.

Hui, K.-P., N. Bean, M. Kraetzl, and D. P. Kroese. 2005. "The cross-entropy method for network reliability estimation". *Annals of Operations Research* 134 (1): 101–118.

Imai, H., M. Iri, and K. Murota. 1985. "Voronoi diagram in the Laguerre geometry and its applications". *SIAM Journal on Computing* 14 (1): 93–105.

Kahn, H., and T. E. Harris. 1951. "Estimation of particle transmission by random sampling". *National Bureau of Standards Applied Mathematics Series* 12:27–30.

Kroese, D. P., T. Taimre, and Z. I. Botev. 2011. *Handbook of Monte Carlo Methods*. New York: J. Wiley & Sons.

L'Ecuyer, P., V. Demers, and B. Tuffin. 2006. "Splitting for rare-event simulation". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 184–191. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Li, J., and M. Östling. 2013. "Percolation thresholds of two-dimensional continuum systems of rectangles". *Phys. Rev. E* 88:012101.

Lomonosov, M. 1994. "On Monte Carlo Estimates in Network Reliability". *Probability in the Engineering and Informational Sciences* 8:245–264.

Quintanilla, J. A., and R. M. Ziff. 2007. "Asymmetry in the percolation thresholds of fully penetrable disks with two different radii". *Phys. Rev. E* 76:051115.

Rubinstein, R. Y., and D. P. Kroese. 2008. *Simulation and the Monte Carlo Method*. 2nd ed. New York: J. Wiley & Sons.

Sahini, M., and M. Sahimi. 1994. *Applications of Percolation Theory*. Bristol: Taylor & Francis.

Stenzel, O., L. J. A. Koster, R. Thiedmann, S. D. Oosterhout, R. A. J. Janssen, and V. Schmidt. 2012. "A new approach to model-based simulation of disordered polymer blend solar cells". *Advanced Functional Materials* 22 (6): 1236–1244.

Stivaros, C., and K. Sutner. 1997. "Computing optimal assignments for residual network reliability". *Discrete Applied Mathematics* 75 (3): 285 – 295.

Sutner, K., A. Satyanarayana, and C. Suffel. 1991. "The Complexity of the Residual Node Connectedness Reliability Problem". *SIAM Journal on Computing* 20 (1): 149–155.

Torquato, S., and Y. Jiao. 2012. "Effect of dimensionality on the continuum percolation of overlapping hyperspheres and hypercubes. II. Simulation results and analyses". *The Journal of Chemical Physics* 137 (7): 074106.

Weichenberg, G., V. W. S. Chan, and M. Medard. 2004. "High-reliability topological architectures for networks under stress". *IEEE Journal on Selected Areas in Communications* 22 (9): 1830–1845.

AUTHOR BIOGRAPHIES

ROHAN SHAH is a PhD student at the School of Mathematics and Physics of the University of Queensland. He has an honors degree in statistics from the University of Western Australia. His research interests include Monte Carlo methods, rare event simulation, stochastic geometry and statistical software. His email address is rohan.shah@uqconnect.edu.au.

CHRISTIAN HIRSCH is a PhD student at the Faculty of Mathematics and Economics of Ulm University. He has a Master (Diploma) in Mathematics from the Ludwig Maximilian University of Munich. His research interests include stochastic geometry, random geometric graphs, and Monte Carlo simulation of spatial stochastic models. His personal website can be found under http://www.uni-ulm.de/stochastik. His email address is christian.hirsch@uni-ulm.de.

DIRK P KROESE is a professor of Mathematics and Statistics at the University of Queensland. He is the co-author of several influential monographs on simulation and Monte Carlo methods, including *Handbook of Monte Carlo Methods and Simulation and the Monte Carlo Method*, (2nd Edition). Dirk is a pioneer of the well-known Cross-Entropy method — an adaptive Monte Carlo technique, invented by Reuven Rubinstein, which is being used around the world to help solve difficult estimation and optimization problems in science, engineering, and finance. His personal website can be found under http://www.maths.uq.edu.au/~kroese. His email address is kroese@maths.uq.edu.au.

VOLKER SCHMIDT is Professor at the Faculty of Mathematics and Economics of Ulm University. His research interests include stochastic geometry, spatial statistics, and Monte Carlo simulation of spatial stochastic models as well as their applications to structural analysis of (microscopic and geographically mapped) image data. He is (co-) author of more than 100 peer-reviewed publications, including several textbooks and monographs. His personal website can be found under http://www.uni-ulm.de/stochastik. His email address is volker.schmidt@uni-ulm.de.